

UNIVERSIDAD TÉCNICA DEL NORTE



Facultad de Ingeniería en Ciencias Aplicadas  
Carrera de Ingeniería en Sistemas Computacionales

**DETECCIÓN DE PATRONES DE CONTRABANDO PARA LA GESTIÓN DE  
INFORMACIÓN DE APREHENSIONES Y RETENCIONES UTILIZANDO TÉCNICAS  
DESCRIPTIVAS DE AGRUPAMIENTO, ASOCIACIÓN Y ATÍPICOS EN MINERÍA DE  
DATOS.**

Trabajo de Titulación previo a la obtención del título de Ingeniera en Sistemas  
Computacionales

Autora:

Diana Carolina Rosero Rea

Director:

PhD. Iván García Santillán

Ibarra - Ecuador

2021

## AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD



### UNIVERSIDAD TÉCNICA DEL NORTE BIBLIOTECA UNIVERSITARIA

#### AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

#### 1. IDENTIFICACIÓN DE LA OBRA

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago la entrega del presente trabajo a la Universidad Técnica del Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a disposición la siguiente información:

DATOS DE CONTACTO			
<b>CÉDULA DE IDENTIDAD:</b>	100356022-2		
<b>APELLIDOS Y NOMBRES:</b>	Rosero Rea Diana Carolina		
<b>DIRECCIÓN:</b>	Ernesto Monge Sandoval 358 – Ibarra - Imbabura		
<b>EMAIL:</b>	dcroseror@utn.edu.ec, droserorea@gmail.com		
<b>TELÉFONO FIJO:</b>	-	<b>TELÉFONO MÓVIL:</b>	099288030

DATOS DE LA OBRA	
<b>TÍTULO:</b>	DETECCIÓN DE PATRONES DE CONTRABANDO PARA LA GESTIÓN DE INFORMACIÓN DE APREHENSIONES Y RETENCIONES UTILIZANDO TÉCNICAS DESCRIPTIVAS DE AGRUPAMIENTO, ASOCIACIÓN Y ATÍPICOS EN MINERÍA DE DATOS.
<b>AUTOR (ES):</b>	Rosero Rea Diana Carolina
<b>FECHA: DD/MM/AAAA</b>	11/01/2021
SOLO PARA TRABAJOS DE GRADO	
<b>PROGRAMA:</b>	<input checked="" type="checkbox"/> PREGRADO <input type="checkbox"/> POSGRADO
<b>TÍTULO POR EL QUE OPTA:</b>	INGENIERA EN SISTEMAS COMPUTACIONALES
<b>ASESOR /DIRECTOR:</b>	PhD. Iván Danilo García Santillán

#### 2. CONSTANCIAS

El autor (es) manifiesta (n) que la obra objeto de la presente autorización es original y se la desarrolló, sin violar derechos de autor de terceros, por lo tanto la obra es original y que es (son) el (los) titular (es) de los derechos patrimoniales, por lo que asume (n) la responsabilidad sobre el contenido de la misma y saldrá (n) en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, a los 11 días del mes de enero de 2021

**EL AUTOR:**

(Firma)   
Nombre: Rosero Rea Diana Carolina

## **CERTIFICACIÓN DEL DIRECTOR DE TRABAJO DE GRADO**

En mi calidad de Tutor del Trabajo de Grado presentado por la egresada, **Rosero Rea Diana Carolina** para optar por el Título de Ingeniero en Sistemas Computacionales, cuyo tema es: **DETECCIÓN DE PATRONES DE CONTRABANDO PARA LA GESTIÓN DE INFORMACIÓN DE APREHENSIONES Y RETENCIONES UTILIZANDO TÉCNICAS DESCRIPTIVAS DE AGRUPAMIENTO, ASOCIACIÓN Y ATÍPICOS EN MINERÍA DE DATOS**. Considero que el presente trabajo reúne los requisitos y méritos suficientes para ser sometido a la presentación pública y evaluación por parte del tribunal examinador.

En la ciudad de Ibarra, a los 15 días del mes de diciembre del 2020.



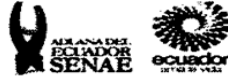
Firmado electrónicamente por:

**IVAN DANILO  
GARCIA  
SANTILLAN**

PhD. Iván Danilo García Santillán

**DIRECTOR TRABAJO DE GRADO**

# CERTIFICACIÓN INSTITUCIÓN BENEFICIARIA



## CUERPO DE VIGILANCIA ADUANERA

### CENTRO DE FORMACION DE VIGILANCIA ADUANERA

CERTIFICA

QUE: La Srta. DIANA CAROLINA ROSERO REA con cédula identidad 1003560222 estudiante de la Universidad Técnica del Norte de la Facultad de Ingeniería en Ciencias Aplicadas de la Carrera de Ingeniería en Sistemas Computacionales, ha desarrollado con los datos entregados por la Dirección Administrativa de la Aduana el Proyecto de Tesis **"DETECCIÓN DE PATRONES DE CONTRABANDO PARA LA GESTIÓN DE INFORMACIÓN DE APREHENSIONES Y RETENCIONES UTILIZANDO TÉCNICAS DESCRIPTIVAS DE AGRUPAMIENTO, ASOCIACIÓN Y ATÍPICOS EN MINERÍA DE DATOS"**.

QUE: El análisis del proyecto fue entregado al Centro de Formación de Vigilancia Aduanera acantonada en la ciudad de San Miguel de Ibarra el 14 de diciembre del 2020.

Es todo cuanto puedo certificar, facultando al Interesado hacer uso de este certificado como estime conveniente, excepto para trámites judiciales.

Ibarra, 14 de diciembre del 2020



Atentamente

Cnel. Mbá. Edgar Duque Ch

Coordinador (E) Centro de Formación de Vigilancia Aduanera



## Dedicatoria

A la mejor mamá del mundo, María del Carmen, quien me ha motivado, me ha llenado de gran valor siempre y me ha enseñado que lo mejor que puede tener una persona es pasión por su profesión y dedicación para llegar al éxito, a mi padre José Stalin por su ejemplo de constancia y trabajo.

A mis hermanas Andrea Fernanda y María José que con su cariño y respaldo están pendientes siempre de mí y me han brindado grandes consejos para llegar a donde hoy me encuentro.

A una de mis personas favoritas, mi compañero de aventuras Harold Andres que me alentó a diario y a cada una de las personas que en mi paso por la universidad hemos compartido gratos e inolvidables momentos mis amigos.

A la memoria de la persona que tuvo el impulso de promover la tecnología, investigación y aplicación de minería de datos en su profesión y trabajo por el bien de una sociedad más justa donde el contrabando no afecte enormemente al trabajo honesto del productor y comerciante ecuatoriano, gracias por su dedicación, esfuerzo y apoyo brindado Crnl. Néstor Marcelo Esparza en nuestra memoria siempre.

*Diana Rosero Rea*

## **Agradecimiento**

El presente trabajo de tesis primeramente me gustaría agradecer a Dios por bendecirme para llegar hasta donde he llegado, porque ya es realidad la una de mis metas tan anheladas, a mi tutor PhD. Iván García Santillán por todos sus conocimientos brindados, su paciencia, apoyo y trabajo para poder culminarlo exitosamente.

Gracias infinitas a todos quienes integran la carrera de Ingeniería en Sistemas Computacionales de quienes me llevo los más gratos recuerdos y a la Universidad Técnica del Norte por darme la oportunidad de estudiar y ser una gran profesional.

Al Cuerpo de Vigilancia Aduanera – Centro de Formación y Vigilancia Aduanera por la apertura y apoyo brindando en el transcurso y desarrollo del proyecto, inmensamente agradecida con una de las personas que conocí dentro su gran institución.

*Diana Rosero Rea*

## Tabla de contenido

AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD.....	II
CERTIFICACIÓN DEL DIRECTOR DE TRABAJO DE GRADO.....	III
CERTIFICACIÓN INSTITUCIÓN BENEFICIARIA .....	IV
Dedicatoria.....	V
Agradecimiento.....	VI
Resumen .....	XIV
Abstract.....	XIV
INTRODUCCIÓN.....	XV
Antecedentes .....	XV
Problema .....	XVI
Objetivos .....	XVI
Objetivo General .....	XVI
Objetivos Específicos.....	XVI
Justificación .....	XVI
Alcance .....	XVIII
CAPÍTULO 1 .....	1
Marco Teórico.....	1
1.1.    Contrabando en Ecuador .....	1
1.1.1    Atribuciones y responsabilidades de la entidad de control .....	1
1.1.2    Aprehensión y retención.....	2
1.1.3    Situación actual del contrabando en Ecuador .....	3
1.2.    Introducción a minería de datos .....	4
1.2.1.    Minería de Datos.....	4
1.2.2.    Características de la Minería de Datos .....	4
1.2.3.    Relación de la minería de datos con otras disciplinas .....	5
1.3.    Tipos de datos .....	6
1.3.1.    Clasificación de los datos .....	7
1.4.    Técnicas de Minería de datos.....	8
1.4.1.    Técnicas Descriptivas .....	8
1.4.2.    Técnicas Predictivas .....	12

1.4.3.	Herramientas para minería de datos.....	12
1.4.4.	Herramientas de análisis empresarial .....	14
1.5.	Proceso KDD (Knowledge Discover in Databases) .....	14
1.5.1.	Fases del proceso KDD .....	15
1.6.	ISO/IEC 25012:2008.....	17
1.6.1.	Calidad de Datos Inherente: .....	17
1.6.2.	Calidad de Datos Dependiente del Sistema:.....	18
1.7.	Trabajos existentes.....	20
CAPÍTULO 2 .....		23
Proceso de descubrimiento del Conocimiento.....		23
2.1.	Visión general del proyecto.....	23
2.2.	Entregables del proyecto .....	23
2.3.	Organización del proyecto .....	24
2.3.1.	Participantes del Proyecto.....	24
2.3.2.	Roles y Responsabilidades .....	24
2.4.	Gestión del proyecto.....	25
2.4.1.	Estimaciones .....	25
2.4.2.	Plan del proyecto .....	26
2.5.	Recopilación de datos .....	27
2.5.1.	Tipos de bases de datos .....	27
2.6.	Fase de selección, limpieza y transformación.....	28
2.6.1.	Selección .....	29
2.6.2.	Transformación.....	31
2.6.3.	Limpieza .....	42
2.6.4.	Aplicación de Normativa ISO25012 .....	46
2.7.	Minería de datos .....	47
2.7.1.	Componentes Principales .....	48
2.7.2.	Agrupamiento (Clustering) .....	49
2.7.3.	Asociación .....	50
2.7.4.	Atípicos.....	52
CAPÍTULO 3 .....		53



Proceso de descubrimiento del Conocimiento.....	53
3.1. Evaluación e interpretación .....	53
3.1.1. Evaluación, análisis e interpretación de tareas de Clustering .....	53
3.1.2. Evaluación, análisis e interpretación de tareas de Asociación .....	71
3.2. Atípicos.....	87
3.3. Obtención del conocimiento.....	97
3.4. Resumen Ejecutivo del conocimiento obtenido .....	107
3.5. Análisis de impacto .....	108
LIMITACIONES.....	112
CONCLUSIONES.....	112
RECOMENDACIONES.....	114
ANEXOS.....	116
REFERENCIAS.....	128

## Índice de Figuras

Figura 1- Porcentaje correspondiente a los artículos decomisados.....	3
Figura 2 - Relación de la Minería de Datos con otras disciplinas .....	5
Figura 3 - Técnicas de minería de datos .....	8
Figura 4 - Proceso KDD.....	15
Figura 5 - ETL realizado en Pentaho.....	28
Figura 6 - Atributos base de datos inicial (22) años 2014-2018 .....	29
Figura 7 - Atributos relevantes (16) años 2014-2018 .....	29
Figura 8 – Atributos (22) base de datos inicial año 2019 .....	30
Figura 9 – Filtro de registro de grupos excluidos.....	31
Figura 10 – Normalización de atributo Grupo .....	32
Figura 11- Categorización por rangos atributo PRECIO .....	39
Figura 12 - Error en atributo DISTRITO con información incorrecta .....	43
Figura 13 - Errores de digitación en atributo BODEGA y MARCA.....	43
Figura 14 - Errores de digitación en atributo SITIO APREHENSION y GRUPO OPERATIVO .....	44
Figura 15 - Mapeo del atributo Distrito con registro erróneo DOM .....	44
Figura 16 - Eliminación de caracteres especiales .....	45
Figura 17 - Registros únicamente en mayúsculas.....	45
Figura 18 – Vista minable para asociación.....	47
Figura 19 - Vista minable para atípicos/agrupamiento .....	48
Figura 20 – Resultados algoritmo PCA en Weka.....	48
Figura 21 – Parámetros del algoritmo <i>SimpleKmeans</i> en Weka.....	50
Figura 22 - Parámetros del algoritmo <i>EM</i> en Weka.....	50
Figura 23 - Parámetros del algoritmo <i>Apriori</i> en Weka.....	51
Figura 24 – Parámetros del algoritmo <i>FP Growth</i> en Weka .....	52
Figura 25 - Resultados algoritmo <i>EM</i> .....	58
Figura 26 - Resultados algoritmo <i>apriori</i> para el conjunto de datos total .....	71
Figura 27 – Resultado algoritmo <i>Apriori</i> conjunto de datos PC1 .....	73
Figura 28 - Resultado algoritmo <i>Apriori</i> conjunto de datos PC2 .....	74
Figura 29 - Resultado algoritmo <i>Apriori</i> conjunto de datos PC3 .....	75
Figura 30 - Resultados algoritmo <i>FP Growth</i> para el conjunto de datos total .....	79
Figura 31 - Resultado algoritmo <i>FP Growth</i> conjunto de datos PC1 .....	81
Figura 32- Resultado algoritmo <i>FP Growth</i> conjunto de datos PC2 .....	82
Figura 33 - Resultado algoritmo <i>FP Growth</i> conjunto de datos PC3 .....	83
Figura 34 – Diagrama de cajas atributo Total.....	88
Figura 35 - Estadística del atributo CAT_CANTIDAD.....	91
Figura 36 - Estadísticas del atributo UNIDADES.....	91
Figura 37 - Estadística del atributo CAT_TOTAL .....	92

Figura 38 – Estadísticas del atributo GRUPO .....	92
Figura 39 – Estadística del atributo SUBGRUPO .....	93
Figura 40 - Estadísticas del atributo BODEGA.....	94
Figura 41 - Estadísticas del atributo SITIO_ APREHENSION.....	94
Figura 42 – Mapa de ubicaciones geográficas de Sitios de Aprehensión.....	95
Figura 43 - Estadísticas del atributo GRUPO_OPERATIVO.....	95
Figura 44 - Estadísticas del atributo DISTRITO.....	96
Figura 45 - Estadísticas del atributo ZONA .....	96

## Índice de Tablas

Tabla 1 - Representación de datos .....	7
Tabla 2 - Simbología Ecuación 1.....	16
Tabla 3 - Simbología Ecuación 2.....	16
Tabla 4 - Simbología Ecuación 3.....	17
Tabla 5 - Entregables del Proyecto .....	23
Tabla 6 - Directores de las Áreas Comprendidas .....	24
Tabla 7 - Participantes Directos del Proyecto.....	24
Tabla 8 – Roles y responsabilidades del proyecto .....	24
Tabla 9 - Talento Humano.....	25
Tabla 10 - Recursos Materiales.....	25
Tabla 11 - Costo total del proyecto.....	26
Tabla 12 - Distribución de horas .....	26
Tabla 13 – Atributos y tipos de datos de la información .....	27
Tabla 14 - Categorías para el atributo UNIDADES .....	31
Tabla 15 - Categorías para el atributo GRUPO.....	32
Tabla 16 - Categorización del atributo SUBGRUPO (muestra del total) .....	33
Tabla 17 - Categorización de atributo PROCEDENCIA .....	33
Tabla 18 - Categorización de atributo BODEGA.....	34
Tabla 19- Categorización de atributo MARCA .....	34
Tabla 20 - Categorización de atributo ESTATUS .....	34
Tabla 21 - Categorización de atributo ORIGEN_ APREHENSION .....	35
Tabla 22 - Categorización de atributo SITIO_ APREHENSION.....	35
Tabla 23 - Categorización de atributo GRUPO_OPERATIVO.....	36
Tabla 24 - Categorización de atributo DISTRITO.....	36
Tabla 25 - Relación atributo Zonas con Distrito.....	36
Tabla 26 - Categorización de atributo CANTIDAD .....	37
Tabla 27 - Categorización atributo TOTAL .....	38

Tabla 28 - Categorización atributo PRECIO.....	38
Tabla 29- Escala de valoración para característica de <i>Consistencia</i> .....	46
Tabla 30 – Valores de aplicación de métrica .....	46
Tabla 31 - Atributos mejor clasificados mediante PCA.....	49
Tabla 32 - Resultado de algoritmo <i>KMeans</i> (Parte 1).....	53
Tabla 33 - Resultado de algoritmo <i>KMeans</i> (Parte 2).....	53
Tabla 34 - Resultado de algoritmo <i>KMeans</i> PC1 (Parte 1).....	54
Tabla 35 - Resultado de algoritmo <i>KMeans</i> PC1 (Parte 2).....	55
Tabla 36 - Resultado de algoritmo <i>KMeans</i> PC2 (Parte 1).....	55
Tabla 37 - Resultado de algoritmo <i>KMeans</i> PC2 (Parte 2).....	55
Tabla 38 - Resultado de algoritmo <i>KMeans</i> PC3 (Parte 1).....	56
Tabla 39 - Resultado de algoritmo <i>KMeans</i> PC3 (Parte 2).....	56
Tabla 40 – Resultado de algoritmo <i>EM</i> (Parte 1).....	58
Tabla 41 - Resultado de algoritmo <i>EM</i> (Parte 2).....	59
Tabla 42 – Clústeres desnormalizados del algoritmo <i>EM</i> (Parte 1).....	62
Tabla 43 - Clústeres desnormalizados del algoritmo <i>EM</i> (Parte 2).....	63
Tabla 44 - Resultado de algoritmo <i>EM</i> conjunto de datos PC1 (Parte 1) .....	64
Tabla 45 - Resultado de algoritmo <i>EM</i> conjunto de datos PC1 (Parte 2) .....	64
Tabla 46 - Clústeres desnormalizados del algoritmo <i>EM</i> para el PC1 (Parte 1) .....	65
Tabla 47 - Clústeres desnormalizados del algoritmo <i>EM</i> para el PC1 (Parte 2) .....	66
Tabla 48 - Resultado de algoritmo <i>EM</i> conjunto de datos PC2 (Parte 1) .....	66
Tabla 49 - Resultado de algoritmo <i>EM</i> conjunto de datos PC2 (Parte 2) .....	66
Tabla 50 - Clústeres desnormalizados del algoritmo <i>EM</i> para el PC2 (Parte 1) .....	68
Tabla 51 - Clústeres desnormalizados del algoritmo <i>EM</i> para el PC2 (Parte 2) .....	68
Tabla 52 -Resultado de algoritmo <i>EM</i> conjunto de datos PC3 (Parte 1) .....	68
Tabla 53 - Resultado de algoritmo <i>EM</i> conjunto de datos PC3 (Parte 2) .....	68
Tabla 54 - Clústeres desnormalizados del algoritmo <i>EM</i> para el PC3 (Parte 1) .....	70
Tabla 55 - Desnormalizados del algoritmo <i>EM</i> para el PC3 (Parte 1).....	70
Tabla 56 - Reglas de asociación, soporte y confianza algoritmo <i>Apriori</i> conjunto de datos completos.....	72
Tabla 57 - Reglas de asociación, soporte y confianza resultado del algoritmo <i>Apriori</i> PC1.....	76
Tabla 58 - Reglas de asociación, soporte y confianza resultado del algoritmo <i>Apriori</i> PC2.....	76
Tabla 59 - Reglas de asociación, soporte y confianza resultado del algoritmo <i>Apriori</i> PC3.....	78
Tabla 60 – Listado de mejores reglas de asociación con el algoritmo <i>Apriori</i> .....	79
Tabla 61 - Reglas de asociación, soporte y confianza resultado del algoritmo <i>FP Growth</i> conjunto de datos completos .....	80
Tabla 62 - Reglas de asociación, soporte y confianza resultado del algoritmo <i>FP Growth</i> PC1...	84
Tabla 63 - Reglas de asociación, soporte y confianza resultado del algoritmo <i>FP Growth</i> PC2...	84

Tabla 64 - Reglas de asociación, soporte y confianza resultado del algoritmo <i>FP Growth</i> PC3...	85
Tabla 65- Listado de mejores reglas de asociación según el algoritmo <i>FP Growth</i> .....	86
Tabla 66 - Formulas de cálculo por cuartil.....	87
Tabla 67 – Calculo de cuartiles, mediana y límites.....	88
Tabla 68 - Niveles de Impacto.....	109
Tabla 69 - Impacto Sociocultural .....	109
Tabla 70 - Impacto Económico.....	110
Tabla 71 - Impacto Tecnológico .....	111

## Índice de Anexos

Anexo 1 Carta de confidencialidad entre las partes .....	116
Anexo 2 Preguntas del negocio para análisis exploratorio (BI) .....	119
Anexo 3 Dashboards para responder preguntas del negocio propuesta en el Anexo2 .....	120
Anexo 4 Formato sugerido para ingreso de datos (Macro Excel) .....	127

## Resumen

El ingreso de mercadería por pasos ilegales conocido como contrabando es delito que aqueja a la sociedad en la que nos desarrollamos diariamente ya que a consecuencia de ellos productores y comerciantes nacionales se ven afectados por la poca competencia e ingresos que generan. El presente trabajo de investigación permitió identificar los principales patrones de contrabando dentro del territorio ecuatoriano, por medio de la aplicación de las técnicas de minería de datos como: agrupamiento, asociación y atípicos. Se analizó el conjunto de datos correspondientes a aprehensiones y retenciones de los años 2014 a 2018 con un total de 102667 instancias se desarrolló cada una de las fases respectivas de minería de datos con el objetivo de obtener el conocimiento deseado usando el software Pentaho Data Integration (PDI) y Weka respectivamente. Los resultados más relevantes se los obtuvo en la técnica de agrupamiento (clustering) donde se identificó información estratégica para la toma de decisiones dentro de la entidad de control.

**Palabras Clave:** contrabando, minería de datos, mercadería ilegal, técnicas descriptivas, delito.

## Abstract

The entry of merchandise through illegal steps known as smuggling is a crime that afflicts the society in which we develop daily since as a result of them national producers and merchants are affected by the little competition and income they generate. The present research work allowed to identify the main smuggling patterns within the Ecuadorian territory, through the application of data mining techniques such as: grouping, association and atypical. The data set corresponding to apprehensions and detentions from 2014 to 2018 was analyzed with a total of 102667 instances, each of the respective data mining phases was developed with the aim of obtaining the desired knowledge using the Pentaho Data Integration software (PDI) and Weka respectively. The most relevant results were obtained in the clustering technique where strategic information is identified for decision-making within the control entity.

**Key Words:** smuggling, data mining, illegal merchandise, descriptive techniques, crime.

## INTRODUCCIÓN

### **Antecedentes**

El ingreso de diversos tipos de mercadería de manera ilícita está presente dentro de todo el territorio ecuatoriano, mismos que son considerados como delitos aduaneros y de evasión tributaria.

El contrabando es considerado como una actividad generalizada en la cultura ecuatoriana. Es reconocido en el medio como uno de los mecanismos de intercambio comercial en el cual se presenta el uso especulativo de dinero, capitales oportunidades, negocio fácil y ganancias desproporcionadas por la intermediación que estas actividades generan.

Los avances y logros significativos alcanzados por la institución pública encargada de la regulación y control de este delito han permitido identificar las zonas de frontera donde la tarea debe continuar con las respectivas mejoras en recursos técnicos para la obtención de mayores resultados (Chiriboga Zambrano, 2015).

El número de aprehensiones totales por año realizadas como productos de los operativos de lucha contra el contrabando y la defraudación fiscal han ido en aumento, según gráficos del Informe de Gestión Ecuador 2017 que indica de manera generalizada cuales son los productos con más altos valores monetarios que se han retenido, evidenciando a los cigarrillos como el producto con un valor aproximado de 11.5 millones equivalente a un aproximado de 33 millones de unidades (SENAE, 2018).

La brecha entre la oferta y demanda se produce por un comercio ilícito que permite ofrecer productos específicos considerados de buena calidad que no están en el mercado convencional en valores inferiores a los reglamentarios. Los tipos de mercancías más concurridas son ropa, calzado, licores, dispositivos electrónicos y cigarrillos, que son ingresados al país por varias rutas que son difíciles de establecer por la capacidad de los contrabandistas para aperturar nuevos pasos clandestinos no habilitados con el objetivo principal de evadir los controles en pasos terrestres, marítimos y aéreos (Salazar Méndez, 2015).

## **Problema**

Anualmente se realizan un gran número de aprehensiones y retenciones de mercadería de diferentes tipos en varios puntos fronterizos del territorio ecuatoriano (Salazar Méndez, 2015). Los distritos de control aduanero se encuentran en constante planificación de operativos y patrullajes para vigilar el ingreso de mercadería de forma ilegal en la lucha contra el contrabando. Se ha realizado reformas sobre consideraciones, normativas y regulaciones con el fin de conseguir mejores resultados, con todo el trabajo realizado no se ha considerado las características potenciales de las situaciones y condiciones de contrabando o los casos singulares que se presentan en el cometimiento de este ilícito.

## **Objetivos**

### **Objetivo General**

Identificar patrones de contrabando para la gestión de información en aprehensiones y retenciones mediante técnicas descriptivas de agrupamiento, asociación y atípicos en minería de datos.

### **Objetivos Específicos**

Sustentar las técnicas descriptivas de minería de datos mediante la composición de una base teórica del proceso de descubrimiento de conocimiento en bases de datos (KDD).

Elaborar un conjunto de datos cualitativos y cuantitativos de las retenciones y aprehensiones realizadas en determinado período de tiempo usando el proceso KDD para la extracción de datos y la herramienta Weka para el análisis.

Validar los resultados obtenidos utilizando métodos estadísticos y la característica de consistencia de la norma ISO/IEC 25012 para calidad de los datos obtenidos.

## **Justificación**

Objetivos de Desarrollo Sostenible (ODS)

Uno de los ODS particularmente el objetivo 16 alude a Paz, Justicia e Instituciones Sólidas , haciendo referencia a la meta número 16.6 orientada a crear a todos los niveles instituciones eficaces y transparentes que rindan cuentas (Naciones Unidas, 2015), por este motivo la



eficiencia en la detección de aprehensiones y retenciones en la institución de regulación y control del ingreso de mercadería por contrabando acrecentará los resultados en la mitigación del ilícito.

#### Justificación Ambiental

Los productos alimenticios que ingresan de forma ilegal al territorio nacional afecta directamente al sector agrícola y ambiental por todas las pérdidas que se presentan por la competencia desleal que provoca el ingreso de productos alimenticios por contrabando (SENAE, 2019), con la identificación de esta problemática permitirá contribuir al sector para mantenerlo adecuadamente y evitar cualquier inconveniente futuro.

#### Justificación Tecnológica

Uno de los programas de implementación de Económica Digital y Tecnologías Emergentes, haciendo énfasis en el último presenta una de sus líneas de acción en promover el uso de servicios de análisis de grandes volúmenes de datos (Ministerio de Telecomunicaciones, 2018), esto permitirá una toma de decisiones con mayor fundamento para una mejor gestión de acción e información por parte de la institución de regulación y control del ingreso de mercadería por contrabando.

#### Justificación Social

El contrabando es un problema con nuevos protagonistas en un mundo globalizado, las altas ganancias se conjugan con riesgos y formas de extorsión, secuestro y sicariato a los que cualquier miembro de la sociedad ecuatoriana puede verse involucrado (Ruiz, 2015), por esta razón de alta importancia prever las posibles características, geográficas, de tiempo, sociales, gubernamentales, climatológicas para casos de contrabando de mercadería, contribuiría un gran aporte para mejorar los resultados de los patrullajes y operativos de control, en donde un enfoque más acertado permitirá que los casos de este ilícito se presenten. Esta investigación será desarrollada para detectar los patrones, clústeres y atípicos que más se presentan en los casos de contrabando, de un período de años pasados.

El presente trabajo de investigación tiene como foco principal la aplicación de técnicas descriptivas de agrupamiento y asociación en minería de datos que permitirá a las personas

encargadas de la planificación de los controles fijos, móviles y patrullajes la toma de decisiones estratégicas para beneficio de la institución encargada con respecto a recursos humanos y económicos en mira de obtener cambios en los índices de contrabando y construir el inicio de una sociedad segura y consciente de la gravedad de este ilícito; no solo se ve afectada la economía sino además la seguridad ciudadana sufre un desbalance por todos los factores humanos que deben estar inmersos en el cometimiento del delito de contrabando.

Esencialmente promover y prevenir a la sociedad de verse involucrado en delitos por contrabando será una tarea que iniciará por parte de la institución de control, con el uso de la información obtenida y la toma de decisiones precisas y preventivas.

### **Alcance**

El estudio propuesto tiene como fin obtener los patrones, clústeres y atípicos provenientes del contrabando con el uso de técnicas descriptivas de agrupamiento, asociación y atípicos de minería de datos. Para la consecución del mismo se realizará un proceso que inicia con la obtención de los datos de las aprehensiones realizadas en un periodo de años pasados (2014-2019), seguidamente se realizará el proceso ETL correspondiente a extracción, transformación y limpieza de la información con la herramienta de Pentaho 7 en un ambiente local, para proseguir con la obtención de la vista minable ,donde se aplicará diferentes algoritmos descriptivos de minería de datos utilizando la herramienta Weka 8.3 para finalmente aplicar métodos estadísticos y verificar la calidad de los datos obtenidos mediante la característica de consistencia de la norma ISO/IEC 25012.

El conjunto de datos obtenidos permitirá realizar una buena interpretación de los requisitos para una mejor gestión de la información en aprehensiones y retenciones de mercadería y personas que ingresa por contrabando, mediante la obtención de reportes con respecto a ubicación (distritos), tiempo, fechas, grupo operativo, rangos de tiempo, tipo de mercadería, situaciones sociales y políticas, entre otros. Adicionalmente, se planteará un nuevo formato de ingreso de datos para la gestión de aprehensiones y retenciones con miras a un análisis más exhaustivo de la información.

# CAPÍTULO 1

## Marco Teórico

### 1.1. Contrabando en Ecuador

El significado de la palabra contrabando hace referencia a la infracción, delito fiscal o tributario que consiste en la producción, circulación, importación o distribución prohibida de bienes o mercancías (Borja-Cevallos, 2020). Todas estas acciones se encuentran enmarcas dentro del contexto de un delito o un ilícito por la evasión de leyes que tiene inmerso, comenzando con la elusión del pago de los impuestos o tasas arancelarias establecidas, y con el cometimiento de otros delitos como actos delincuenciales, violencia, chantaje entre otros.

Cada estado mediante leyes comerciales o aduaneras establecen las reglas sobre la importación o exportación de productos y por ende los tributos que se deben pagar, dentro del Código Orgánico de la Producción, Comercio e Inversiones Capítulo III - De las Contravenciones en su Artículo 190 detalla lo que respecta contravenciones aduaneras sobre el ingreso de mercadería ilícitamente especificando lo siguiente: “las conductas tipificadas en el Código Orgánico Integral Penal, respecto de mercancías cuya cuantía sea inferior a diez salarios básicos unificados del trabajador en general” (Dirección Nacional Jurídica, 2018), misma acción se encuentra sancionada con una multa de tres veces el valor de la mercancía materia de la infracción según el Artículo 191 literal g.

Las atribuciones de la entidad de control se encuentran establecidas en los Arts. 176 y 211 Literales a) y b) del Código Orgánico de la Producción, Comercio e Inversiones, así como el Art. 259 del Código Orgánico de Entidades de Seguridad Ciudadana y Orden Público “COESCOP” sobre las Funciones y Responsabilidades del Cuerpo de Vigilancia Aduanera y en virtud de lo dispuesto en el numeral 1 del Art. 479 del Código Orgánico Integral Penal.

#### 1.1.1 Atribuciones y responsabilidades de la entidad de control

La entidad responsable de control de ingreso de mercadería ilegal ejecuta sus estrategias que permitan proteger el comercio formal y aportar con el desarrollo económico del país; por

este motivo tiene a su cargo el cumplimiento las siguientes responsabilidades sobre cada una de sus jurisdicciones dentro del todo el territorio nacional (Aduana del Ecuador, 2020):

- Vigilar a las personas, mercancías y medios de transporte en las Zonas Secundarias.
- Realizar las investigaciones técnicas correspondientes para la detección del delito aduanero, además de planificar y ejecutar las operaciones correspondientes en sectores establecidos.
- Aprender y retener provisionalmente mercaderías tanto abandonados como rezagas o que ingresen al territorio por pasos no legales.
- Aprender mercancías y elementos de convicción o evidencia de una infracción aduanera y ponerlas a disposición de la Dirección Distrital.
- Capturas a los responsables de los ilícitos cometidos y ponerlos a disposición de las autoridades competentes.
- Colaborar en el control del tráfico ilícito de sustancias estupefacientes, psicotrópicas y precursores, además armas, municiones, explosivos y control del lavado de activos todo lo antes mencionado en conformidad con la ley.

### **1.1.2 Aprehensión y retención**

Se considera como aprehensión a la toma forzosa por parte de la unidad operativa de la Entidad de control, siendo la responsable del control posterior de mercancías, medios de transporte o cualquier otro bien que pueda ser elemento de convicción o evidencia de la comisión como parte de una infracción aduanera según a lo que estable la normativa vigente (SENAE, 2016).

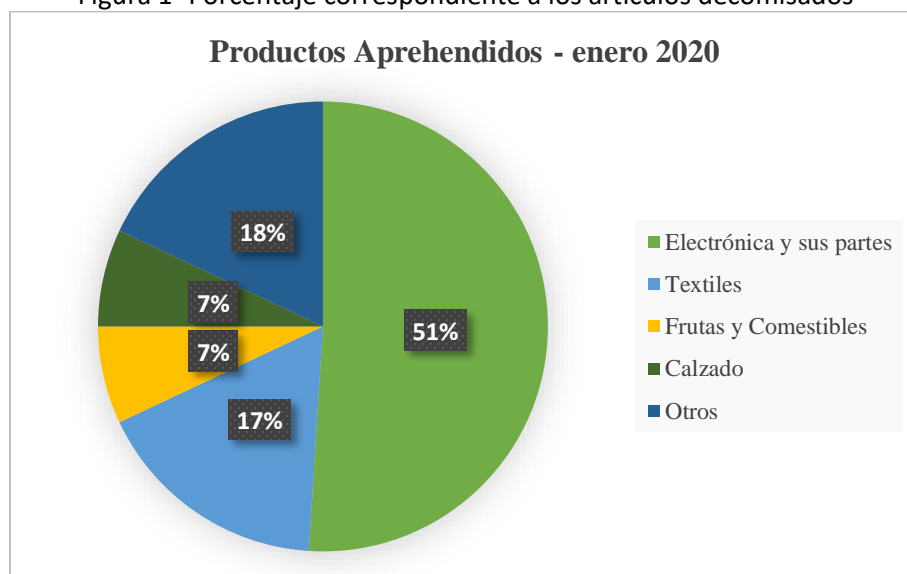
Dentro de Artículo 31 correspondiente al Procedimiento general para el despacho de equipaje de pasajeros y tripulantes que ingresan o salen de los aeropuertos internacionales del Ecuador correspondiente a la retención de mercancías, manifiesta que la retención se generará como resultado del aforo físico en donde se identificaron bienes tributables; la retención de dichos bienes será de 5 días hábiles, contados desde su arribo al país (Resolución Nro. SENAE-SENAE-2017-0345-RE, 2017). Además, durante el tiempo permitido se debe realizar la cancelación de los tributos correspondientes, presentar las autorizaciones,

permisos y registros según la normativa vigente, se puede solicitar o disponer de la devolución de mercancías al exterior dependiente del caso, sino todos los bienes retenidos caerán un abandono definitivo.

### 1.1.3 Situación actual del contrabando en Ecuador

Las cifras de aprehensiones de mercancía ilegal van en incremento por la intensificación de controles aduaneros a nivel nacional. De igual manera, la renovación del personal a cargo de los operativos hacen que se registre un aumento en el decomiso de mercancía ilegales que intentaban entrar al país (SENAE, 2020). Tan solo en el mes de enero del año 2020 la institución encargada del control de este ilícito aprehendió una cantidad de USD 6,40 millones entre diferentes tipos de mercaderías que se detalla en la Fig. 1 a continuación:

Figura 1- Porcentaje correspondiente a los artículos decomisados



Fuente: (SENAE, 2020)

El ingreso de mercadería en los primeros días del año 2020 denota como el incremento del delito de contrabando está presente y como el trabajo de la entidad de control juega un papel importante para mantener el orden en el territorio nacional. Además, se puede evidenciar en el gráfico como la categorización de la mercancía es global para un conocimiento general y como la categoría denominada como Electrónica y sus partes abarca más del 50% de las aprehensiones realizadas.

## **1.2. Introducción a minería de datos**

La minería de datos nace de la unión de dos términos como son dato y minería, siendo el primero formalmente un valor que toma una variable y el segundo considerado un arte para la extracción de metales preciados (Lara Torralbo, 2014), dando lugar a un término que genera revolución en el mundo de la informática y en desarrollo de las tecnologías orientada a la analítica de datos para la solución de problemas dentro de otras disciplinas.

### **1.2.1. Minería de Datos**

Según Lara (2014) define a la minería de datos como: “Una disciplina de la informática que estudia el análisis de grandes cantidades de datos con el objetivo de obtener conocimiento a partir de ellos” (p.11). El descubrimiento del conocimiento es la esencia de esta disciplina y la aplicabilidad en diferentes vertientes de datos de cualquier índole que se presentan en grandes cantidades.

La evolución de la concepciones y denominación de minería de datos como arqueología de datos, minería de base de datos; en la década de 1990 varios investigadores llegan al consenso de denominarla como el conjunto de base de datos, la información, extracción y conocimiento (Camana, 2016).

### **1.2.2. Características de la Minería de Datos**

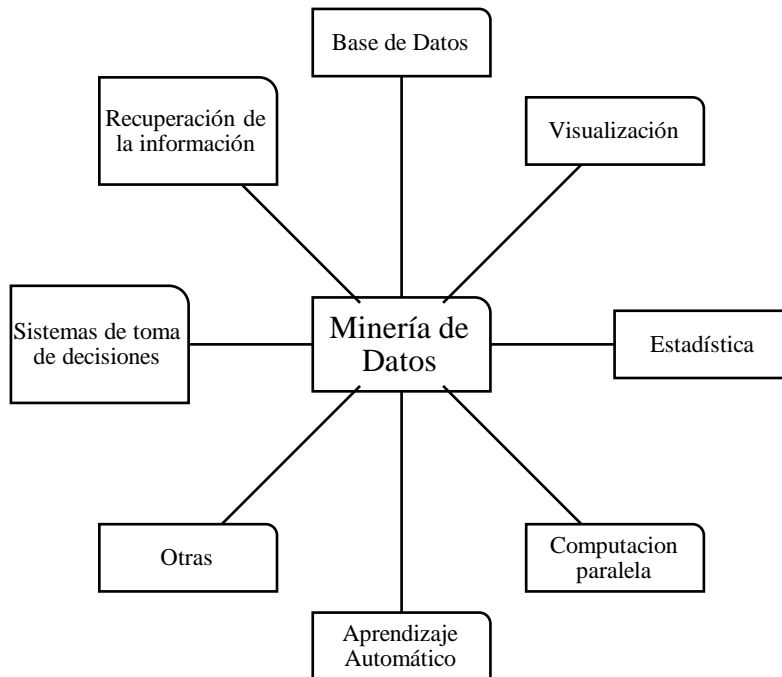
Jiménez Berríos (2015) detalla las siguientes características sobre minería de datos:

- Su principal característica es la capacidad inductiva de descubrimiento de información y patrones clases (conocimiento) que se encuentran ocultos en los datos.
- La exploración que realiza basada en la información encontrada en almacenes de datos de gran tamaño, en bases de datos y en repositorios correspondiente a un grupo de años.
- Es un proceso que invierte la dinámica del método científico, es decir, se formula la hipótesis de manera inicial y luego diseña el experimento para coleccionar los datos que confirmen o refuten la hipótesis.

### 1.2.3. Relación de la minería de datos con otras disciplinas

El surgimiento de la minería de datos procede de la mezcla de conceptos procedentes de muchas y diferentes disciplinas (Lara Torralbo, 2014), por ello se puede relacionar con otras disciplinas como son:

Figura 2 - Relación de la Minería de Datos con otras disciplinas



Fuente: (Hernández et al., 2004)

- **Base de datos**

Envuelve los conceptos de almacenes de datos, mismo en que su mayoría son de tipo transaccional como los gestores de bases de datos. Su importancia radica en las técnicas de indexación que permiten el acceso eficiente a los datos e información (Cisneros Buitrón, 2019).

- **Visualización**

El requisito fundamental para considerar útil los resultados de la minería de datos es que pueda ser visualizado por los expertos de cada dominio empleando diagramas, gráficos resúmenes entre otros (Lara Torralbo, 2014).

- **Estadística**

Muchos de los conceptos y técnicas que se emplean en la estadística, tales como modelos matemáticos o inferencias basadas en probabilidades, se utilizan también en minería de datos (Vila Espinosa, 2019).

- **Computación paralela y distribuida**

Permite repartir entre diferentes procesadores o computadores las tareas más complejas de minería de datos, incluyendo las tecnologías de procesamiento paralelo, distribuido o computación grid (Jiménez Berríos, 2015).

- **Aprendizaje Automático**

Tanto la minería de datos como el aprendizaje automático persiguen la obtención de modelos por medio de mecanismos automáticos (Lara Torralbo, 2014).

- **Sistema de toma de decisiones**

Son herramientas y sistemas informatizados que tienen como objetivo proporcionar información necesaria para realizar decisiones efectivas en el ámbito empresarial, es decir, ayudan a los directivos en la resolución de problemas y toma de decisiones (Jiménez Berríos, 2015).

- **Recuperación de la información**

Busca obtener información desde datos textuales, basando su desarrollo histórico en el uso de bibliotecas (actualmente bibliotecas digitales) (Jiménez Berríos, 2015).

- **Otras**

La relación de la minería de datos con otras disciplinas depende de los tipos de datos a ser minados, como por ejemplo: el análisis de imágenes, procesamiento de señales entre otros (Cisneros Buitrón, 2019).

### **1.3. Tipos de datos**

Se los considera una corrección de entidades mapeadas en un dominio de interés, su representación se base en la relación existente entre un conjunto de atributos de datos. Estos representan las propiedades y características básicas de los datos y se los conoce como: variables, campos o características (Hernández Valadez, 2006). Un dato se describe en función del conjunto de atributos como se muestra en la Tabla 1 un ejemplo sencillo de datos de personas con su edad y ciudad de procedencia, además se lo conoce como punto, muestra, entidad o instancia.



Tabla 1 - Representación de datos  
ATRIBUTOS

ID	NOMBRE	EDAD	CUIDAD
1	Andrea	37	Ibarra
2	José	56	Quito
3	Fernando	44	Ibarra
4	Julián	26	Otavalo

Fuente - Elaboración propia

Dentro del campo la informática el proceso que corresponde al almacenamiento de datos se lo efectúa en repositorios más conocidos como base de datos, comúnmente son de tipo relacional y se las representa por medio de tablas compuestas de un conjunto de instancias.

Las bases de datos relacionales son las más eficientes para el almacenamiento de información de empresas e instituciones por la estructura normalizada que usan previniendo la redundancia o duplicación de esta información. La aplicación y ejecución de técnicas de minería de datos para el análisis de datos no necesariamente se necesita la estructura mencionada anteriormente, se puede emplear una estructura en formato desnormalizado con los procesos correspondiente para obtener la información necesaria.

### 1.3.1. Clasificación de los datos

La clasificación de los datos se genera acorde la naturaleza del atributo que se representa, clasificándolos en:

#### 1.3.1.1. *Cuantitativos*

Aquellos atributos cuyos valores representan magnitudes, siendo los siguientes:

- *Discretos*: pueden tomar un número limitado de valores diferentes.
- *Continuos*: cumplen que para cualquier par de valores siempre existirá un intermedio.

#### 1.3.1.2. *Cualitativos*

Aquellos que representan una categoría y no una cantidad, siendo:

- *Nominales*: existe una asignación un puramente arbitraria de números o símbolos.
- *Ordinales*: existe una relación de orden entre categorías.

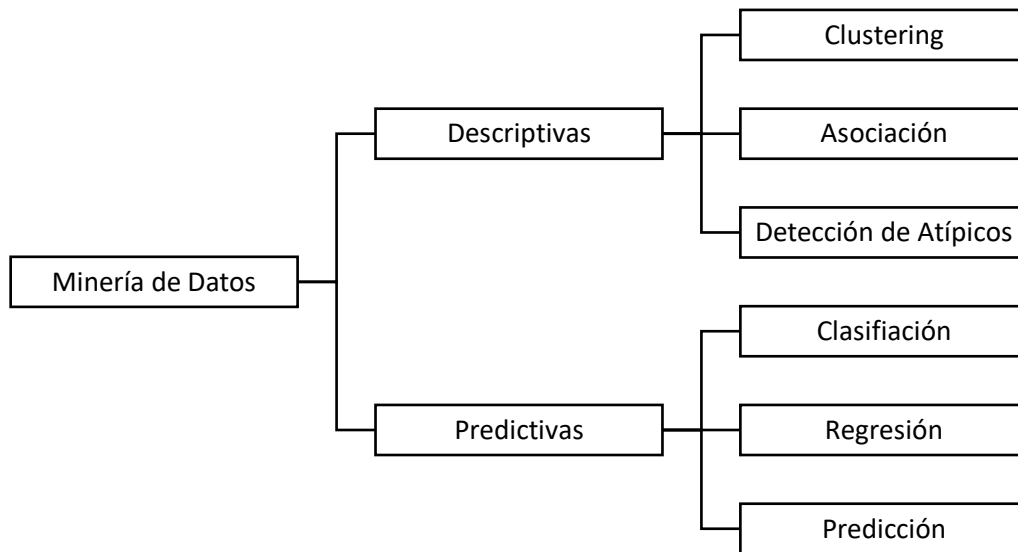
Los tipos de datos básicos no tan convencionales, pero también aplicables para la minería de datos, estos datos deben ser almacenados en repositorios especiales, y pueden ser:

- *Series temporales*: sucesión de valores que representan la evolución de una característica.
- *Datos espaciales*: representan una estructura espacial de un objeto.
- *Datos multimedia*: datos de tipo audio, imagen y video.
- *Documentos*: descripción textual de un objeto.
- *Datos procedentes de la web*: gran cantidad de datos en sitios web.

#### 1.4. Técnicas de Minería de datos

Las técnicas de minería de datos principalmente se dividen en 2 categorías: predictivas (aprendizaje supervisado) y descriptivas (aprendizaje no supervisados) (Hernández Valadez, 2006). En la figura 3 se muestra la taxonomía general de las categorías.

Figura 3 - Técnicas de minería de datos



Fuente: (Hernández Valadez, 2006)

##### 1.4.1. Técnicas Descriptivas

Cada una de las tareas de minería de datos pueden hacer referencia a distintos problemas que se pueden abordar de los cuales se destacan los siguientes:

#### **1.4.1.1. Clustering (Agrupamiento)**

Su objetivo es identificar y describir conjunto de datos, para que los elementos asignados a un mismo grupo sean similares entre sí, y los elementos de grupos distintos sean disimiles, para generar perfiles según los contextos de aplicación, permitiendo identificar grupos de datos organizados (Castillo-Rojas et al., 2017).

Según Lara (2014) la aplicación de la medida de similaridad se construye los clústers, dividiéndose en cuatro tipos:

- *Clustering particional*

Su objetivo es obtener una partición de los objetos el grupos o clústers de tal manera que todos los objetos pertenezcan a un grupo de los clústers posibles y por otra parte los clústers sean disjuntos (Larrañaga et al., 2012).

- *Clustering jerárquico*

Una jerarquía de divisiones del conjunto de elementos conglomerados se presenta mediante este procedimiento, con sus algoritmos más conocidos el aglomerativo y el disociativo (de la Fuente Fernández, 2001). Respectivamente, el primero parte de formar conglomerados e ir uniendo en las siguientes observaciones, mientras que el otro es en sentido inversa que va desde un gran conglomerado que en cada observación queda como uno distinto.

- *Clustering basado en densidad*

Algoritmos basados en la densidad de un punto, midiendo el número de puntos que son alcanzables desde él considerando un radio. DBSCAN es el algoritmo más conocido (Lara Torralbo, 2014).

- *Clustering basado en grid*

Algoritmos de clustering dedicados para datos espaciales específicamente, denominados como algoritmos basados en grid diseñados mismo que permiten cuantificar el espacio en un número finito de celdas y aplican operaciones sobre dicho espacio (Hernández Valadez, 2006).

Los principales algoritmos que se usan para el análisis de clúster son (Santamaria Ruiz, 2006):

- Algoritmo K-MEANS

Se basa en clústers donde principalmente se toma el centro, en términos de un centroide el cual en la mayoría de casos es la media de un conjunto de puntos y se aplica a objetos en espacios continuos n-dimensionales debiéndose especificar el número de clúster a encontrar.

- Algoritmo DBSCAN

Se toma en cuenta las regiones de alta densidad y se localizan grupos estratégicos para ser agrupados por regiones de baja densidad según los principios de clúster de densidad, obteniendo como resultado automático el número del clúster, además los puntos en baja densidad son únicamente ruido y son ignorando.

#### **1.4.1.2. Asociación**

Tiene como objetivo identificar relaciones no explícitas entre atributos categóricos y permite expresar patrones de datos de diferentes fuentes. Estos patrones sirven para conocer el comportamiento general del problema que genera la fuente de datos (Adriana et al., 2005). Existen algunos algoritmos para las reglas de asociación siendo los más utilizados:

- *Algoritmo A priori*

Realiza una búsqueda exhaustiva por niveles (menor a mayor tamaño de *itemsets*), permitiendo reducir el área de búsqueda, donde se debe tomar en cuenta la regla primordial que si un *itemset* no es frecuente, ninguno de sus *supersets* (el mayor tamaño que contengan al primero) puede ser frecuente (Amat Rodrigo, 2018). Este algoritmo está estructurado en 2 fases para su ejecución, siendo la primera la encargada de identificar los *itemsets* que se presentan con una frecuencia superior a la del límite establecido o también llamados *itemsets* frecuentes, la segunda corresponde a estos últimos *itemsets* a reglas de asociación.

- *Algoritmo FP-Growth*

Extrae reglas de asociación partiendo de los *itemsets*

Permite extraer reglas de asociación a partir de *itemsets* presentes con mayor número de repeticiones que se identifican sin necesidad de generar candidatos para cada

tamaño. El algoritmo emplea una estructura de árbol (*Frequent Pattern Tree*) donde almacena toda la información de las transacciones (Amat Rodrigo, 2018). Al permitir que la información se comprima se puede dividir en varias bases de datos con un patrón frecuente para cada una de ellas para concluir con el análisis individual para la concatenación de resultados.

#### **1.4.1.3. Detección de atípicos**

Busca encontrar dentro de un conjunto de objetos los que muestren características significativamente diferentes al resto. Se debe considerar el criterio para considerar este tipo de datos como atípicos (Flores Urgiles & Ortiz Amoroso, 2018), y las diferentes aproximaciones que permiten dar lugar a esta técnica de detección son :

- *Técnicas basadas en aproximaciones estadísticas*

Se usa modelos estadísticos creados a partir de datos para posteriormente evaluar la probabilidad de pertenencia, siendo los modelos más habituales lo de las distribuciones Normal, Poisson o Binomial.

- *Técnicas basadas en proximidad*

En esta técnica se considera como un valor atípico si dista demasiado de la mayoría del resto de objetos, la técnica se denomina k-vecinos más próximos.

- *Técnicas basadas en densidad*

Se calcula la densidad de regiones en los datos y declaran los objetos en regiones de baja densidad como atípicos. Se puede mencionar varios algoritmos de detección de valores atípicos como: resampling by half-means (RHM), smallest half volume (SHV), and ellipsoidal multivariate trimming (MVT) y LOF (Local Outlier Factor) (Chiang et al., 2003; Santamaria Ruiz, 2006).

- *Técnicas basadas en clustering*

Se basan en la suposición clave de la pertenencia de los objetos a grandes y densos clústers para considerar como atípicos a los muy pequeños. Los algoritmos empleados en esta técnica son algoritmo K-Means y algoritmo PAM (Partitioning Around Medoides), siendo el primero un método particional en el cual cada clúster está

asociado con un centroide (punto central) y en donde cada punto se asigna al clúster con el centroide más próximo, se debe especificar el número de clúster, K (Flores Urgiles & Ortiz Amoroso, 2018) permitiendo crear grupos de la partición a partir de un conjunto de datos. El segundo algoritmo se considera más confiable por su robustez y se basa en usar objetos centrales encontrados en un clúster también llamado *medoide* y no la media del clúster logrando así mayor eficiencia en sus resultados.

#### **1.4.2. Técnicas Predictivas**

Tienen como objetivo predecir un valor de un conjunto de datos a futuro en función de los datos conocidos, los mismos que son inducidos a una relación entre el dato y una serie de atributos, según Hernández Valadez (2006) las técnicas predictivas están clasificadas y consisten en:

##### **1.4.2.1. Clasificación:**

Esclarecer la dependencia de una clasificación en donde los atributos inmersos pueden tomar valores entre un grupo de clases ya conocidas.

##### **1.4.2.2. Regresión**

Su objetivo es la predicción de valores de una variable continua a partir de la evolución sobre otra variable continua que en la mayoría de los casos es la variable tiempo.

#### **1.4.3. Herramientas para minería de datos**

Existen un conjunto de herramientas para aplicar técnicas de minería de datos que han sido desarrolladas con el objetivo de aplicar las diferentes técnicas de minería de datos existentes, mismas que permiten gestionar los datos para identificar los posibles patrones (conocimiento). Algunas de estas herramientas son:

##### **1.4.3.1. RapidMiner**

Es una herramienta de acceso gratuito hasta 10.000 filas de datos y un fácil manejo al no necesitar un conocimiento muy elaborado en programación (RapidMiner Inc., 2020), es una plataforma basada en la colaboración de extremo a extremo y de código abierto y extensible,

permitiendo una profundidad para las necesidades del trabajo de los científicos de datos, simplificada para todos.

#### **1.4.3.2. Orange**

Este software permite realizar visualizaciones de datos para descubrir patrones de datos ocultos. Proporciona un plataforma limpia y de código abierto y la posibilidad de agregar más funcionalidades para todos los campos de la ciencia (Borondics, 2020). El objetivo de esta herramienta es proveer de una fácil construcción de flujos de trabajos y visualizaciones para el análisis de gran cantidad de información.

#### **1.4.3.3. Knime**

Es una plataforma de software de código abierto de nivel empresarial que consta con de una gran cantidad de funciones, módulos y paquetes de aplicaciones en su plataforma de análisis KNIME y el servidor KNIME que permiten descubrir estructuras de datos ocultas (Knime, 2019). El análisis de datos integrativos y la integración de numerosos procedimientos de aprendizaje automático son los que destacan en esta herramienta.

#### **1.4.3.4. SAS**

Provee de una visualización interactiva de los datos mediante el pronóstico y evaluación de los datos, posee todos los elementos necesarios para cumplir su propósito de detección de conocimiento oculto, mediante sus herramientas para procesamiento de datos, análisis estadístico, previsión, análisis de texto, optimización y simulación. (SAS Institute Inc., 2020). No obstante, a todos sus beneficios es una herramienta que únicamente funciona con la compra de una licencia para su acceso.

#### **1.4.3.5. Weka**

Es un software desarrollado por la Universidad de Waikato, de código abierto mediante el acceso a una interfaz gráfica brinda acceso a una gran variedad de algoritmos de aprendizaje automático para tareas de minería de datos (Lara Gutiérrez et al., 2016).

Su compatibilidad es para los sistemas operativo Windows, macOS y Linux, además con la posibilidad de conexión a bases de datos SQL. Las funciones que presenta esta herramienta

permiten la paliación de las técnicas para clústeres, de correlación o de regresión, la clasificación de datos, siendo este último un pilar de la minería de datos, al usar redes de neuronas artificiales, árboles de decisión y algoritmos ID3 o C4.5. Su gran acogida es por su gran utilidad en el ámbito de la enseñanza y la investigación (Machine Learning Group, 2011).

#### **1.4.4. Herramientas de análisis empresarial**

##### **1.4.4.1. *Pentaho Data Integration***

Pentaho Data Integration (PDI) es una herramienta de ETL (Extracción, Transformación, Carga) de Pentaho, permite realizar la integración de datos mediante una interfaz gráfica simplificada mediante su opción de arrastrar y soltar, permite la conexión a cualquier fuente de datos además permite una gran escalabilidad de los datos. Permite operacionalizar los modelos R, Python, Scala y Weka (Ventara-Hitachi, 2019).

##### **1.4.4.2. *Power BI***

Es un conjunto de servicios de software que trabajan en conjunto para la conversión de orígenes de datos sin relación específica en información organizada, coherente y comprensible visualmente, provee herramientas para análisis de información empresarial para la toma de decisiones (Microsoft, 2020). Posee tres partes para explotar sus mayores ventajas como son Power BI Desktop, servicio Power BI y Power BI Mobile.

#### **1.5. Proceso KDD (Knowledge Discover in Databases)**

El término KDD traducido al español significa *Descubrimiento de Conocimiento en Bases de Datos* fue nombrado por Gregory Piatetsky-Shapiro al realizar el primer taller de sobre el tema, partiendo desde ahí el uso del término en el ámbito científico y educativo (Camana, 2016).

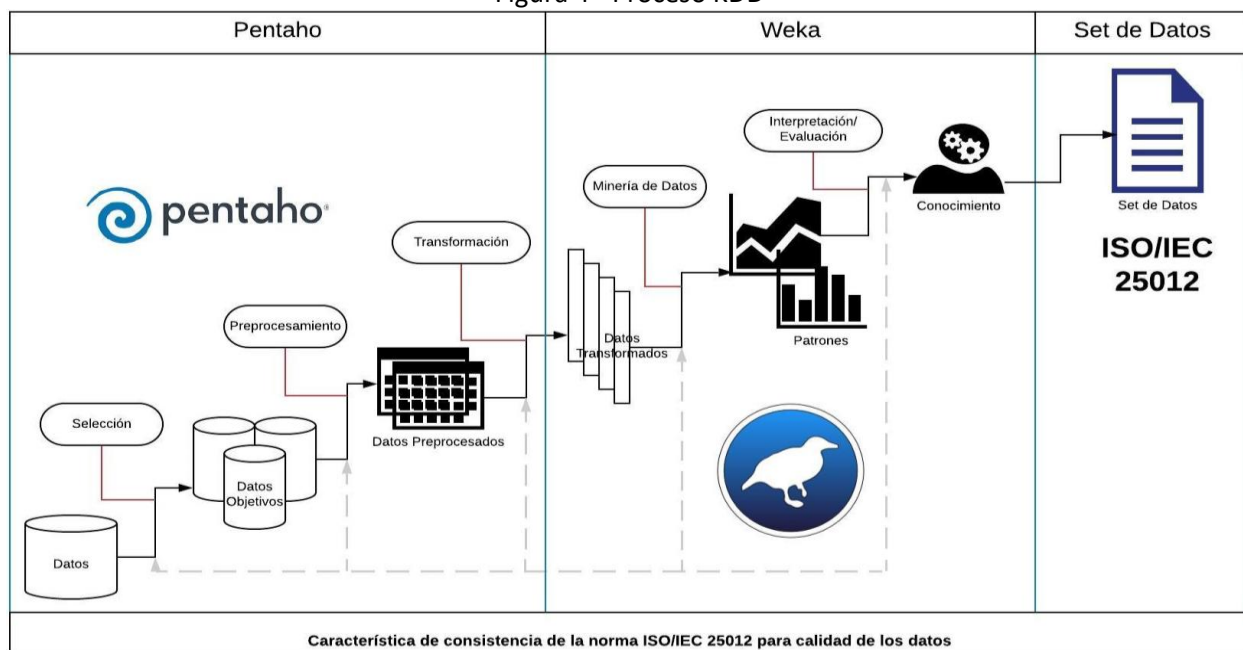
El descubrimiento de conocimiento en bases de datos pertenece al campo de la inteligencia artificial mismo que presenta rápido crecimiento, combina técnicas del aprendizaje de máquina, reconocimiento de patrones, estadística, bases de datos y visualización para una extracción automática de conocimiento (o información), de un nivel bajo de datos (bases de datos) (Nigro et al., 2003).



### 1.5.1. Fases del proceso KDD

El proceso KDD inicia con la entrada de los datos que se desea analizar, seguido de ello los datos son preparados siendo esto un paso fundamental para lograr éxito en el descubrimiento del conocimiento. Después de las transformaciones necesarias sigue la denominada fase de minería de datos en donde aplican un conjunto de técnicas para obtener los modelos. Según Lara (2014) un modelo es: “Una representación simbólica de la realidad que representan los datos de entrada” (p.15). Finalmente, los modelos obtenidos pasan a la siguiente fase donde se evalúa e interpreta los resultados y llegar al conocimiento (patrones) de lo oculto en la información.

Figura 4 - Proceso KDD



Fuente: Adaptación de Timarán-Pereira, Hernández-Arteaga, Caicedo-Zambrano, Hidalgo-Troya, & Alvarado-Pérez, 2016

- **Selección**

Esta es la fase inicial donde comienza la recopilación e integración de la información donde se obtiene datos relevantes de diferentes fuentes para ser utilizados en la fase de procesamiento (Camana, 2016).

- **Procesamiento**

Los datos serán procesados con la eliminación de ruido, estrategias de valores ausentes, normalización de datos, entre otros (Beltrán Martínez, 2001), para obtener un conjunto limpio para la seguir con la siguiente fase.

- **Transformación**

Consiste en la transformación de los datos, creando la posibilidad de tener nuevas variables desde las existentes previamente (Camana, 2016); además, dentro de este proceso se debe considerar la métodos de *normalización y discretización* de datos para su posterior análisis que consiste en la simplificación de los datos dentro de los campos de registros para estructurarlos de una forma precisa , entendible y manejable para representar las relaciones necesarias (Montenegro Cochás, 2005), algunos métodos de normalización según Gironés, Casas, & Minguillón et al., (2017) son:

- Normalización por el máximo

Encuentra el valor máximo del atributo a normalizar y divide el resto de valores por este valor máximo, asegurando los valores en el rango de [0,1] su fórmula es la siguiente:

$$z_i = \frac{x_i}{x_{max}} \quad \text{Ec. 1}$$

Tabla 2 - Simbología Ecuación 1

<b>Simbología</b>	<b>Descripción</b>
$Z_i$	Valor normalizado
$x_i$	Valor original del atributo
$x_{max}$	Valor máximo para el ese atributo de todo el conjunto de datos

Fuente: (Gironés et al., 2017)

- Normalización por la diferencia

Compensa el efecto de la distancia del valor tratado acorde al máximo de los valores observados, su fórmula es la siguiente:

$$z_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad \text{Ec. 2}$$

Tabla 3 - Simbología Ecuación 2

<b>Simbología</b>	<b>Descripción</b>
$Z_i$	Valor normalizado
$x_i$	Valor original del atributo
$x_{max}$	Valor máximo para el ese atributo de todo el conjunto de datos

$x_{min}$	Valor mínimo para ese atributo en el conjunto de datos
-----------	--

Fuente: (Gironés et al., 2017)

- Normalización basada en la desviación estándar (estandarización de valores)  
Asegura la obtención de valores dentro de un rango determinado que tiene como propiedad valor 0 para su media y 1 para su desviación estándar.

$$z_i = \frac{x_i - \mu}{\sigma} \quad \text{Ec. 3}$$

Tabla 4 - Simbología Ecuación 3

Simbología	Descripción
$z_i$	Valor normalizado
$x_i$	Valor original del atributo
$\mu$	Media
$\sigma$	Varianza

Fuente: (Gironés et al., 2017)

#### ▪ Minería de Datos

Por medio de esta fase se encuentran los patrones según el propósito de la exploración de la información. Los algoritmos de minería de datos contemplan 3 elementos: modelo, criterio de preferencia y algoritmo de búsqueda (Rodríguez Miranda, 2018).

#### ▪ Evaluación e interpretación

Se evalúan los patrones obtenidos y se analizan por los expertos en el dominio de aplicación. En caso de ser necesario, se regresa a las fases anteriores para generar una nueva iteración (Perez Lopez & Gonzáles Satin, 2007).

### 1.6. ISO/IEC 25012:2008

Presenta un modelo de Calidad de Datos, como la base fundamental para la construcción de un sistema para la evaluación de un producto de datos. Se establecen las características que se deben tomar en cuenta al momento de evaluar las propiedades de un producto de este tipo (ISO 25000, 2018). Las características de calidad de datos están clasificadas en las siguientes:

#### 1.6.1. Calidad de Datos Inherente:

Hace referencia al grado de las características de calidad que se ven reflejadas en el potencial intrínseco para satisfacer determinadas necesidades bajo un contexto específico, haciendo

énfasis en: valores de dominios de datos y restricciones, relaciones entre valores de datos y metadatos.

Dentro de la calidad de datos inherentes se debe considerar (Instituto Ecuatoriano de Normalización, 2014):

#### **1.6.1.1. Exactitud**

Grado en el que los datos representan correctamente el verdadero valor del atributo deseado de un concepto un contexto de uso específico con aspectos de exactitud sintáctica y exactitud semántica, correspondientes a la cercanía de los valores de los datos a un conjunto de valores considerados correctos dentro del dominio sintáctico o semántico.

#### **1.6.1.2. Completitud**

Consiste en que todos los datos asociados a una entidad tengan sus valores completos para todos sus atributos de entidades relacionadas dentro de un contexto determinado.

#### **1.6.1.3. Consistencia**

Verifica que los datos estén sin contradicción y sean coherentes dentro de un determinado contexto consiguiendo ser analizados en una o varias entidades comparables.

#### **1.6.1.4. Credibilidad**

Hace referencia a que los datos de las atribuciones se los consideran verdaderos y creíbles dentro de determinado contexto incluyendo su autenticidad en la veracidad de los orígenes de datos y sus atribuciones.

#### **1.6.1.5. Actualidad**

Consiste en que los datos de los atributos tienen el tiempo y la época correcta dentro de un contexto determinado.

### **1.6.2. Calidad de Datos Dependiente del Sistema:**

En este caso hace referencia al grado con el que la Calidad de Datos se consigue y mantiene en un sistema informático con datos sometidos a determinadas condiciones y dependiendo del

dominio tecnológico en donde los datos serán usados. Se logra gracias a las capacidades de los componentes del sistema informático tales como: dispositivos hardware y otro software. Considerando las siguientes características (Insituto Ecuatoriano de Normalización, 2014):

**1.6.2.1. Accesibilidad**

Capacidad en el que los datos pueden ser accedidos en un determinado contexto.

**1.6.2.2. Conformidad**

Solvencia de los datos para adherirse a estándares, convenciones o normativas y reglas similares con respecto a la calidad de datos en un determinado contexto.

**1.6.2.3. Confidencialidad**

Permite que los datos tengan atributos que aseguran el acceso e interpretación únicamente por aquellos usuarios acreditados dentro de un determinado contexto.

**1.6.2.4. Eficiencia**

Hace referencia a como los datos pueden ser procesados y proporcionados con niveles de alto rendimiento usando las cantidades y tipos adecuados de recursos dentro de un contexto determinado.

**1.6.2.5. Precisión**

Como los datos tienen atributos exactos y proporcionan acierto dentro de un contexto determinado.

**1.6.2.6. Trazabilidad**

Solvencia de los datos para proporcionar acceso auditado a los datos o cualquier otro cambio realizado sobre los datos dentro de un contexto determinado.

### **1.6.2.7. *Comprensibilidad***

Capacidad de los datos para ser leídos y comprendidos por los usuarios, son expresados utilizando lenguajes, símbolos y unidades apropiados dentro de un contexto determinado.

### **1.7. Trabajos existentes**

Algunos trabajos relevantes para el proyecto se enumeran a continuación por la aplicación a delitos y otras, usando técnicas minería de datos descriptivas.

Camaná & Torres, (2017) usaron una base de datos de la información de estudiantes de la Carrera de Tecnología en Análisis de Sistemas (CTAS) para el descubrimiento del estilo de aprendizaje de los evaluados, utilizaron una plantilla de Excel con datos extraídos de una encuesta posteriormente convertido a un archivo tipo .arff para la aplicación de clustering con su algoritmo FarthestFirs en la herramienta Weka, obteniendo como resultados recomendaciones necesarias para diferentes dimensiones que corresponden a un grupo de asignaturas correspondientes a las materias de la carrera CTAS, finalmente proponen parámetros con los cuales se obtiene un grupo de lineamientos con el propósito que el estilo de aprendizaje de los estudiantes responda al estilo de enseñanza del docente para un correcto encaje y resultados del proceso educativo.

Orellana & Cedillo (2020) dividieron en 3 fases la detección de valores atípicos en el análisis de datos financieros de una entidad bancaria en el cual consideró el uso de una experiencia con técnicas estadísticas y técnicas de minería de datos. Utilizaron una base de datos de tipo nominal con 15 columnas de una base de datos de crédito, aplicando el método estadístico chi-cuadrado para definir la relevancia de la columna del perfil del cliente en base a la estimación de un peso denominado W para generar la comparación con de cantidad de crédito otorgado. Además, se aplicó el algoritmo de agrupación KNN Global Anomaly (KNN) mismo que permitió identificar la puntuación de los valores atípicos y concluir con el uso del algoritmo K-Means para disgregación correcta de los valores atípicos de los valores falsos-positivos. Obteniendo como resultado un total de 48 tuplas en 11 columnas con valores anómalos y tras aplicar el algoritmo establecieron que el 80 % de las columnas contienen valores anómalos, y un restante 20 % resultaron una mezcla entre valores anómalos y falsos-positivos.

Aguirre Yacup & Walteros Alcázar (2018) identificó patrones delictivos en Colombia mediante el uso del algoritmo de agrupamiento o clustering k-modes de Zhexue Huang mismo que permite utilizar datos cualitativos, reemplazó las medias por modos y usa un método basado en frecuencia con el objetivo principal la sumatoria de las distancias de los objetos a los modos, se usó un data set con 21 atributos y 402.631 registros , obteniendo como resultado la identificación de cuáles son los barrios de Popayán con hurtos de diferentes clases en el periodo 2010-2016, organizando toda la información resultante en modelos de 3 y 5 clústers.

Suárez Rodríguez (2015) usó algoritmos de agrupamiento de minería de datos espaciales implementando en el proyecto 2 algoritmos el K-means y de DBSCAN (Density-based spatial clustering of applications with noise) en bases de datos alfanuméricas de los hurtos a personas en Bogotá, obteniendo como resultado los focos de la localidad con mayores actividades delincuenciales , además aporó con la distancia para más corta a estaciones policías en cada punto de su respectivo clúster.

Ramírez-Murillo, Torres-Pinzón, & Forero-García, (2019) encontraron el potencial fotovoltaico por encima de los márgenes recomendados en base a los datos obtenidos de cuatro ciudades de Colombia (Bogotá, Cúcuta, Manizales y Pasto), usaron archivos de más de 270.000 datos por cada estación meteorológica, realizando una limpieza de datos minuciosa en Excel para poder realizar en el procesamiento de la información en la herramienta Matlab, aplicando el método de clustering con los algoritmos K-means y Fuzzy C-means.

Velandia Rodríguez, Guataquira, & Prada, (2018) usaron el modelo de clustering y el algoritmo k-means para realizar minería de datos apoyados con la herramienta Weka para determinar cómo se encuentra la percepción de seguridad en Bogotá con la aplicación de un serie de encuestas en la plataforma Google Drive con el fin de obtener la percepción del ciudadano común, misma que fue aceptada por 49 personas de diferentes localidades, obteniendo como información relevante que para la ciudadanía no es suficiente la seguridad que se está brindando por parte del Estado Colombiano y además existe una decadencia en lugar de una mejoría en el sistema de seguridad.

R. Timarán-Pereira, Hernández-Garzón, & Quemá-Taimbud, (2017) seleccionaron la tarea de aprendizaje no supervisado de agrupamiento o clustering para la identificación de lesiones no fatales dentro del municipio de Pasto logrando determinar que el 79,7% de lesiones son de tipo no intencional o accidental y el porcentaje restante corresponde a violencia intrafamiliar, esto fue obtenido con el uso de la herramienta Weka aplicando el algoritmo K-means.

Arce, Lima, & Orellana et al., (2018), basados en la metodología CRISP-DM experimentaron sobre la información recopilada en la estación de monitoreo y contaminantes del aire en Cuenca, Ecuador con el fin de descubrir los patrones de comportamiento entre los contaminantes del aire, usando un conjunto de datos correspondientes a 25.877 después de los respectivos pasos de limpieza para la aplicación de algoritmos como X-means y K-means, obteniendo como resultado que el contaminante más relevante en la ciudad es el O<sub>3</sub> (ozono) y las correlaciones con otros contaminantes evaluados, reflejando resultados positivos en el tiempo de procesamiento adecuado para tareas de agrupación.



## CAPÍTULO 2

### Proceso de descubrimiento del Conocimiento

#### 2.1. Visión general del proyecto

##### ▪ Suposiciones

El desarrollo del proyecto de inicia con la recopilación de la información por parte de la institución de control (anónima) para luego proceder a la entrega de esta al jefe del proyecto y al analista de sistemas, para proceder con el tratamiento de la información y la posterior implementación de técnicas de minería de datos con el objetivo principal de generar un análisis en base a la experticia de los encargados de todo el proceso de control para determinar las acciones correspondientes para mitigar la problemática antes descrita.

##### ▪ Restricciones

El desarrollo de proyecto se encuentra estimado en un tiempo de 6 meses organizados acorde al cronograma definido mismo que consta de un grupo de actividades secuenciales, tomando en cuenta la disponibilidad de los expertos y primordialmente de la disponibilidad de la información al ser de tipo confidencial. La ejecución de cada una de las fases debe contar el cumplimiento de permisos de acceso a la información y el acuerdo de responsabilidad sobre la misma.

#### 2.2. Entregables del proyecto

Los entregables del proyecto se mencionan en la Tabla 5 a continuación, tomando en cuenta que mientras duran el proceso KDD estos pueden ir cambiando.

Tabla 5 - Entregables del Proyecto

ENTREGABLE	DETALLE
Vista Minable	Fase de recopilación de datos
Modelos descriptivos (asociación, agrupación y atípicos)	Fase de minería de datos
Conocimiento	Fase de interpretación y evaluación
Set de datos	Entregable
Interpretación de la información	Visualización de resultados

Fuente: Elaboración Propia

## 2.3. Organización del proyecto

### 2.3.1. Participantes del Proyecto

En la tabla a continuación se detalla los directores de las áreas:

Tabla 6 - Directores de las Áreas Comprendidas

Dependencia	Participante	Función
Coordinador Carrera de Ingeniería en Sistemas Computacionales	Msc. Pedro Granda	Asignar especialista en minería de datos
Entidad de Control	Anónimo	Especialista en la problemática

Fuente: Elaboración Propia

En la Tabla 7 a continuación se detalla los participantes directos del proyecto:

Tabla 7 - Participantes Directos del Proyecto

Rol	Dependencia	Nombre
Jefe del proyecto	Docente Carrera de Ingeniería en Sistemas Computacionales	PhD. Iván García S.
Especialista del negocio	Entidad de Control	Anónimo
Analista de Sistemas	Carrera de Ingeniería en Sistemas Computacionales	Diana Rosero Rea

Fuente: Elaboración Propia

### 2.3.2. Roles y Responsabilidades

En la Tabla 8 a continuación se describen cada uno de los roles con sus respectivas funciones y responsabilidades.

Tabla 8 – Roles y responsabilidades del proyecto

Rol	Responsabilidad
<b>Jefe de Proyecto</b>	Encargado de planificar, designar, ejecutar y llevar control de las actividades con el propósito de integrar dinámicamente el cumplimiento de los objetivos planteados (Porrás & Conde, 2019).
<b>Especialista del negocio</b>	Encargado de proporcionar el conocimiento requerido al analista de sistemas la comprensión del contexto en el cual se obtiene la información que provee.
<b>Analista de Sistemas</b>	Encargado del análisis de la información proporcionada por el especialista del negocio, validar

Fuente: Elaboración Propia

## 2.4. Gestión del proyecto

### 2.4.1. Estimaciones

En las tablas a continuación se detalla los valores y presupuestos de talento humano y los recursos materiales para el desarrollo del proyecto, en la Tabla 9 se puede apreciar la estimación del costo del talento humano para el desarrollo del proyecto.

Tabla 9 - Talento Humano

DESCRIPCIÓN	N. DE HORAS	COSTO POR HORA (\$)	COSTO TOTAL (\$)
Horas de investigación del proyecto	200	20,00	4000,00
Horas de desarrollo del proyecto	220	20,00	4400,00
		<b>TOTAL</b>	<b>8400,00</b>

Fuente: Elaboración Propia

El número total de horas empleadas en el desarrollo del proyecto corresponde a un trabajo diario de 2 horas, durante 22 semanas de los meses de febrero, marzo, junio, julio, agosto y septiembre, tiempo en el cual se debió coordinar cuestiones logísticas sobre el acceso a la información y el acuerdo entre las partes sobre el uso e interpretación correcta de la información.

En caso de las estimaciones de los recursos materiales necesarios se consideró aspectos de hardware, software, material de oficina e investigación como se aprecia en la Tabla 9.

Tabla 10 - Recursos Materiales

DESCRIPCIÓN	COSTO ESTIMADO	COSTO REAL
<b>HARDWARE</b>		
Computadora Portátil	900,00	0,00
Impresora	180,00	0,00
<b>SOFTWARE</b>		
Microsoft Excel	0,00	0,00
Microsoft Word	0,00	0,00
Pentaho 7	0,00	0,00
Weka	0,00	0,00
<b>MATERIALES DE OFICINA</b>		
Tinta para impresora	48,00	40,00

Hojas A4	5,00	5,00
Internet	200,00	200,00
Esferos	2,00	2,00
Flash Memory	12,00	12,00
<b>INVESTIGACIÓN</b>		
Textos, fotocopias	25,00	25,00
ISO/IEC 25012	88,00	0,00
<b>TOTAL</b>	<b>1460,00</b>	<b>284,00</b>

Fuente: Elaboración Propia

Finalmente, en la Tabla 11 se aprecia la estimación total del costo de proyecto.

Tabla 11 - Costo total del proyecto

DESCRIPCIÓN	COSTO
Talento Humano	8000,00
Recursos Materiales	1460,00
<b>Total</b>	<b>9460,00</b>

Fuente: Elaboración Propia

## 2.4.2. Plan del proyecto

El presente proyecto será desarrollado en fases con el fin de identificar cada una de ellas para el cumplimiento de los objetivos propuestos, la obtención de resultados que contribuirá al trabajo planteado para la mitigación del delito de contrabando, en la Tabla 12 a continuación se aprecia cada una de las fases con el tiempo de duración en horas de cada una de ellas.

Tabla 12 - Distribución de horas

FASE	DURACIÓN EN HORAS
Fase de recopilación de datos	30
Fase de selección, limpieza y transformación de datos	30
Implementación de la norma ISO/IEC 25012	10
Fase de minería de datos	50
Fase de evaluación e interpretación	170
Documentación	40
Análisis de Resultados	40
Análisis de Impactos	40

BI	10
<b>TOTAL</b>	<b>420</b>

Fuente: Elaboración Propia

## 2.5. Recopilación de datos

### 2.5.1. Tipos de bases de datos

EL desarrollo del proceso KDD, se emplearán datos descriptivos, económicos, de ubicación que se han almacenado en un documento de Microsoft Excel administrado por la institución anónima. Los tipos de datos encontrados dentro de los registros son: fechas, enteros, decimales y cadenas de texto.

En la Tabla 13 a continuación se detallan los tipos de datos que presentan la base de datos acorde a cada uno de los archivos categorizados por año que corresponden en conjunto al periodo 2014-2019.

Tabla 13 – Atributos y tipos de datos de la información

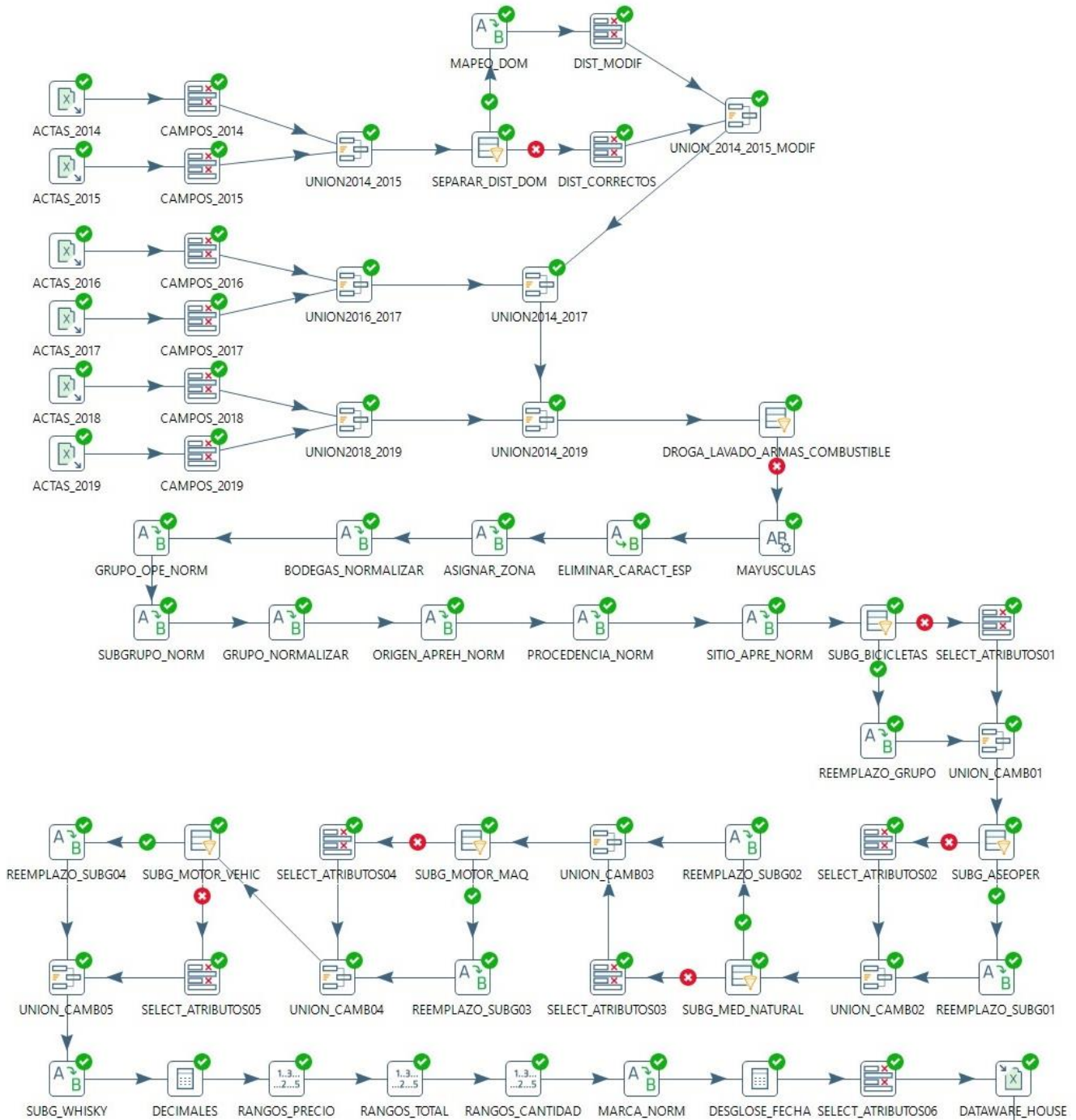
Nº	Atributo	Tipo de dato	ACTAS DE APREHENSIÓN					
			2014	2015	2016	2017	2018	2019
1	Fecha	Fecha	x	x	x	x	x	x
2	Cantidad	Cadena de texto	x	x	x	x	x	x
3	Unidades	Real	x	x	x	x	x	x
4	Precio	Real	x	x	x	x	x	x
5	Total	Cadena de texto	x	x	x	x	x	x
6	Grupo	Cadena de texto	x	x	x	x	x	x
7	SubGrupo	Cadena de texto	x	x	x	x	x	x
8	Descripción	Cadena de texto	x	x	x	x	x	x
9	Procedencia	Cadena de texto	x	x	x	x	x	x
10	Bodega	Cadena de texto	x	x	x	x	x	x
11	Marca	Cadena de texto	x	x	x	x	x	x
12	Status de la mercancía	Cadena de texto	x	x	x	x	x	x
13	Lo que originó la aprehensión	Cadena de texto	x	x	x	x	x	x
14	Sitio de la aprehensión	Cadena de texto	x	x	x	x	x	x
15	Grupo Operativo	Cadena de texto	x	x	x	x	x	x
16	Distrito	Cadena de texto	x	x	x	x	x	x
17	Zona	Cadena de texto						x

Fuente: Elaboración Propia

## 2.6. Fase de selección, limpieza y transformación

En la Fig. 5 se muestra el proceso ETL realizado con la herramienta Pentaho para la realización del proceso de selección, transformación y limpieza de la información.

Figura 5 - ETL realizado en Pentaho

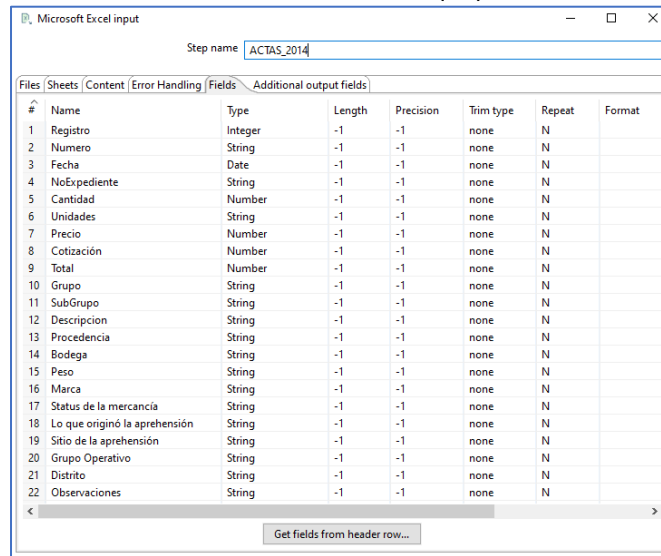


## 2.6.1. Selección

### ▪ Extracción de atributos

Al proceder con la etapa de selección y limpieza de la información entregada por la entidad de control se inició seleccionando los campos útiles para el análisis descartando aquellos que no presentan información relevante, obteniendo así **16 campos relevantes** de los 22 iniciales para el periodo 2014-2019 como se muestra en la Fig. 6.

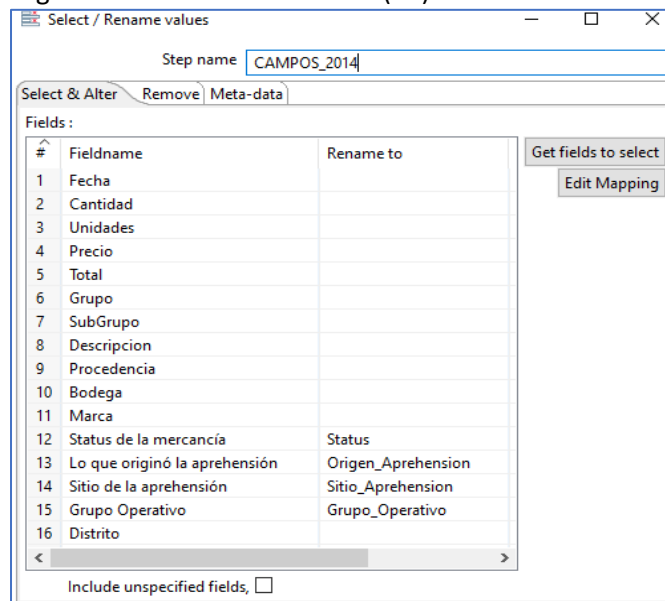
Figura 6 - Atributos base de datos inicial (22) años 2014-2018



The screenshot shows a window titled "Microsoft Excel input" with a step name "ACTAS\_2014". The "Fields" tab is active, displaying a table of 22 attributes. The table has columns for #, Name, Type, Length, Precision, Trim type, Repeat, and Format.

#	Name	Type	Length	Precision	Trim type	Repeat	Format
1	Registro	Integer	-1	-1	none	N	
2	Numero	String	-1	-1	none	N	
3	Fecha	Date	-1	-1	none	N	
4	NoExpediente	String	-1	-1	none	N	
5	Cantidad	Number	-1	-1	none	N	
6	Unidades	String	-1	-1	none	N	
7	Precio	Number	-1	-1	none	N	
8	Cotización	Number	-1	-1	none	N	
9	Total	Number	-1	-1	none	N	
10	Grupo	String	-1	-1	none	N	
11	SubGrupo	String	-1	-1	none	N	
12	Descripcion	String	-1	-1	none	N	
13	Procedencia	String	-1	-1	none	N	
14	Bodega	String	-1	-1	none	N	
15	Peso	String	-1	-1	none	N	
16	Marca	String	-1	-1	none	N	
17	Status de la mercancía	String	-1	-1	none	N	
18	Lo que originó la aprehensión	String	-1	-1	none	N	
19	Sitio de la aprehensión	String	-1	-1	none	N	
20	Grupo Operativo	String	-1	-1	none	N	
21	Distrito	String	-1	-1	none	N	
22	Observaciones	String	-1	-1	none	N	

Figura 7 - Atributos relevantes (16) años 2014-2018



The screenshot shows a window titled "Select / Rename values" with a step name "CAMPOS\_2014". The "Select & Alter" tab is active, displaying a table of 16 selected attributes. The table has columns for #, Fieldname, and Rename to. There are buttons for "Get fields to select" and "Edit Mapping".

#	Fieldname	Rename to
1	Fecha	
2	Cantidad	
3	Unidades	
4	Precio	
5	Total	
6	Grupo	
7	SubGrupo	
8	Descripcion	
9	Procedencia	
10	Bodega	
11	Marca	
12	Status de la mercancía	Status
13	Lo que originó la aprehensión	Origen_Aprehension
14	Sitio de la aprehensión	Sitio_Aprehension
15	Grupo Operativo	Grupo_Operativo
16	Distrito	

En el caso de la información correspondiente al año 2019 los campos son los mismos de los años anteriores, pero se presentó adicionalmente el campo Zonas siendo relevante.

Figura 8 – Atributos (22) base de datos inicial año 2019

The screenshot shows a window titled 'Microsoft Excel input' with a 'Step name' field containing 'ACTAS\_2019'. Below this is a tabbed interface with 'Fields' selected. A table lists 22 fields with their respective types, lengths, and other attributes.

#	Name	Type	Length	Precision	Trim type	Repeat
1	Registro	Integer	-1	-1	none	N
2	Numero	String	-1	-1	none	N
3	Fecha	Date	-1	-1	none	N
4	NoExpediente	String	-1	-1	none	N
5	Cantidad	Number	-1	-1	none	N
6	Unidades	String	-1	-1	none	N
7	Precio	Number	-1	-1	none	N
8	Cotización	Number	-1	-1	none	N
9	Total	Number	-1	-1	none	N
10	Grupo	String	-1	-1	none	N
11	SubGrupo	String	-1	-1	none	N
12	Descripcion	String	-1	-1	none	N
13	Procedencia	String	-1	-1	none	N
14	Bodega	String	-1	-1	none	N
15	Peso	String	-1	-1	none	N
16	Marca	String	-1	-1	none	N
17	Status de la mercancía	String	-1	-1	none	N
18	Lo que originó la aprehensión	String	-1	-1	none	N
19	Sitio de la aprehensión	String	-1	-1	none	N
20	Grupo Operativo	String	-1	-1	none	N
21	Distrito	String	-1	-1	none	N
22	ZONAS	String	-1	-1	none	N

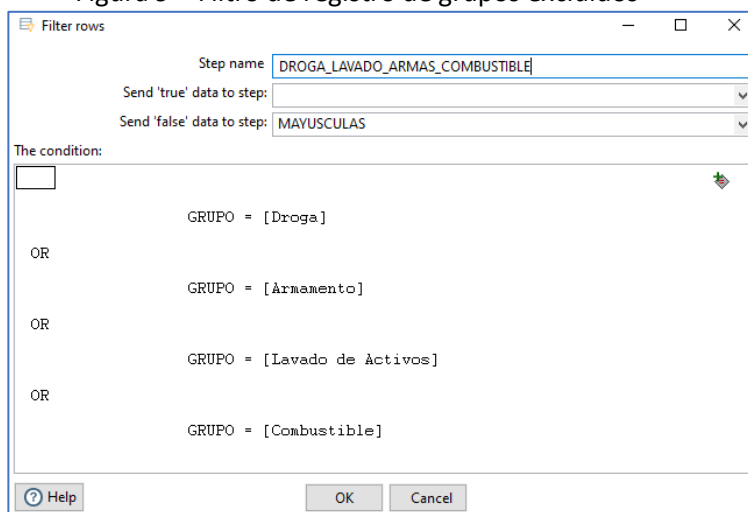
At the bottom of the table is a scroll bar and a button labeled 'Get fields from header row...'.

Obteniendo de esta forma un total de 103135 instancias correspondientes a los registros **correspondientes a los 17 campos relevantes** de la base de datos del periodo correspondiente a 2014-2019 almacenada en la herramienta Excel entregados por parte de entidad de control.

En base los diálogos mantenidos con el experto de la entidad de control se acordó excluir los registros correspondientes a los grupos de *Armamento, Droga, Combustible y Lavado de Activos* por motivos de vulnerabilidad y riesgo de uso de la información catalogada como confidencial acorde a la carta de confidencialidad firmada por las partes (Anexo 1), para ello se usó la opción de *Filter rows* como se muestra en la Fig. 9.



Figura 9 – Filtro de registro de grupos excluidos



Después de haber realizado el filtrado se obtuvo un conjunto total de datos correspondiente a 102667 instancias para poder continuar con los pasos correspondientes al preprocesamiento de la información

### 2.6.2. Transformación

Para realizar la transformación se categorizó cada uno de los campos donde existían criterios específicos previos, después de la respectiva explicación del experto en tema obteniendo así las siguientes categorías para ciertos atributos relevantes.

*Nota:* Las tablas de información de los tributos correspondientes al proceso de transformación de la información no muestran la totalidad de opciones presentes por motivos de confidencialidad de la información para lo cual se optó por tomar una pequeña muestra simbólica para fines de redacción y presentación del trabajo realizado.

- **Atributo Unidades**

Para el caso del atributo unidades de normalizo los datos presentes en la información correspondiente al período 2014-2019, tomando en cuenta que las opciones resultantes no contengan repeticiones obteniendo de esa manera una lista de opciones de la cual se una muestra simbólica se encuentra en la Tabla 14 que se encuentra a continuación:

Tabla 14 - Categorías para el atributo UNIDADES

UNIDADES	
DESCONOCIDA	BANDEJA
AMPOLLA	BIDÓN
ARROBA	BILLETE

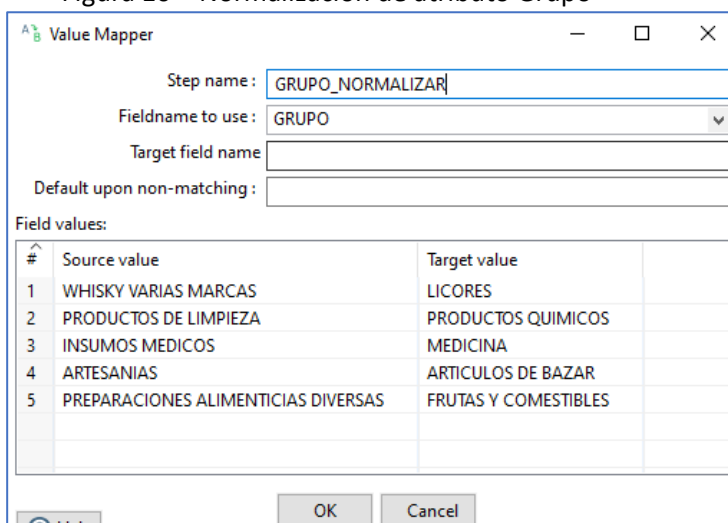
ATADO                      BOLSA  
BALDE                      BOTELLA

Fuente: Elaboración Propia

▪ **Atributo Grupo**

Para el atributo Grupo se normalizó las opciones presentadas en la tabla, acorde a la información del período 2014-2019, tomando en cuenta la eliminación de datos repetidos. Además, se tomó en cuenta la existencia de grupos que eran categorizados como subgrupos de igual forma se realizó la transformación correspondiente para evitar errores en la información. Además, acorde a la información de la base de datos se agrupó la información de los grupos según el contexto de cada uno de ellos para lo cual se usó *Value Mapper* para eliminar grupos que se contenían en otros como se muestra en la Fig. 10 a continuación:

Figura 10 – Normalización de atributo Grupo



Para finalmente obtener un total de 27 datos para el atributo GRUPO, algunos de ellos se muestran en la Tabla 15 presentada a continuación:

Tabla 15 - Categorías para el atributo GRUPO

<b>GRUPO</b>	
ANIMALES VIVOS Y SUS PARTES	CONFITERIA
ARTEFACTOS ELECTRONICOS	COSMÉTICOS
ARTICULOS DE BAZAR	COURIER
CALZADO	ELECTRÓNICA Y SUS ACCESORIOS
CIGARRILLOS	EMBARCACIONES

Fuente: Elaboración Propia

- **Atributo Subgrupo**

Para el atributo subgrupo se identificó los subgrupos que se encontraban repetidos por errores de digitación y hacía referencia a un solo subgrupo y se organizó de tal manera que los elementos de subgrupo tengan relación con el grupo como se puede apreciar en la Tabla 16.

Tabla 16 - Categorización del atributo SUBGRUPO (muestra del total)

<b>SUBGRUPO</b>	
AVES EN GENERAL	ARTEFACTOS ELECTRONICOS
CHANCHOS	MERCADERIA SURTIDA
CABEZAS DE GANADO	RELOJES
HUEVOS	ARTESANIAS
PATAS DE BOBINO	JUGUETES
POLLITOS/POLLOS/GALLOS	IMPRESIONES DE PAPEL
PAVOS	BISUTERIA
INSECTOS VARIOS	GAFAS_ LENTES
CABALLOS EQUINOS	MATERIAL DIDACTICO KILOS
BORREGOS	CETROS_ TIARAS_ CORONAS BISUTERIA FINA

Fuente: Elaboración Propia

En la tabla anterior se enlistan parte del contenido de la categorización del atributo subgrupo ya que por razones de confidencialidad y una lista muy extensa se ha optado por mostrar en la Tabla 16.

- **Atributo Procedencia**

Para el atributo de procedencia se pudo identificar las opciones presentadas en la Tabla 17 a continuación:

Tabla 17 - Categorización de atributo PROCEDENCIA

<b>PROCEDENCIA</b>	
ALEMANIA	ESPAÑA
BRASIL	EXTRANJERA
CHINA	FILIPINAS
COLOMBIA	PERU
ECUADOR	USA

Fuente: Elaboración Propia

- **Atributo Bodega**

Para el atributo Bodega se normalizó las opciones según la organización macro que maneja la entidad de control, por ello se agrupó las diferentes formas de hacer mención

a una misma bodega para así evitar redundancia de la información correspondiente a este campo obteniendo las opciones que se aprecia en la Tabla 18.

Tabla 18 - Categorización de atributo BODEGA

<b>SUBGRUPO</b>	
ALMACEN TEMPORAL BOLIVARIANA	BODEGA DE ADUANA - PTO. DE MANTA
ARETINA	BODEGA DE BOMASA
BASE OPERATIVA TABACUNDO	BODEGA DE SAI-SENAE AEROPUERTO JJO GYE
BODEGA 22 DE LA CAE PUERTO MARÍTIMO	BODEGA DE SAI-SENAE AEROPUERTO MS UIO
BODEGA CONTECON	BODEGA HUAQUILLAS

Fuente: Elaboración Propia

▪ **Atributo Marca**

En el caso del atributo Marca se identificó una lista muy amplia de marcas registradas en los registros por lo que en este atributo se realizó una limpieza para evitar incongruencias en la información, más no una categorización de estas por el volumen de información presentada en los datos a tratar. De esta forma se obtuvo registros de marcas libres de errores ortográficos y caracteres especiales que impidan su correcta identificación para su análisis. En la Tabla 19 se enlista parte del contenido del atributo marca.

Tabla 19- Categorización de atributo MARCA

<b>MARCA</b>		
7VEN	FORYOU	OLIVE_OIL
ABBA	FORZA	OLIVERS
ABBOTT	FOSFOCEREBRIN	OLK_005
ABC	FOSFOYOVAR	OLMECA
ABDUZEEDO	FOSKAPRIM	OLYMPIAN
ABEJA_REAL	FOSSIL	OLYMPUS
ABERCROMBIE	FOV_TECH	OMAS
ABISAI	FOX	OMEGA
ABLOY_CLIQ	FPL	OMER
ABOCOL_10-30-10	FRACO	OMINY

Fuente: Elaboración Propia

▪ **Atributo Estatus**

Para el atributo de Estatus se pudo identificar tres estados para valorar el estado en el que la mercadería se encuentra como se muestra en la Tabla 20.

Tabla 20 - Categorización de atributo ESTATUS

<b>ESTATUS</b>
BUENO
DAÑADO

USADO

Fuente: Elaboración Propia

▪ **Atributo Origen\_Aprehension**

Para el atributo correspondiente al Origen de la Aprehensión se identificó según a la información cotejada con el experto de la entidad de control, se eliminó redundancia y errores de digitación que hacían referencia a un mismo origen, pero constaba como registros diferentes. Obteniendo los datos que se muestra en la Tabla 21.

Tabla 21 - Categorización de atributo ORIGEN\_APREHENSION

<b>ORIGEN APREHENSION</b>	
ACCIONES INTELIGENCIA Y PROTECCION - DIP	CONTROL DE INGRESO DE DIVISAS
ACTA DE ENTREGA - RECEPCION	CONTROL DE RUTINA
ALLANAMIENTO	CONTROL EN CARRETERA
CONTROL AEROPUERTO	CONTROL FIJO
CONTROL CONJUNTO DE CONTENEDORES	CONTROL MOVIL

Fuente: Elaboración Propia

▪ **Atributo Sitio Aprehensión**

Para el atributo correspondiente al Sitio de Aprehensión se tomó en cuenta el texto insertado en los registros para poder determinar un campo que describa un lugar en específico dentro del territorio nacional, para posteriormente ubicarlo en un mapa que permita generar un mejor análisis de la información. Se tomó en cuenta la información en base lugares, cantones, ciudades, ríos, provincias pertenecientes al Ecuador. Por la extensión de la información a continuación en la Tabla 22 se presenta parte de los datos correspondientes al atributo.

Tabla 22 - Categorización de atributo SITIO\_APREHENSION

<b>SITIO_APREHENSION</b>	
ALOAG-ECUADOR	EL QUINCHE-ECUADOR
AMALUZA-ECUADOR	EL TAMBO-ECUADOR
AMBATO-ECUADOR	EL TELÉGRAFO-ECUADOR
AMBUQUI-ECUADOR	ESPEJO-ECUADOR
ANTONIO ANTE-ECUADOR	SARAGURO-ECUADOR
ARENILLAS-ECUADOR	SAUCILLO-ECUADOR
AZCAZUBI-ECUADOR	SAN LORENZO-ECUADOR
EL GUABO-ECUADOR	UCHUCAY-ECUADOR
EL PORTON-ECUADOR	RIO SIETE-ECUADOR
EL PROGRESO-ECUADOR	SOZORANGA-ECUADOR

Fuente: Elaboración Propia

- **Atributo Grupo Operativo**

El Grupo Operativo que realizó el proceso de retención y aprehensión se lo identificó en base a la información registrada y analizada con el experto de la entidad obteniendo los grupos operativos específicos como se muestra a continuación en la Tabla 23.

Tabla 23 - Categorización de atributo GRUPO\_OPERATIVO

<b>GRUPO_OPERATIVO</b>	
CONTROL CONJUNTO	CUERPO DE VIGILANCIA
INTERINSTITUCIONAL	ADUANERA
AGROCALIDAD	FUERZAS ARMADAS
ARCSA	POLICIA NACIONAL
DESCONOCIDO	INSPECTORIA DE PESCA

Fuente: Elaboración Propia

- **Atributo Distrito**

Para el atributo distrito se identificó los diez distritos acorde la organización y jurisdicción propia de la institución (Aduana del Ecuador, 2020). En la Tabla 24 se muestran las siglas correspondientes a cada distrito.

Tabla 24 - Categorización de atributo DISTRITO

<b>DISTRITO</b>	
IDG	VIDL
IIDM	VIIDPB
IIIDE	VIII DC
IVDQ	IXDH
VDA	XDLAT

Fuente:(Aduana del Ecuador, 2020)

- **Atributo Zona**

Acorde a la información facilitada por el experto de la entidad de control se agregó el campo Zonas que únicamente se encontraba en los registros del año 2019, obteniendo información de gran importancia. Para ello se determinó la relación entre el atributo Zona y el atributo Distrito de cada registro según la información de las direcciones zonales correspondiente a cada jurisdicción (Aduana del Ecuador, 2020), en la Tabla 25 de muestra la información correspondiente a cada Zona con sus Distritos.

Tabla 25 - Relación atributo Zonas con Distrito

<b>ZONAS</b>	<b>DISTRITO</b>	<b>NOMBRE DISTRITO</b>
ZONA1	III DE	Distrito Esmeraldas
	VIII DC	Distrito Carchi

ZONA2	IX DH V DA VII DPB	Distrito Huaquillas Distrito Azuay Distrito Puerto Bolívar
ZONA3	VI DL	Distrito Loja
ZONA4	I DG II DM	Distrito Guayas Distrito Manta
ZONA5	IV DQ X DLAT	Distrito Loja Distrito Latacunga

Fuente: (Aduana del Ecuador, 2020)

Acorde la relación del atributo Zonas con Distritos se pudo asignar el valor correspondiente a cada registro de la información mediante la herramienta de mapeo de Pentaho PDI.

#### ▪ **Atributo Cantidad**

Para el atributo cantidad se categorizó numéricamente acorde al volumen de los registros desde valores unitarios hasta millones acorde a la información obtenida en los registros de la información en el periodo 2014-2019 obteniendo como resultado las categorías que se detallan en la Tabla 26.

Tabla 26 - Categorización de atributo CANTIDAD

CATEGORÍA	VALORES (\$)
UNIDAD	0 a 9
DECENAS	10 a 100
CENTENAS	100 a 999
MILES	1000 a 999999
MILLONES	1000000 a 99999999

Fuente: Elaboración Propia

#### ▪ **Atributo Total**

Se categorizó el atributo total tomando en cuenta como base el monto del salario básico unificado ecuatoriano valorado en \$400 (Universo, 2019) obteniendo las categorías que se presentan en la Tabla 27. Además se consideró las normativas legal correspondientes al Código Orgánico Integral Penal (COIP), en el artículo 301 el cual menciona el contrabando que la persona que evada controles para el ingreso de mercadería por una cuantía igual o superior a diez salarios básicos será sancionada con pena privativa de la libertad de 3 a 5 años y una multa de hasta tres veces el valor en aduana de la mercancía del objeto del delito (Código Orgánico Integral Penal, 2014) .

El artículo 302 del COIP menciona el mal uso de exenciones o suspensiones tributarias aduaneras persona que venda, transfiera o use indebidamente mercancías cuya cuantía sea superior a ciento cincuenta salarios básicos unificados del trabajador en general misma que será sancionada con pena privativa de libertad de tres a cinco años y multa de hasta diez veces el valor de los tributos que se pretendió evadir considerando este delito de mayor gravedad (Código Orgánico Integral Penal, 2014).

Tabla 27 - Categorización atributo TOTAL

CATEGORÍA	VALORES (\$)
BAJO	0 a 400
MEDIO	400 a 3999.99
ALTO	4000.0 a 59999.99
MUY ALTO	60000.0 a 999999.99

Fuente: Elaboración Propia

- **Atributo Precio**

Para el atributo precio se tomó en cuenta las mismas consideraciones que para el atributo total por lo cual se generó la misma tabla para su categorización, por lo que las unidades pueden ir desde un solo artículo hasta millones con valores muy diferentes como se aprecia en la Tabla 28.

Tabla 28 - Categorización atributo PRECIO

CATEGORÍA	VALORES (\$)
BAJO	0 a 400
MEDIO	400.0 a 3999.99
ALTO	4000.0 a 59999.99
MUY ALTO	60000.0 a 999999.99

Fuente: Elaboración Propia

Para realizar la categorización de los atributos numéricos previamente mencionados se usó la opción *Number Ranges* en Petaho PDI permitiendo crear el campo adecuado para almacenar la información de cantidad, precio y total para su posterior categorización con datos cuantitativos, para ello se usó los valores máximos y mínimos mencionados en las tablas 26, 27 y 28. En la Fig. 11 se detalla como empleo el rango numérico para el caso de precio:



Figura 11- Categorización por rangos atributo PRECIO

Number ranges

Step name: RANGOS\_PRECIO

Input field: PRECIO

Output field: CAT\_PRECIO

Default value(if no range): unknown

Ranges (min <= x < max):

#	Lower Bound	Upper Bound	Value
1	0.0	399.99	BAJO
2	400.0	3999.99	MEDIO
3	4000.0	59999.99	ALTO
4	60000.0	9.999999999E9	MUY ALTO

OK Cancel Help

La aplicación de la técnica de agrupamiento necesita el uso de datos cuantitativos por lo cual se normalizo todos los datos disponibles, por cuestiones de confidencialidad acordados entre la entidad de control y el desarrollo del presente proyecto se presenta un parte de la información mas no su totalidad acorde a los parámetros a continuación:

- UNIDADES
  - Totalidad de parámetros: 59
  - DESCONOCIDA=1
  - AMPOLLA=2
  - ARROBA=3
  - ATADO=4
  - BALDE=5
  - BANDEJA=6
  - BIDÓN=7
  - BILLETE=8
  - BOLSA=9
  - BOTELLA=10
  
- GRUPO
  - Totalidad de parámetros: 27
  - ANIMALES VIVOS Y SUS PARTES=1
  - ARTEFACTOS ELECTRONICOS=2
  - ARTICULOS DE BAZAR=3
  - CALZADO=4
  - CIGARRILLOS=5
  - CONFITERIA=6

- COSMETICOS=7
- COURIER=8
- ELECTRONICA Y SUS ACCESORIOS=9
- EMBARCACIONES=10
  
- PROCEDENCIA
 

Totalidad de parámetros: 12

  - ALEMANIA=1
  - BRASIL=2
  - CHINA=3
  - COLOMBIA=4
  - DESCONOCIDA=5
  
- STATUS
 

Totalidad de parámetros: 3

  - BUENO=1
  - DAÑADO=2
  - USADO=3
  
- SUBGRUPO
 

Totalidad de parámetros: 218

  - AVES EN GENERAL=1
  - CHANCHOS=2
  - CABEZAS DE GANADO=3
  - HUEVOS=4
  - PATAS DE BOBINO=5
  - POLLITOS/POLLOS/GALLOS=6
  - PAVOS=7
  - INSECTOS VARIOS=8
  - CABALLOS EQUINOS=9
  - BORREGOS=10
  
- BODEGA
 

Totalidad de parámetros: 39

  - ALMACEN TEMPORAL BOLIVARIANA=1
  - ARETINA=2
  - BASE OPERATIVA TABACUNDO =3
  - BODEGA 22 DE LA CAE PUERTO MARÍTIMO=4
  - BODEGA CONTECON=5
  - BODEGA DE ADUANA - PTO. DE MANTA=6

- BODEGA DE BOMASA=7
- BODEGA DE SAI-SENAE AEROPUERTO JJO GYE=8
- BODEGA DE SAI-SENAE AEROPUERTO MS UIO=9
- BODEGA HUAQUILLAS=10
  
- ORIGEN\_APREHENSION
 

Totalidad de parámetros: 20

  - ACCIONES INTELIGENCIA Y PROTECCION - DIP=1
  - ACTA DE ENTREGA - RECEPCION=2
  - ALLANAMIENTO=3
  - CONTROL AEROPUERTO=4
  - CONTROL CONJUNTO DE CONTENEDORES=5
  - CONTROL DE INGRESO DE DIVISAS =6
  - CONTROL DE RUTINA=7
  - CONTROL EN CARRETERA=8
  - CONTROL FIJO=9
  - CONTROL MOVIL=10
  
- GRUPO OPERATIVO
 

Totalidad de parámetros: 8

  - DESCONOCIDO=1
  - AGROCALIDAD=2
  - ARCSA=3
  - CONTROL CONJUNTO INTERINSTITUCIONAL=4
  - CUERPO DE VIGILANCIA ADUANERA=5
  - FUERZAS ARMADAS=6
  - POLICIA NACIONAL=7
  - INSPECTORIA DE PESCA=8
  
- DISTRITO
 

Totalidad de parámetros: 10

  - IDG=1
  - IIDM=2
  - IIIIDE=3
  - IVDQ=4
  - VDA=5
  - VIDL=6
  - VIIDPB=7
  - VIIIIDC=8
  - IXDH=9

- XDLAT=10
- ZONA
  - Totalidad de parámetros: 5
  - ZONA1=1
  - ZONA2=2
  - ZONA3=3
  - ZONA4=4
  - ZONA5=5
- CANTIDAD
  - Totalidad de parámetros: 5
  - UNIDADES=1
  - DECENAS=2
  - CENTENAS=3
  - MILES=4
  - MILLON=5
- PRECIO
  - Totalidad de parámetros: 5
  - BAJO=1
  - MEDIO=2
  - ALTO=3
  - MUY ALTO=4
- TOTAL
  - Totalidad de parámetros: 5
  - BAJO=1
  - MEDIO=2
  - ALTO=3
  - MUY ALTO=4

### 2.6.3. Limpieza

En la fase de limpieza se corrigieron inconsistencias de la información encontradas generalmente en el atributo distrito del periodo 2014-2015 como se muestra en la Fig.12, también se transformó todos los campos de tipo *String* a Mayúsculas, para evitar redundancia de información como en la Fig. 13.

Figura 12 - Error en atributo DISTRITO con información incorrecta

Descripcion	Procedencia	Bodega	Status	Origen_Aprehension	Grupo_Operativo	Distrito
Reloj	Extranjera	Dirección Distrital "Manabi"	Bueno	Patrullaje de rutina (Orden de patrulla)	Patrulla del II Distrito "Manabi"	II DM
Televisor Plasma 50PB560D Serie 409RM...	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
LICOR BRANDY	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
Tabaco	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
Vino	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
Cerveza	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
PELUCAS	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
Televisor SONY LED BRAVIA DE 32" MOD K...	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
Televisor SONY LED BRAVIA DE 40" MOD K...	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
Televisor LED FHD SMART TV DE 40" MOD. ...	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
TELEVISOR DE 32" KALLEY K-LED32HDT O...	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
Productos naturales	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
PLAY STATION 4	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
CD PS4 EA SPORTS FIFA 15	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
LICOR BRANDY	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
CHIVAS REGAL	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
Whisky	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
TELEVISIÓN DE 32", SERIE 4000668, MODEL...	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
DINERO EN EFECTIVO	Extranjera	Direccion Distrital Esmeraldas	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
Prendas de vestir para hombre, mujer, niñ...	Extranjera	Dirección General SENAE	Bueno	Operativo conjunto CAE - FFAA	Patrulla CAE - FF.AA.	DOM
PANTALON CALENTADOR PARA MUJER	Extranjera	Direccion Distrital Cuenca	Bueno	Patrullaje de rutina (Orden de patrulla)	Patrulla del V Distrito "Azuay"	V DA
Fuegos artificiales	Extranjera	Direccion Distrital Cuenca	Bueno	Patrullaje de rutina (Orden de patrulla)	Patrulla del V Distrito "Azuay"	V DA
Pop Pop Snaps	Extranjera	Direccion Distrital Cuenca	Bueno	Patrullaje de rutina (Orden de patrulla)	Patrulla del V Distrito "Azuay"	V DA

En el atributo distrito se encontraron 112 registros con el error en de registro usando el dato DOM para ello.

De igual manera para los atributos de Bodega y Marca se encontró errores de digitación como se aprecia en la Fig. 13 por lo que se aplicó el proceso respectivo de limpieza.

Figura 13 - Errores de digitación en atributo BODEGA y MARCA

Descripcion	Procedencia	Bodega	Marca	Status	Origen_Aprehension
Prendas de Vestir	Extranjera	BODEGA DE CONTRABANDO BOMASA	VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)
Repuestos	Extranjera	BODEGA DE CONTRABANDO BOMASA	Ford	Bueno	Patrullaje de rutina (Orden de patrulla)
Whisky	Extranjera	BODEGA DE CONTRABANDO BOMASA	Chivas Regal	Bueno	Patrullaje de rutina (Orden de patrulla)
Whisky	Extranjera	BODEGA DE CONTRABANDO BOMASA	VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)
Prendas de Vestir	Extranjera	BODEGA DE CONTRABANDO BOMASA	VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)
STICKER	Extranjera	BODEGA DE CONTRABANDO BOMASA	S/M	Bueno	Patrullaje de rutina (Orden de patrulla)
MONOPATIN ELECTRONICO	Extranjera	Bodega del IX Distrito "Huaquillas"	MI	Bueno	Control de Rutina
TELEVISOR 50"	Extranjera	BODEGA DE CONTRABANDO BOMASA	Samsung	Bueno	Patrullaje de rutina (Orden de patrulla)
Prendas de Vestir	Extranjera	Dirección Distrital "Manabi"	VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)
Televisor	Extranjera	Dirección Distrital "Manabi"	SAMNSUNG	Bueno	Patrullaje de rutina (Orden de patrulla)
Cerveza	Extranjera	Bodega del IX Distrito "Huaquillas"	TRES CRUCES	Bueno	Control de Rutina
Cerveza	Extranjera	Bodega del IX Distrito "Huaquillas"	CUSQUEÑA TRIG-	Bueno	Control de Rutina
Celulares	Extranjera	Dirección Distrital "Manabi"	VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)
VEHICULO DE PLACAS BVH981	Extranjera	Dirección Distrital "Manabi"	Renault	Bueno	Patrullaje de rutina (Orden de patrulla)
Prendas de Vestir	Extranjera	Bodega del IX Distrito "Huaquillas"	VM	Bueno	Control de Rutina
Prendas de Vestir	Extranjera	Dirección Distrital "Manabi"	VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)
CALZADO	Extranjera	Bodega del IX Distrito "Huaquillas"	VM	Bueno	Control de Rutina

Para los atributos de Sitio de Aprehensión y Grupo operativo se pudo apreciar igualmente errores de digitación al igual que el uso de mayúsculas y minúsculas sin una normalización como se muestra en la Fig. 14.

Figura 14 - Errores de digitación en atributo SITIO APREHENSION y GRUPO OPERATIVO

Marca	Status	Origen_Aprehension	Sitio_Aprehension	Grupo_Operativo
VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)	La Pila Provincia de Manabí	Patrulla del II Distrito "Manabí"
Renault	Bueno	Patrullaje de rutina (Orden de patrulla)	Ciudad de Manta	Patrulla del II Distrito "Manabí"
VM	Bueno	Control de Rutina	Control Buses del Destacamento de Chacras	Turno de Guardia
VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)	Tierra Bonita - Canton Rocafuerte - Manabi	Patrulla del II Distrito "Manabí"
VM	Bueno	Control de Rutina	CONTROL DE CARRETERA DEL DESTACAMENTO CHACRAS	Turno de Guardia
RCF	Bueno	Patrullaje de rutina (Orden de patrulla)	Sector Puerto Inca	Patrulla del I Distrito "Guayas"
SM	Bueno	Control de Rutina	CONTROL DE CARRETERA DEL DESTACAMENTO CHACRAS	Turno de Guardia
VM	Bueno	Control de Rutina	CONTROL DE CARRETERA DEL DESTACAMENTO CHACRAS	Turno de Guardia
VM	Bueno	Control de Rutina	CONTROL DE CARRETERA DEL DESTACAMENTO CHACRAS	Turno de Guardia
S/M	Bueno	Patrullaje de rutina (Orden de patrulla)	PEAJE NARANJAL	Patrulla del I Distrito "Guayas"
Renault	Bueno	Patrullaje de rutina (Orden de patrulla)	Destacamento Yahuarcocha	PATRULLA YAHUARCOCHA
YAMAHA	Bueno	Patrullaje de rutina (Orden de patrulla)	Destacamento Yahuarcocha	PATRULLA YAHUARCOCHA
VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)	Destacamento Yahuarcocha	PATRULLA YAHUARCOCHA
Renault	Bueno	Patrullaje de rutina (Orden de patrulla)	Destacamento Yahuarcocha	PATRULLA YAHUARCOCHA
Renault	Bueno	Patrullaje de rutina (Orden de patrulla)	Destacamento Yahuarcocha	PATRULLA YAHUARCOCHA
VM	Bueno	Control de Rutina	CONTROL DE CARRETERA DEL DESTACAMENTO CHACRAS	Turno de Guardia
VARIAS	Bueno	Patrullaje de rutina (Orden de patrulla)	REDONDEL DE GONZALES SUAREZ	PATRULLA TABACUNDO

Para la corrección del error en Distrito se consultó con el experto de la empresa para poder filtrar los registros para separarlos y asignar el distrito a cada uno de ellos en base a su relación con el atributo de Sitio de Aprehensión mediante la herramienta de mapeo en Pentaho como muestra la Fig. 15.

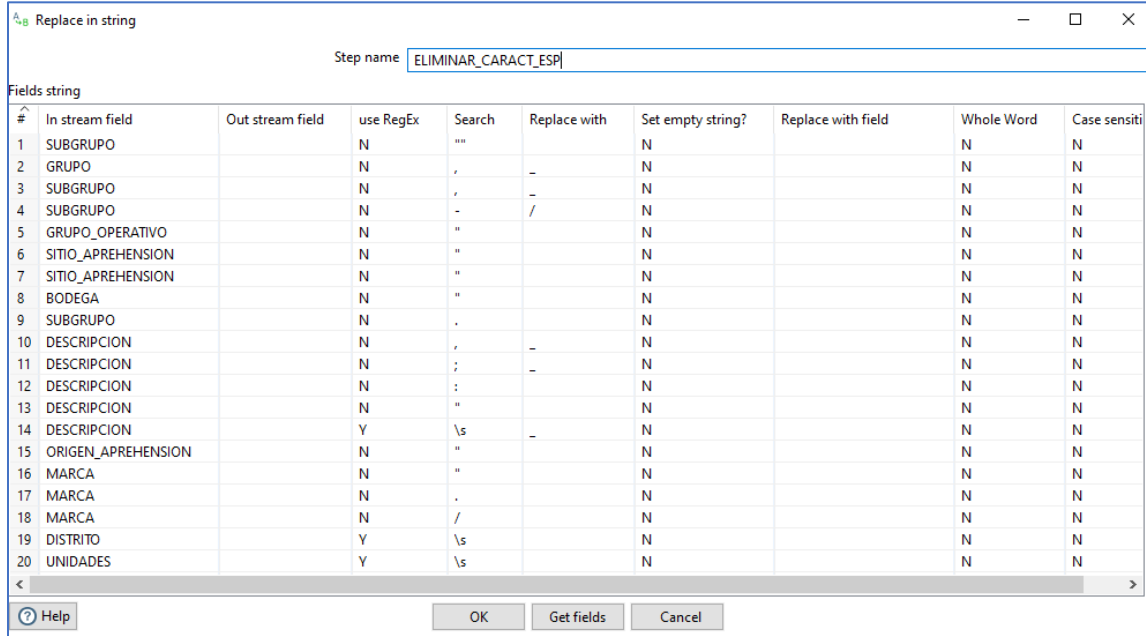
Figura 15 - Mapeo del atributo Distrito con registro erróneo DOM

#	Source value	Target value
1	Puerto de cabotaje Esmeraldas	III DE
2	Ciudad de Esmeraldas	III DE
3	Sector Esteros del Viento	III DE
4	CANAL DEL VIENTO	III DE
5	sector de Bucay	III DE
6	Puerto de Cabotaje de la ciudad de Esmeraldas	III DE
7	Alta mar	III DE
8	Población de Viche - Esmeraldas	III DE
9	Vía Esmeraldas - San Lorenzo	III DE
10	La Pila Provincia de Manabí	II DM
11	Tierra Bonita - Canton Rocafuerte - Manabi	II DM
12	Puertos Privados de Guayaquil	I DG

En el caso de los caracteres especiales usados en los campos de tipo texto se usó la opción de *Replace String* Fig. 16 para evitar que existan 2 opciones que se las atribuya a un solo lugar en específico además para el atributo SubGrupo se eliminó las comas, espacios en blanco y se

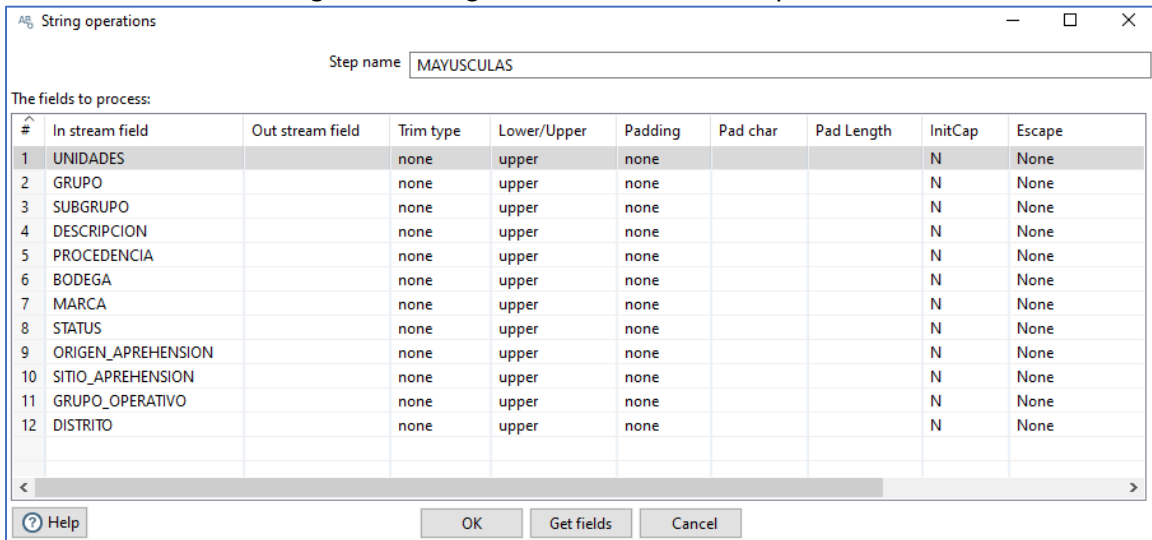
reemplazó con guión bajo ciertos caracteres especiales para evitar erros posteriores en minería de datos.

Figura 16 - Eliminación de caracteres especiales



Con el mismo propósito anterior de mantener una información libre de errores se usó *String Operations* Fig. 17 para que todos los registros sean manejados en letras mayúsculas, en el caso de los atributos de tipo texto con los que se dispone.

Figura 17 - Registros únicamente en mayúsculas



#### 2.6.4. Aplicación de Normativa ISO25012

La aplicación de la ISO25012 es un factor clave para poder considerar la calidad de la información en base a la característica seleccionada para el presente proyecto, la característica de **consistencia** permite identificar cuan coherentes y libres de contradicción se encuentra el conjunto de datos. Los valores obtenidos, como resultado de la aplicación de la métrica se los considera aceptable según la escala que se observa en la Tabla 29.

Tabla 29- Escala de valoración para característica de *Consistencia*

Escala	Interpretación
Porcentaje% <= 10%	Casi ningún dato se corresponde con las reglas definidas
Porcentaje%>10% & Porcentaje%<=45%).	Algunos datos se corresponden con las reglas, pero se observan muchos errores
Porcentaje% >45% & Porcentaje% <=85%	Muchos datos se corresponden con las reglas definidas y se observan algunos errores
Porcentaje% Obtenido >85%	La gran totalidad de los datos se corresponden con las reglas definidas

Fuente: adaptación de Calabrese et al., 2019

Al tener una base de datos de una única tabla se evalúa la consistencia de los datos en cada atributo, en la Tabla 30 se aprecia los resultados obtenidos:

Tabla 30 – Valores de aplicación de métrica

ATRIBUTO	CANTIDAD DE DATOS NO COHERENTES	VALOR DE FUNCIÓN DE MEDICIÓN	PORCENTAJE%
CAT_CANTIDAD	0	1	100,00
UNIDADES	10	0,999902598	99,99
CAT_PRECIO	0	1	100,00
TOTAL	0	1	100,00
CAT_TOTAL	0	1	100,00
GRUPO	3807	0,962918952	96,29
SUBGRUPO	1301	0,987327963	98,73
DESCRIPCION	0	1	100,00
PROCEDENCIA	11753	0,885523099	88,55
BODEGA	2949	0,971276067	97,13
MARCA	0	1	100,00
STATUS	0	1	100,00
ORIGEN_APREHENSION	19	0,999814936	99,98
SITIO_APREHENSION	0	1	100,00
GRUPO_OPERATIVO	106	0,998967536	99,90
DISTRITO	0	1	100,00
ZONA	0	1	100,00

Fuente: Elaboración Propia



Con los resultados obtenidos en la tabla previa se puede evidenciar que la característica de consistencia para cada uno de los atributos se encuentra dentro de la escala más alta de valoración dando a entender que el conjunto de datos posee información consistente en un promedio de 98,86% de su totalidad de variables.

## 2.7. Minería de datos

La presente investigación tiene como objetivo principal la identificación de patrones de contrabando dentro del territorio nacional, por este motivo se usó la totalidad de la información correspondiente a 102667 registros, de esta manera se encontrará las probables relaciones que tienen los diferentes atributos entre sí para identificar las situaciones en las cuales este tipo de ilícito de desarrolla en el Ecuador. Para la aplicación de los diferentes algoritmos propuestos de uso la herramienta **Weka**, misma que recibe un tipo de archivo en \*.csv este se lo obtuvo al transformar el archivo Excel resultante del preprocesamiento de la información.

En las figuras a continuación se aprecia la vista de los archivos tanto para la técnica de agrupamiento misma que requiere de información de tipo *Nominal* para la aplicación de los algoritmos como se aprecia en la Fig. 18, mientras que para el caso de las técnicas de asociación los algoritmos requieren que la información sea de tipo *Numeric* por lo que utilizó la información normalizada en la sección 2.6.2 como se muestra en la Fig. 19.

Figura 18 – Vista minable para asociación

1	ANIO, MES, DIA, CAT_CANTIDAD, UNIDADES, CAT_PRECIO, TOTAL, CAT_TOTAL, GRUPO, SUBGRUPO, DESCRIPCION, PROCEDENCIA, BODEGA, MARCA, S
2	2019, 1, 9, MILES, UNIDAD, BAJO, 7500, ALTO, CIGARRILLOS, CIGARRILLOS, CIGARRILLOS, EXTRANJERA, DIRECCION DISTRITAL LATACUNGA, VM
3	2019, 1, 9, UNIDADES, UNIDAD, MEDIO, 450, MEDIO, ELECTRONICA Y SUS ACCESORIOS, TELEVISORES, TELEVISOR, EXTRANJERA, DIRECCION DI
4	2019, 1, 9, UNIDADES, UNIDAD, MEDIO, 500, MEDIO, ELECTRONICA Y SUS ACCESORIOS, TELEVISORES, TELEVISOR, EXTRANJERA, DIRECCION DI
5	2019, 1, 9, UNIDADES, UNIDAD, MEDIO, 3400, MEDIO, REPUESTOS Y ACCESORIOS, DE VEHICULOS, AIRBAG, EXTRANJERA, DIRECCION DISTRITAL
6	2019, 1, 9, DECENAS, UNIDAD, BAJO, 2830.000001, MEDIO, ARTICULOS DE BAZAR, MERCADERIA SURTIDA, PRODUCTOS_RELIGIOSOS, EXTRANJER
7	2019, 1, 9, UNIDADES, UNIDAD, BAJO, 300, BAJO, ELECTRONICA Y SUS ACCESORIOS, CELULARES_ TELEFONOS_ FAXES_ HIPHONE, CELULAR, EX
8	2019, 1, 9, CENTENAS, UNIDAD, BAJO, 2110.000001, MEDIO, TEXTILES, PRENDAS DE VESTIR/NUEVAS, PRENDAS_DE_VESTIR, EXTRANJERA, DIRE
9	2019, 1, 9, UNIDADES, UNIDAD, BAJO, 300, BAJO, ELECTRONICA Y SUS ACCESORIOS, CELULARES_ TELEFONOS_ FAXES_ HIPHONE, CELULAR, EX
10	2019, 1, 9, UNIDADES, UNIDAD, MEDIO, 450, MEDIO, ELECTRONICA Y SUS ACCESORIOS, TELEVISORES, TELEVISOR, EXTRANJERA, DIRECCION GEI
11	2019, 1, 9, UNIDADES, UNIDAD, BAJO, 250, BAJO, ELECTRONICA Y SUS ACCESORIOS, TELEVISORES, TELEVISOR, EXTRANJERA, BODEGA HUAQUILL
12	2019, 1, 9, UNIDADES, UNIDAD, BAJO, 250, BAJO, ELECTRONICA Y SUS ACCESORIOS, COMPUTADORAS_ LAPTOPS_ TABLETS, COMPUTADOR_ALL_I
13	2019, 1, 9, UNIDADES, UNIDAD, BAJO, 50, BAJO, ELECTRONICA Y SUS ACCESORIOS, PARLANTES/AMPLIFICADORES, PARLANTE, EXTRANJERA, BODI
14	2019, 1, 9, DECENAS, UNIDAD, BAJO, 520, MEDIO, ARTICULOS DE BAZAR, MERCADERIA SURTIDA, DISFRACES, EXTRANJERA, BODEGA HUAQUILLAS
15	2019, 1, 9, DECENAS, BOTELLA, BAJO, 300, BAJO, LICORES, WHISKY VARIAS MARCAS, WISKY, EXTRANJERA, BODEGA HUAQUILLAS, JOHNNIE_WALKI
16	2019, 1, 9, UNIDADES, UNIDAD, ALTO, 4500, ALTO, MATERIAL DE TRANSPORTE, VEHICULOS, VEHICULO_COLOMBIANO_DE_PLACA_XEJ-541, EXTRAI
17	2019, 1, 9, UNIDADES, BOTELLA, BAJO, 120, BAJO, LICORES, LICORES VARIOS, LICOR, EXTRANJERA, BODEGA HUAQUILLAS, ANTIOQUENO, BUENO, (
18	2019, 1, 9, DECENAS, KILOGRAMO, BAJO, 33.075, BAJO, FRUTAS Y COMESTIBLES, TAMARINDO, TAMARINDO, EXTRANJERA, BODEGA HUAQUILLAS, SI
19	2019, 1, 9, UNIDADES, SACO, BAJO, 105, BAJO, FRUTAS Y COMESTIBLES, CEBOLLA, CEBOLLA_ROJA, EXTRANJERA, BODEGA HUAQUILLAS, SM, BUEN
20	2019, 1, 9, UNIDADES, UNIDAD, BAJO, 350, BAJO, ELECTRONICA Y SUS ACCESORIOS, TELEVISORES, TELEVISOR, EXTRANJERA, DESTACAMENTO Y/
21	2019, 1, 9, MILES, UNIDAD, BAJO, 10000, ALTO, CIGARRILLOS, CIGARRILLOS, CIGARRILLOS, EXTRANJERA, DESTACAMENTO YAHUARCOCHA, MARSHAI
22	2019, 1, 9, UNIDADES, UNIDAD, MUY ALTO, 65000, MUY ALTO, MATERIAL DE TRANSPORTE, VEHICULOS, VEHICULO_DE_PLACAS_IIL-213_COLOMB
23	2019, 1, 9, DECENAS, UNIDAD, BAJO, 300, BAJO, LICORES, WHISKY VARIAS MARCAS, WISKY, EXTRANJERA, BODEGA HUAQUILLAS, SOMETHING_SPE
24	2019, 1, 9, UNIDADES, UNIDAD, BAJO, 1200, MEDIO, VARIOS, LLANTAS/NEUMATICOS, LLANTAS, EXTRANJERA, BODEGA HUAQUILLAS, VIKRANT, BUEI
25	2019, 1, 9, UNIDADES, UNIDAD, MEDIO, 1800, MEDIO, VARIOS, LLANTAS/NEUMATICOS, LLANTAS, EXTRANJERA, BODEGA HUAQUILLAS, HIFLY, BUEN
26	2019, 1, 9, UNIDADES, UNIDAD, BAJO, 750, MEDIO, VARIOS, LLANTAS/NEUMATICOS, LLANTAS, EXTRANJERA, BODEGA HUAQUILLAS, LINGLONG, BUEI

Figura 19 - Vista minable para atípicos/agrupamiento

1	ANIO, MES, DIA, CAT_CANTIDAD, UNIDADES, CAT_PRECIO, CAT_TOTAL, GRUPO, SUBGRUPO, PROCEDENCIA, BODEGA, STATUS, ORIGEN_APREHENSION
2	2019,1,9,4,59,1,3,5,25,9,25,1,8,45,5,10,5
3	2019,1,9,1,59,2,2,9,32,9,30,1,8,228,5,8,1
4	2019,1,9,1,59,2,2,9,32,9,30,1,8,228,5,8,1
5	2019,1,9,1,59,2,2,24,181,9,30,1,8,207,5,8,1
6	2019,1,9,2,59,1,2,3,12,9,30,1,8,207,5,8,1
7	2019,1,9,1,59,1,1,9,33,9,11,1,8,71,5,9,2
8	2019,1,9,3,59,1,2,26,188,9,30,1,8,207,5,8,1
9	2019,1,9,1,59,1,1,9,33,9,11,1,8,71,5,9,2
10	2019,1,9,1,59,2,2,9,32,9,31,1,8,207,5,8,1
11	2019,1,9,1,59,1,1,9,32,9,11,1,8,71,5,9,2
12	2019,1,9,1,59,1,1,9,34,9,11,1,8,71,5,9,2
13	2019,1,9,1,59,1,1,9,35,9,11,1,8,71,5,9,2
14	2019,1,9,2,59,1,2,3,12,9,11,1,8,71,5,9,2
15	2019,1,9,2,10,1,1,13,118,9,11,1,8,71,5,9,2
16	2019,1,9,1,59,3,3,16,147,9,21,1,8,127,5,4,5
17	2019,1,9,1,10,1,1,13,119,9,11,1,8,71,5,9,2
18	2019,1,9,2,33,1,1,12,56,9,11,1,8,71,5,9,2
19	2019,1,9,1,50,1,1,12,57,9,11,1,8,71,5,9,2
20	2019,1,9,1,59,1,1,9,32,9,21,1,8,127,5,4,5
21	2019,1,9,4,59,1,3,5,25,9,21,1,8,127,5,4,5
22	2019,1,9,1,59,4,4,16,147,9,21,1,8,127,5,4,5
23	2019,1,9,2,59,1,1,13,118,9,11,1,8,71,5,9,2

### 2.7.1. Componentes Principales

El algoritmo PCA o también denominado Análisis de Componentes Principales, agrupa variables que tienen correlación entre sí de las que no, con el propósito de obtener un conjunto de variables reducido que permita un análisis más fácil y una mejor su interpretación (Quiroga-Juárez & Villalobos-Escobedo, 2015).

Para la aplicación del algoritmo PCA se necesita un conjunto de datos categorizados de tipo numérico, para este caso se utilizó la data obtenida previamente y almacenada en el archivo de tipo csv correspondiente a la vista minable para atípicos y agrupamiento Fig.20, en la figura a continuación se muestran los resultados obtenidos del algoritmo en Weka.

Figura 20 – Resultados algoritmo PCA en Weka

The screenshot displays the Weka software interface for the PCA algorithm. The 'Attribute Evaluator' is set to 'PrincipalComponents -R 0.95-A 5'. The 'Search Method' is 'Ranker -T-1.7976931348623157E308-N-1'. The 'Attribute Selection Mode' is 'Use full training set'. The 'Attribute selection output' shows a list of ranked attributes with their respective weights. The 'Result list' shows three ranked results for 'Ranker + PrincipalComponent'.

Los parámetros usados para el PCA son las 17 variables categorizadas, sin clase seleccionada y con una varianza de 0.95 para la obtención de resultados más acertados, de esta forma los conjuntos de datos más acertados tenemos los siguientes reflejados en la Tabla 31.

Tabla 31 - Atributos mejor clasificados mediante PCA

N°	%REPRESENTATIVO	ATRIBUTOS
PC1	0.8725	SUBGRUPO
		GRUPO
		SITIO_APREHENSION
		GRUPO_OPERATIVO
		BODEGA
PC2	0.7561	DISTRITO
		ZONA
		GRUPO_OPERATIVO
		ORIGEN_APREHENSION
PC3	0.6509	CAT_TOTAL
		GRUPO
		GRUPO_OPERATIVO
		ORIGEN_APREHENSION
		SUBGRUPO
		SITIO_APREHENSION

Fuente: Weka

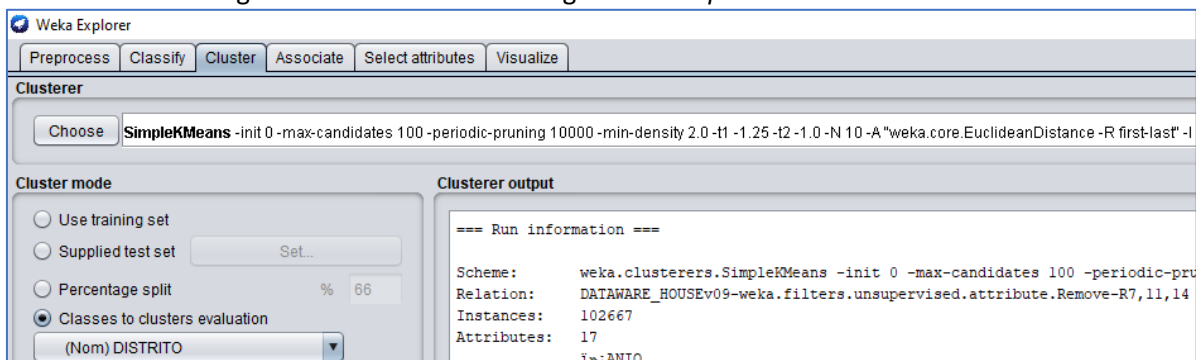
## 2.7.2. Agrupamiento (Clustering)

### Algoritmo K-means

Para el algoritmo *k-means* se aplicó la data en su totalidad con el fin de obtener el mayor provecho posible de los datos existentes. En esencia este algoritmo permite clasificar o segmentar la información con el propósito de identificar las estructuras dentro del conjunto de datos inicialmente sin etiquetas o clases (Minguillón & Casas, 2017). Cada clúster se lo define por un punto identificado como centroide para posteriormente usar el cálculo de la distancia euclídea para encontrar la menor distancia.

Para la aplicación de este algoritmo se estableció 10 clústers acorde al atributo distrito, permitiendo de esta manera identificar específicamente los conjuntos de datos para cada uno de los puntos de vigilancia de la entidad de control, como se aprecia en la Fig. 21.

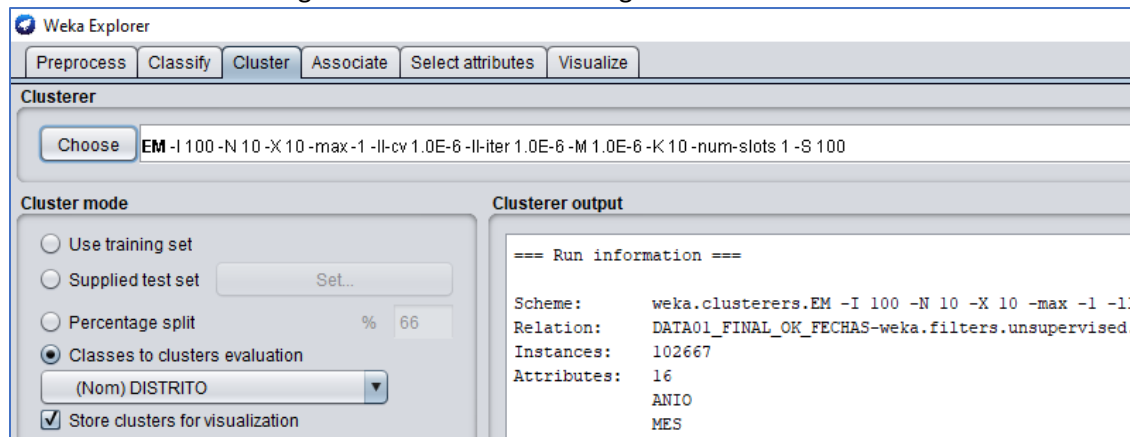
Figura 21 – Parámetros del algoritmo *SimpleKmeans* en Weka



### Algoritmo EM

El algoritmo conocido como *EM* o Expectativa Maximización, se lo usa para identificar estimaciones de máxima verosimilitud de fácil implementación y estabilidad numérica (Huang & Chen, 2017), se realiza una asignación de una distribución de probabilidad a cada instancia indicando su pertenencia a los diferentes clústers creados según la validación cruzada o específicamente por el número de clústers a utilizar. Al igual que el algoritmo se estableció 10 clústers para el agrupamiento de la información en base al atributo distrito, los parámetros para la ejecución del algoritmo se los puede apreciar en la Fig. 22 a continuación:

Figura 22 - Parámetros del algoritmo *EM* en Weka



### 2.7.3. Asociación

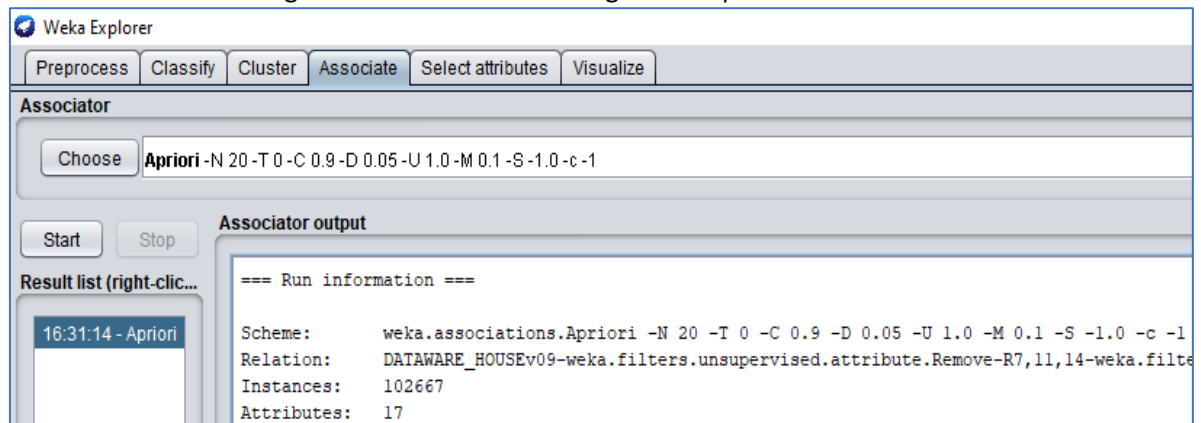
#### Algoritmo Apriori

El algoritmo *apriori* propuesto desde el año 1994 se lo considera el más robusto para la definir reglas de asociación, aplica un enfoque iterativo para la exploración del conjunto de datos (Han et al., 2013). En primer lugar, se identifica el conjunto de frecuentes y partiendo

de ello analiza la data completa para acumular el recuento de cada elemento y recopila aquellos que satisfacen las métricas establecidas (soporte y confianza) ejecutando este proceso hasta encontrar los conjuntos de elementos más frecuentes para establecer las reglas de asociación.

En el caso del algoritmo *apriori* se estableció el valor de la confianza en un mínimo de 0.9 y un soporte mínimo de 0.4 en base a pruebas experimentales de la información descriptiva con el fin de obtener las reglas más adecuadas, los parámetros definidos se los puede apreciar en la Fig. 23.

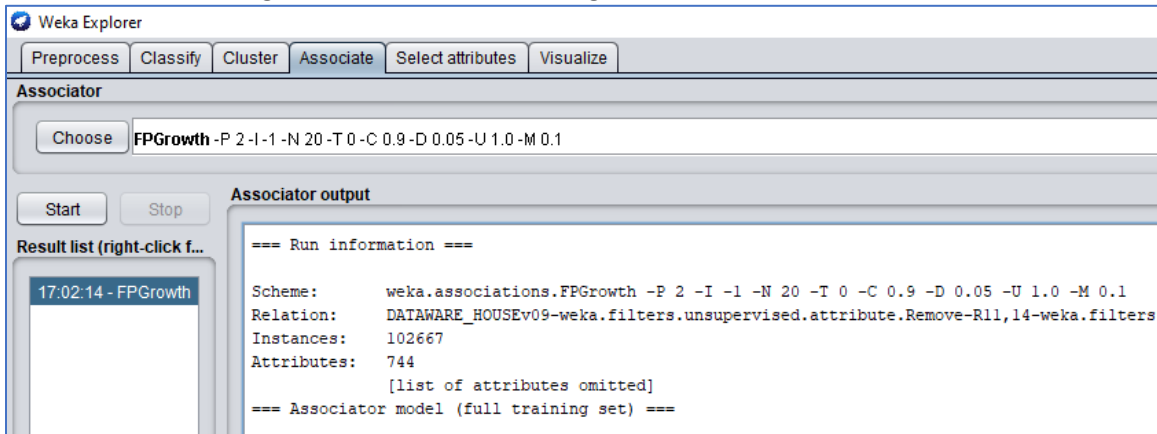
Figura 23 - Parámetros del algoritmo *Apriori* en Weka



### Algoritmo FP Growth

Para la aplicación del algoritmo FP Growth se estableció el valor de la confianza en un mínimo de 0.9 y un soporte mínimo de 0.1 como se puede apreciar en el Fig. 24, estos parámetros se establecieron conforme a las pruebas realizadas con la información binarizada, la misma que consiste en el proceso de mapeo de la información en donde se realizó un proceso de clasificación para cada multiclase obteniendo así una serie de opciones binarias, para la identificación de los resultados en clasificadores binarios (Pineda, 2009), con el fin de obtener las reglas más adecuadas en el caso de este algoritmo.

Figura 24 – Parámetros del algoritmo *FP Growth* en Weka



#### 2.7.4. Atípicos

En el proceso de análisis de datos, se puede observar varias inconsistencias, anomalías o datos que a primera impresión no tienen ninguna relación con la mayoría de la información. Para la identificación de los valores denominados atípicos se debe tomar en cuenta las medidas de dispersión como el recorrido intercuartílico en conjunto con los cuartiles para poder considerar cuales con los valores atípicos presentes en la información (San-Segundo & Marv, 2016). Este anlisis se realizar con el total de atributos para obtener datos relevantes de cada atributo disponible. Para la identificacin de los valores atpicos se realiz un anlisis estadstico (Anexo 3) al disponer de informacin de tipo cualitativa.

## CAPÍTULO 3

### Proceso de descubrimiento del Conocimiento

#### 3.1. Evaluación e interpretación

##### 3.1.1. Evaluación, análisis e interpretación de tareas de Clustering

###### Algoritmo *Kmeans*

Para la aplicación de la técnica de clustering se tomó en cuenta como atributo clase a DISTRITO, obteniendo como resultados la siguiente información:

Tabla 32 - Resultado de algoritmo *KMeans* (Parte 1)

	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
AÑO	2015	2016	2015	2016	2016
MES	2	3	8	8	1
DIA	4	1	31	7	18
CAT_CANTIDAD	DECENAS	UNIDADES	UNIDADES	CENTENAS	DECENAS
UNIDADES	UNIDAD	UNIDAD	UNIDAD	PAR	UNIDAD
CAT_PRECIO	BAJO	BAJO	BAJO	BAJO	BAJO
CAT_TOTAL	BAJO	BAJO	BAJO	MEDIO	BAJO
GRUPO	TEXTILES	TEXTILES	LICORES	TEXTILES	TEXTILES
SUBGRUPO	PRENDAS DE VESTIR/NUEVAS	PRENDAS DE VESTIR/NUEVAS	WHISKY VARIAS MARCAS	MEDIAS DE VESTIR_NYLON_PANTYS_TOBILLERAS	PRENDAS DE VESTIR/NUEVAS
PROCEDENCIA	EXTRANJERA	EXTRANJERA	EXTRANJERA	EXTRANJERA	EXTRANJERA
BODEGA	DESTACAMENTO CHACRAS	DIRECCION DISTRITAL ESMERALDAS	DESTACAMENTO CHACRAS	BODEGA HUAQUILLAS	BODEGA HUAQUILLAS
STATUS	BUENO	BUENO	BUENO	BUENO	BUENO
ORIGEN_APREHENSION	CONTROL DE RUTINA	ACTA DE ENTREGA - RECEPCION	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE RUTINA
SITIO_APREHENSION	CHACRAS-ECUADOR	ESMERALDAS-ECUADOR	CHACRAS-ECUADOR	CHACRAS-ECUADOR	CHACRAS-ECUADOR
GRUPO_OPERATIVO	CUERPO DE VIGILANCIA ADUANERA	POLICIA NACIONAL	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA
ZONA	ZONA2	ZONA1	ZONA2	ZONA2	ZONA2

Fuente: Weka

Tabla 33 - Resultado de algoritmo *KMeans* (Parte 2)

	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
AÑO	2015	2016	2015	2016	2016
MES	4	6	7	11	2
DIA	26	6	24	22	28
CAT_CANTIDAD	UNIDADES	UNIDADES	DECENAS	DECENAS	UNIDADES
UNIDADES	SACO	UNIDAD	PAR	CAJA	PAR
CAT_PRECIO	BAJO	MEDIO	BAJO	BAJO	BAJO
CAT_TOTAL	BAJO	MEDIO	MEDIO	MEDIO	BAJO
GRUPO	FRUTAS Y COMESTIBLES	MAQUINAS_ EQUIPOS Y APARATOS COMPLETOS	CALZADO	FRUTAS Y COMESTIBLES	TEXTILES
SUBGRUPO	ARROZ	DE LABORATORIO Y SUS ACCESORIOS	DEPORTIVO	PEPARACIONES ALIMENTICIAS DIVERSAS/ENLATADOS	MEDIAS DE VESTIR_NYLON_

PROCEDENCIA	EXTRANJERA	DESCONOCIDA	EXTRANJERA	EXTRANJERA	PANTYS_TOBILLERAS... EXTRANJERA
BODEGA	DESTACAMENTO CHACRAS	DIRECCION DISTRITAL PUERTO BOLIVAR	BODEGA CONTECON	DIRECCION DISTRITAL TULCAN	DIRECCION DISTRITAL TULCAN
STATUS	BUENO	BUENO	BUENO	BUENO	BUENO
ORIGEN_APREHENSION	CONTROL DE RUTINA	CONTROL DE RUTINA	ALLANAMIENTO	CONTROL DE RUTINA	CONTROL DE RUTINA
SITIO_APREHENSION	CHACRAS-ECUADOR	RIO SIETE-ECUADOR	GUAYAQUIL-ECUADOR	TULCAN-ECUADOR	RUMICHACA-ECUADOR
GRUPO_OPERATIVO	CUERPO DE VIGILANCIA	CUERPO DE VIGILANCIA	CUERPO DE VIGILANCIA	CUERPO DE VIGILANCIA	CUERPO DE VIGILANCIA
ZONA	ADUANERA ZONA2	ADUANERA ZONA2	ADUANERA ZONA4	ADUANERA ZONA1	ADUANERA ZONA1

Fuente: Weka

### Análisis e interpretación

Para analizar los resultados obtenidos se debe tomar en cuenta que en el algoritmo *KMeans* se lleva a cabo en una partición de conjunto de datos a consideración con el propósito de obtener subconjuntos que alcancen particiones óptimas logrando encontrar la información más representativa para cada clúster (Lara Torralbo, 2014). En la sección 3.3 se detallarán la interpretación de los datos anteriores.

Con el propósito de obtener una mayor cantidad de patrones que permitan una mejor interpretación y análisis de la información con la que se trabaja, se aplicó a cada Componente Principal (PC) obtenido en la sección 2.7.4 el algoritmo *KMeans*.

Para el caso del **PC1** se puede identificar los siguientes resultados para cada uno de los clústers:

Interpretar tablas, párrafo que explique cada una de ellas

Tabla 34 - Resultado de algoritmo *KMeans* PC1 (Parte 1)

	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
<b>GRUPO</b>	TEXTILES	FRUTAS Y COMESTIBLES	LICORES	FRUTAS Y COMESTIBLES	TEXTILES
<b>SUBGRUPO</b>	PRENDAS DE VESTIR/NUEVAS	CIGARILLOS	WHISKY VARIAS MARCAS	LLANTAS/ NEUMATICOS	PRENDAS DE VESTIR/NUEVAS
<b>BODEGA</b>	DESTACAMENTO CHACRAS	DIRECCION DISTRITAL PUERTO BOLIVAR	DESTACAMENTO CHACRAS	BODEGA HUAQUILLAS	BODEGA HUAQUILLAS
<b>SITIO_APREHENSION</b>	CHACRAS-ECUADOR	PUERTO BOLIVAR-ECUADOR	CHACRAS-ECUADOR	CHACRAS-ECUADOR	CHACRAS-ECUADOR
<b>GRUPO_OPERATIVO</b>	CUERPO DE VIGILANCIA ADUANERA	POLICIA NACIONAL	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA

Fuente: Weka



En la Tabla 34 que se aprecia anteriormente se presentan los datos correspondientes a los 5 primeros clústers resultantes de la aplicación del algoritmo KMeans al conjunto de datos correspondientes al PC1.

Tabla 35 - Resultado de algoritmo *KMeans* PC1 (Parte 2)

	CLUSTER 5	CLUSTER 6	CLUSTER 7	CLUSTER 8	CLUSTER 9
<b>GRUPO</b>	FRUTAS Y COMESTIBLES	ARTICULOS DE BAZAR	CALZADO	ELECTRONICA Y SUS ACCESORIOS	ELECTRONICA Y SUS ACCESORIOS
<b>SUBGRUPO</b>	CEBOLLA	MERCADERIA SURTIDA	CALZADO EN GENERAL	CELULARES_ TELEFONOS_ FAXES_ HIPHONE	TELEVISORES
<b>BODEGA</b>	DESTACAMENTO CHACRAS	DIRECCION DISTRITAL PUERTO BOLIVAR	BODEGA CONTECON	DIRECCION DISTRITAL TULCAN	DIRECCION DISTRITAL TULCAN
<b>SITIO_APREHENSION</b>	CHACRAS- ECUADOR	RIO SIETE- ECUADOR	GUAYAQUIL- ECUADOR	TULCAN- ECUADOR	RUMICHACA- ECUADOR
<b>GRUPO_OPERATIVO</b>	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA

Fuente: Weka

Para el caso de la Tabla 35 se aprecia datos correspondientes a los clústers del 5 al 9 resultantes de la aplicación del algoritmo KMeans al conjunto de datos correspondientes al PC1.

En el caso del **PC2** se puede identificar los siguientes resultados para cada uno de los clústers:

Tabla 36 - Resultado de algoritmo *KMeans* PC2 (Parte 1)

	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
<b>CAT_TOTAL</b>	BAJO	BAJO	MEDIO	MEDIO	MEDIO
<b>ORIGEN_APREHENSION</b>	CONTROL DE RUTINA	ACTA DE ENTREGA - RECEPCION	CONTROL DE RUTINA	ALLANAMIENTO	CONTROL DE RUTINA
<b>GRUPO_OPERATIVO</b>	CUERPO DE VIGILANCIA ADUANERA	POLICIA NACIONAL	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA
<b>ZONA</b>	ZONA2	ZONA2	ZONA2	ZONA4	ZONA1

Fuente: Weka

La Tabla 36 presenta los datos correspondientes a los 5 primeros clústers resultantes de la aplicación del algoritmo *KMeans* al conjunto de datos correspondientes al PC2.

Tabla 37 - Resultado de algoritmo *KMeans* PC2 (Parte 2)

	CLUSTER 5	CLUSTER 6	CLUSTER 7	CLUSTER 8	CLUSTER 9
<b>CAT_TOTAL</b>	BAJO	BAJO	ALTO	BAJO	BAJO
<b>ORIGEN_APREHENSION</b>	CONTROL DE RUTINA	CONTROL DE RUTINA	ACCIONES INTELIGENCIA Y PROTECCION - DIP	ACCIONES INTELIGENCIA Y PROTECCION - DIP	OPERATIVO CONJUNTO
<b>GRUPO_OPERATIVO</b>	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CONTROL CONJUNTO INTERINSTITUCIONAL
<b>ZONA</b>	ZONA1	ZONA5	ZONA1	ZONA2	ZONA5

Fuente: Weka

En la Tabla 37 que se aprecia anteriormente se presentan los datos correspondientes a los 5 clústers 5 al 9, resultantes de la aplicación del algoritmo *KMeans* al conjunto de datos correspondientes al PC2.

En el caso del **PC3** se puede identificar los siguientes resultados para cada uno de los clústers:

Tabla 38 - Resultado de algoritmo *KMeans* PC3 (Parte 1)

	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
<b>GRUPO</b>	TEXTILES	FRUTAS Y COMESTIBLES	LICORES	TEXTILES	FRUTAS Y COMESTIBLES
<b>SUBGRUPO</b>	PRENDAS DE VESTIR/NUEVAS	PRENDAS DE VESTIR/NUEVAS	WHISKY VARIAS MARCAS	MEDIAS DE VESTIR_ NYLON_ PANTYS_ TOBILLERAS	CEBOLLA
<b>ORIGEN_APREHENSION</b>	CONTROL DE RUTINA	ACTA DE ENTREGA - RECEPCION	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE RUTINA
<b>SITIO_APREHENSION</b>	CHACRAS-ECUADOR	PUERTO BOLIVAR-ECUADOR	CHACRAS-ECUADOR	CHACRAS-ECUADOR	CHACRAS-ECUADOR
<b>GRUPO_OPERATIVO</b>	CUERPO DE VIGILANCIA ADUANERA	POLICIA NACIONAL	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA

Fuente: Weka

Para el caso de la Tabla 38 se aprecia los datos correspondientes a los clústers 0 al 4, resultantes de la aplicación del algoritmo *KMeans* al conjunto de datos correspondientes al PC3.

Tabla 39 - Resultado de algoritmo *KMeans* PC3 (Parte 2)

	CLUSTER 5	CLUSTER 6	CLUSTER 7	CLUSTER 8	CLUSTER 9
<b>GRUPO</b>	CALZADO	CALZADO	ELECTRONICA Y SUS ACCESORIOS	ELECTRONICA Y SUS ACCESORIOS	TEXTILES
<b>SUBGRUPO</b>	CALZADO EN GENERAL	CALZADO EN GENERAL	CELULARES_ TELEFONOS_ FAXES_ HIPHONE	TELEVISORES	PRENDAS DE VESTIR/NUEVAS
<b>ORIGEN_APREHENSION</b>	CONTROL DE RUTINA	ALLANAMIENTO	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE RUTINA
<b>SITIO_APREHENSION</b>	CHACRAS-ECUADOR	GUAYAQUIL-ECUADOR	TULCAN-ECUADOR	RUMICHACA-ECUADOR	CANTON EL GUABO-ECUADOR
<b>GRUPO_OPERATIVO</b>	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA

Fuente: Weka

La Tabla 39 presenta los datos correspondientes a los clústers 5 al 9, resultantes de la aplicación del algoritmo *KMeans* al conjunto de datos correspondientes al PC2.

## Algoritmo EM

En la aplicación de las técnicas de clustering toman en cuenta como atributo clase a DISTRITO, por lo que este permite segmentar la información de una forma precisas para el posterior análisis y toma de decisiones.

```

==== Run information ====
Scheme:   weka.clusterers.EM -I 100 -N 10 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Relation: CAT_CUANTI_woutTEXTO_v09-weka.filters.unsupervised.attribute.NumericToNominal-R16
Instances: 102667
Attributes: 17
  ↳¿ANIO
  MES
  DIA
  CAT_CANTIDAD
  UNIDADES
  CAT_PRECIO
  CAT_TOTAL
  GRUPO
  SUBGRUPO
  PROCEDENCIA
  BODEGA
  STATUS
  ORIGEN_APREHENSION
  SITIO_APREHENSION
  GRUPO_OPERATIVO
  ZONA
Ignored:
  DISTRITO

```

Attribute	Cluster									
	0	1	2	3	4	5	6	7	8	9
	(0.09)	(0.06)	(0.08)	(0.06)	(0.15)	(0.03)	(0.06)	(0.19)	(0.08)	(0.19)
=====										
↳¿ANIO										
mean	2016.2241	2016.1262	2016.6687	2016.9708	2015.9705	2015.4676	2015.8759	2016.2345	2015.3659	2016.3761
std. dev.	1.3734	1.1708	1.3987	1.0784	1.5479	0.9362	1.6031	1.5719	0.5109	1.3899
MES										
mean	6.5615	7.0758	6.6712	6.1171	4.4728	8.4832	10.5213	5.9017	7.5783	6.8359
std. dev.	3.2611	3.1495	3.3632	3.6661	2.8807	3.0073	1.398	3.4446	2.7278	3.2914
DIA										
mean	19.1759	20.1572	20.3846	20.2727	18.2729	18.3898	19.7384	19.7828	19.3572	20.0472
std. dev.	9.272	8.0032	8.1015	8.1131	8.8905	8.4509	8.3274	8.7169	8.6936	9.102
CAT_CANTIDAD										
mean	1.9054	1.3158	2.7076	1.5326	1.7046	2.1259	1.4897	1.9881	1.8504	1.8944
std. dev.	0.9095	0.4965	1.0187	0.6698	0.7928	0.8527	0.6945	0.8171	0.9183	0.9407
UNIDADES										
mean	57.404	56.2774	52.7286	11.378	54.1042	57.9147	57.6233	57.4593	46.1123	50.3082
std. dev.	6.5866	6.3132	8.981	2.1124	8.2483	3.9195	4.1077	5.3076	18.2561	14.4471
CAT_PRECIO										
mean	1.1907	1.1835	1.0585	1.0002	1.0196	1.0912	1.0928	1.0005	1.1524	1.1437
std. dev.	0.5764	0.3942	0.2358	0.0123	0.1388	0.2878	0.3127	0.0217	0.3671	0.3676
CAT_TOTAL										
mean	1.6496	1.3541	1.9775	1.3411	1.2937	1.6368	1.3358	1.5088	1.6504	1.624
std. dev.	0.7813	0.5149	0.7206	0.5398	0.4868	0.6264	0.5056	0.6122	0.6432	0.6541
GRUPO										
mean	23.9121	7.944	5.6034	13.188	9.8551	26.0922	11.5381	25.9467	9.7285	8.7988
std. dev.	3.7418	3.1531	2.7543	3.6685	4.665	0.5413	4.0771	0.9686	4.8444	4.826
SUBGRUPO										
mean	181.4739	41.0276	26.9659	102.682	70.2407	190.0018	92.174	189.3555	63.6589	53.3531

std. dev.	14.9313	30.5101	14.1392	38.3872	49.3107	4.9672	47.1472	5.5405	48.9033	46.4536
<b>PROCEDENCIA</b>										
mean	8.259	8.9995	8.9959	9	8.4532	4.0188	9.1536	9	1.137	9
std. dev.	2.3086	0.0233	0.2533	2.5778	2.133	4.3253	0.5366	0	0.8689	2.5778
<b>BODEGA</b>										
mean	19.4428	19.1638	16.7854	15.8907	15.2014	14.5753	13.9301	19.8262	16.3176	29.2139
std. dev.	6.8302	6.9994	6.9975	5.363	2.7985	9.4694	2.558	7.208	11.1998	2.4418
<b>STATUS</b>										
mean	1	1	1	1.0005	1	1	1	1	1	1
std. dev.	0.007	0.007	0.007	0.0274	0.007	0.007	0.007	0.007	0.007	0.007
<b>ORIGEN_APREHENSION</b>										
mean	8.9011	9.8677	7.6569	7.8876	7.7864	7.0931	8.055	7.6367	7.5809	7.224
std. dev.	4.5705	5.5451	4.7146	3.1116	2.1547	2.9894	0.4852	3.0851	3.8646	4.5902
<b>SITIO_APREHENSION</b>										
mean	138.5484	126.8171	121.2139	86.439	70.8441	101.2275	71.0032	113.3466	131.9696	194.0116
std. dev.	73.9199	80.3088	66.88	45.8002	15.0995	61.0807	1.2564	67.7987	74.4341	64.3018
<b>GRUPO_OPERATIVO</b>										
mean	4.944	4.753	5.2358	5.1357	5.1262	5.1387	4.998	5.1326	5.241	5.4581
std. dev.	0.6206	0.5608	0.9679	0.7572	0.5653	0.6565	0.0584	0.6269	0.8744	1.0419
<b>ZONA</b>										
mean	4.6401	4.7061	3.8309	2.2415	2.006	1.9333	2	1.9952	2.1889	1.414
std. dev.	0.5973	0.4954	1.2341	0.7887	0.0806	0.4635	1.29	0.5889	1.1674	0.6013
Time taken to build model (full training data) : 75.01 seconds										

Figura 25 - Resultados algoritmo EM

Los resultados arrojados con el algoritmo EM permiten realizar un análisis estadístico de las variables categorizadas con el propósito de identificar las similitudes entre los clústers además permite conocer si los datos son variados a razón de la desviación estándar.

Tabla 40 – Resultado de algoritmo EM (Parte 1)

ATRIBUTOS	CLUSTER									
	0		1		2		3		4	
	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar
AÑO	2016,22	1,37	2016,13	1,17	2016,67	1,40	2016,97	1,08	2015,97	1,55
MES	6,56	3,26	7,08	3,15	6,67	3,36	6,12	3,67	4,47	2,88
DIA	19,18	0,93	20,16	8,00	20,38	8,10	20,27	8,11	18,27	8,89
CAT_CANTIDAD	1,91	0,91	1,32	0,50	2,71	10187,00	1,53	0,67	1,70	0,79
UNIDADES	57,40	6,59	56,28	6,31	52,73	8,98	11,38	2,11	54,10	8,25
CAT_PRECIO	1,19	0,58	1,18	0,39	1,06	0,24	1,00	0,01	1,02	0,14
CAT_TOTAL	1,65	0,78	1,35	0,51	1,98	0,72	1,34	0,54	1,29	0,49
GRUPO	23,91	3,74	7,94	3,15	5,60	2,75	13,19	3,67	9,86	4,67
SUBGRUPO	181,47	14,93	41,03	30,51	26,97	14,14	102,68	38,39	70,24	49,31
PROCEDENCIA	8,26	2,31	9,00	0,02	9,00	0,25	9,00	2,58	8,45	2,13
BODEGA	19,44	6,83	19,16	7,00	16,79	7,00	15,89	5,36	15,20	2,80
STATUS	1,00	0,01	1,00	0,01	1,00	0,01	1,00	0,03	1,00	0,01
ORIGEN_APREHENSION	8,90	4,57	9,87	5,55	7,66	4,71	7,89	3,11	7,79	2,15
GRUPO_OPERATIVO	4,94	0,62	4,75	0,56	5,24	0,97	5,14	0,76	5,13	0,57
ZONA	4,64	0,60	4,71	0,50	3,83	1,23	2,24	0,79	2,01	0,08

Fuente: Weka

Tabla 41 - Resultado de algoritmo EM (Parte 2)

ATRIBUTOS	CLUSTER									
	5		6		7		8		9	
	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar
AÑO	2015,47	0,94	2015,88	1,60	2016,23	1,57	2015,37	0,51	2016,38	1,39
MES	8,48	3,01	10,52	1,40	5,90	3,44	7,58	2,73	6,84	3,29
DIA	18,39	8,45	19,74	8,33	19,78	8,72	19,36	8,69	20,05	0,91
CAT_CANTIDAD	2,13	0,85	1,49	0,69	1,99	0,82	1,85	0,92	1,89	0,94
UNIDADES	57,91	3,92	57,62	4,11	57,46	5,31	46,11	18,26	50,31	14,45
CAT_PRECIO	1,09	0,29	1,09	0,31	1,00	0,02	1,15	0,37	1,14	0,37
CAT_TOTAL	1,64	0,63	1,34	0,51	1,51	0,61	1,65	0,64	1,62	0,65
GRUPO	26,09	0,54	11,54	4,08	25,95	0,97	9,73	4,84	8,80	4,83
SUBGRUPO	190,00	4,97	92,17	47,15	189,36	5,54	63,66	48,90	53,35	46,45
PROCEDENDIA	4,02	4,33	9,15	0,54	9,00	0,00	1,14	0,87	9,00	2,58
BODEGA	14,58	9,47	13,93	2,56	19,83	7,21	16,32	11,20	29,21	2,44
STATUS	1,00	0,01	1,00	0,01	1,00	0,01	1,00	0,01	1,00	0,01
ORIGEN_APREHENSION	7,09	2,99	8,06	0,49	7,64	3,09	7,58	3,86	7,22	4,90
GRUPO_OPERATIVO	5,14	0,66	5,00	0,06	5,13	0,63	5,24	0,87	5,46	1,04
ZONA	1,93	0,46	2,00	1,29	2,00	0,59	2,19	1,17	1,41	0,60

Fuente: Weka

### Análisis e interpretación

El análisis de los datos se los realizó por cada atributo presente en el conjunto de datos para lograr identificar la similitud presente en los datos de un clúster con respecto a los datos de otro clúster. Se utilizó dos decimales para la aproximación al inmediato superior en los casos necesario y así proceder con la desnormalización de los datos para una interpretación más certera. De acuerdo con la Tabla 39 y Tabla 40 se obtuvo los siguientes resultados:

#### AÑO

En este atributo los clústers 4,5,6, y 8 poseen la categoría de 2015 con los valores de su desviación estándar muy similares y no muy altos, mientras que el resto almacenan la categoría 2016 y de igual manera su desviación estándar es muy similares entendiendo que hay poca variabilidad de datos de este atributo.

#### MES

Para este atributo los clústers 0,2,3 y 9 almacenan la categoría Junio con una desviación estándar por encima de 3 respectivamente, entendiéndose que los datos de cada clúster son variados, mientras que para los clústers 1 y 8 poseen la categoría Julio con los valores de su desviación estándar diferentes, de igual manera que los casos anteriores se considera que los datos de los clústers son muy variados. Finalmente, los clústers restantes almacenan las categorías abril, agosto, octubre y mayo respectivamente.

#### **DIA**

En este atributo los clústers 0,6,7 y 8 almacenan en su categoría el día 19 del mes con el caso particular que para el clúster 0 su desviación estándar es menor a 1 mientras que para los clústers restantes es superior a 8 entendiéndose que los datos registrados son muy variados, además para los clústers 1,2,3 y 9 poseen en su categoría el día 20 del mes igualmente se presenta que uno de los clústers poseen una desviación estándar por debajo de 1 mientras que los demás tiene su desviación superior a 8 que igualmente permite entender que los datos de estos clústers son muy variados, finalmente los clústers restantes corresponden al día 18 del mes.

#### **CAT\_CANTIDAD**

Para este atributo los clústers 0,1,3,4,6,8 y 9 presentan en sus datos la categoría Unidades con los valores sus desviaciones estándar muy similares y bajas por lo que se entiende que los datos no son muy variados, mientras que para los clústers 2,5 y 7 almacenan la categoría Decenas con sus desviaciones estándar muy similares y acercándose a 1 entendiéndose así que los registros son similares.

#### **UNIDADES**

En el caso de este atributo los clústers 0,6 y 7 almacenan en sus datos la categoría tonelada presentando su desviación estándar con valores altos entendiéndose de esta manera que los datos son muy variados, mientras que para el resto de clústers registro muy variados como tira, tableta, bulto, quintal y saco respectivamente igualmente con valores variados y mayores a 1 en su desviación estándar dando a entender que los datos registrados son muy variados.

#### **CAT\_PRECIO**

En este caso todos los clústers poseen la categoría de Bajo en los rangos definidos para categorizar el precio, la desviación estándar de todos los clústers se encuentra por debajo de 1 entendiéndose que los datos de los clústers son similares.

#### **CAT\_TOTAL**

Para este atributo en los clústers 0,2,5 y 8 almacenan la categoría Medio en los rangos definidos para categorizar el total con una desviación estándar por debajo de 1 dando a

entender que no existes mucha variedad en sus datos, mientras que para los clústers 1,3,4,6 y 7 poseen la categoría Bajo con desviaciones estándar de igualmente por debajo de 1 entendiéndose que los datos no tienen mucha variación.

#### **GRUPO**

En este atributo los clústers 1,4,8 y 9 se almacenan la categoría de Electrónica y sus Accesorios con sus valores de desviación estándar superiores a 3 dando a entender que los datos son sumamente variados mientras que para los clústers 5 y 7 la categoría de Textiles con su desviación estándar de 0.54 y 0.97 respectivamente, entendiéndose que los datos de estos clústers son similares. Finalmente para los restantes se presentan las categorías Confeitería, Licores y Frutas y Comestibles con valores de desviación estándar superiores a 2 entendiéndose que los datos registramos son variados.

#### **SUBGRUPO**

Este atributo almacena datos totalmente variados con las categorías puesto que para clústers se presenta una categoría diferente para el clúster 0 se tiene como resultados De Vehículos, 1 con Acc Celular, 2 con Dulces Varios, 3 con Toronja, 4 con Manzanas, 5 con Medias de vestir\_Nylon\_Pantys\_Tobilleras, 6 con Abono, 7 con Algodón\_Hilos\_Elasticos\_Encajes, 8 con Aguacate y por último el clúster 9 con Fósforos.

#### **PROCEDENCIA**

Para los clústers 1,2,3,6,7 y 9 los clústers almacenan la categoría Extranjera con desviaciones estándar muy variadas desde de 0 hasta mayor de 2 lo que da a entender que existe mucha variedad de datos, mientras que para los clústers 0 y 4 corresponde a España y finalmente para el 5 y 8 corresponde a China y Desconocida con desviaciones estándar igualmente como en los anterior se presentan datos completamente variados.

#### **STATUS**

En este atributo todos los clústers almacenan la categoría nueva con una desviación estándar mínima de 0,01 por lo que se entiende que los datos tienen datos prácticamente similares.

### **ORIGEN\_APREHENSION**

Para este atributo los clústers 0,2,3,4,6,7 y 8 almacenan la categoría Control del Rutina con sus desviaciones estándar con un valor superior a 4 respectivamente para cada caso por lo que se entiende que los datos son completamente variados, mientras que para los clústers 5 y 9 corresponde a Control de Ingreso de Divisas y para el restante a Control de rutina con sus desviaciones estándar correspondientes muy altas que de igual manera dan a entender que los datos son muy variados.

### **GRUPO\_OPERATIVO**

En este caso los clústers del 2 al 9 poseen toda la categoría Cuerpo de Vigilancia Aduanera con los valores de desviación estándar para cada uno de ellos que no exceden en su mayoría a 1 entendiéndose que los registros de estos clústers son similares, mientras que para los clústers restantes 0 y 1 almacenan la categoría de Control Conjunto Interinstitucional con desviación estantes por debajo de 1 que muestra que los datos de igual manera son similares.

### **ZONA**

Para este atributo los clústers 3,4,6,7 y 8 almacena la categoría de Zona 2 con los valores para su desviación estándar muy similares entendiéndose así que sus registros son similares mientras que para los clústers 5 y 9 corresponden a la categoría Zona1, los clústers 0 y 1 a Zona4 y finalmente el clústers 2 con la categoría de Zona 3 igualmente con los valores de desviación estándar muy similares entendiéndose que los datos de los clústers para este atributo son similares.

## Clústeres finales del Algoritmo *EM*

Tabla 42 – Clústeres desnormalizados del algoritmo *EM* (Parte 1)

	CLUSTER	CLUSTER	CLUSTER	CLUSTER	CLUSTER
	0	1	2	3	4
AÑO	2016	2016	2016	2016	2015
MES	JUNIO	JULIO	JUNIO	JUNIO	ABRIL
DIA	19	20	20	20	18
CAT_CANTIDAD	UNIDADES	UNIDADES	DECENAS	UNIDADES	UNIDADES
UNIDADES	57,40	56,28	52,73	11,38	54,10
CAT_PRECIO	BAJO	BAJO	BAJO	BAJO	BAJO
CAT_TOTAL	MEDIO	BAJO	MEDIO	BAJO	BAJO



GRUPO	REPUESTOS Y ACCESORIOS	ELECTRONICA Y SUS ACCESORIOS	CONFITERIA	LICORES	ELECTRONICA Y SUS ACCESORIOS
SUBGRUPO	DE VEHICULOS	ACC CELULAR	DULCES VARIOS	TORONJA	MANZANAS
PROCEDENCIA	ESPAÑA	EXTRANJERA	EXTRANJERA	EXTRANJERA	ESPAÑA
BODEGA	DESTACAMENTO SAUCILLO	DESTACAMENTO SAUCILLO	DESTACAMENTO EL EMPALME	DESTACAMENTO CHACRAS	DESTACAMENTO CATAMAYO
STATUS	NUEVO	NUEVO	NUEVO	NUEVO	NUEVO
ORIGEN_APREHENSION	CONTROL DE RUTINA	CONTROL EN CARRETERA	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE RUTINA
GRUPO_OPERATIVO	CUERPO DE VIGILANCIA ADUANERA	CONTROL CONJUNTO INTERINSTITUCIONAL	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA
ZONA	ZONA4	ZONA4	ZONA 3	ZONA2	ZONA2

Fuente: Weka

Tabla 43 - Clústeres desnormalizados del algoritmo EM (Parte 2)

	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
AÑO	2015	2015	2016	2015	2016
MES	AGOSTO	OCTUBRE	MAYO	JULIO	JUNIO
DIA	18	19	19	19	20
CAT_CANTIDAD UNIDADES	DECENAS	UNIDADES	DECENAS	UNIDADES	UNIDADES
CAT_PRECIO	57,91	57,62	57,46	46,11	50,31
CAT_TOTAL	BAJO	BAJO	BAJO	BAJO	BAJO
GRUPO	MEDIO	BAJO	BAJO	MEDIO	MEDIO
SUBGRUPO	TEXTILES	FRUTAS Y COMESTIBLES	TEXTILES	ELECTRONICA Y SUS ACCESORIOS	ELECTRONICA Y SUS ACCESORIOS
PROCEDENCIA	MEDIAS DE VESTIR/NYLON/ PANTYS/ TOBILLERAS	ABONO	ALGODON/HILOS/ ELASTICOS/ENCAJES	AGUACATE	FOSFOROS
BODEGA	CHINA	EXTRANJERA	EXTRANJERA	DESCONOCIDA	EXTRANJERA
STATUS	DESTACAMENTO CATAMAYO	DESTACAMENTO AMAZONAS	DESTACAMENTO TABACUNDO	DESTACAMENTO CHACRAS	DIRECCION DISTRITAL QUITO
ORIGEN_APREHENSION	NUEVO	NUEVO	NUEVO	NUEVO	NUEVO
GRUPO_OPERATIVO	CONTROL DE INGRESO DE DIVISAS	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE INGRESO DE DIVISAS
ZONA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA
	ZONA1	ZONA2	ZONA2	ZONA2	ZONA1

Fuente: Weka

Según los resultados del software se asignó un clúster a cada uno de los distritos ubicados en el territorio nacional como son: Distrito Latacunga, Distrito Quito, Distrito Guayas, Distrito Loja, Distrito Huaquillas, Distrito Azuay, Distrito Puerto Bolívar, Distrito Esmeraldas, Distrito Carchi respectivamente.

Con el propósito de obtener una mayor cantidad de patrones que permitan una mejor interpretación y análisis de la información con la que se trabaja, se aplicó a cada PC obtenido en la sección 2.7.4 el algoritmo EM con los mismos criterios que la sección anterior.

Para el caso del PC1 se obtuvo los resultados que se detallan en la Tabla 44 y Tabla 45 que se muestran a continuación:

Tabla 44 - Resultado de algoritmo *EM* conjunto de datos PC1 (Parte 1)

ATRIBUTOS	CLUSTER									
	0		1		2		3		4	
	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar
GRUPO	25,77	1,40	7,63	1,95	14,34	2,77	10,51	1,55	5,21	2,14
SUBGRUPO	192,64	10,17	30,72	6,12	116,63	39,03	51,38	17,20	23,89	8,31
BODEGA	20,06	8,25	15,28	7,66	23,51	7,92	17,79	6,42	27,98	4,06
SITIO_APREHENSION	118,09	65,90	143,59	76,70	174,96	67,60	67,67	37,69	200,08	43,00
GRUPO_OPERATIVO	5,24	0,92	4,93	0,27	5,52	1,12	5,15	1,07	6,88	0,41

Fuente: Weka

La Tabla 42 presenta los valores correspondientes a cada uno de los atributos pertenecientes a los 5 primeros clústers (0-4) resultantes de la aplicación del algoritmo *EM* al conjunto de datos correspondientes al PC1.

Tabla 45 - Resultado de algoritmo *EM* conjunto de datos PC1 (Parte 2)

ATRIBUTOS	CLUSTER									
	5		6		7		8		9	
	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar
GRUPO	13,50	1,92	6,89	2,49	3,76	0,93	26,08	0,28	25,97	0,29
SUBGRUPO	114,11	28,89	28,66	8,74	18,71	5,71	189,20	3,08	188,69	2,21
BODEGA	14,65	4,23	29,88	1,53	16,89	7,22	16,57	6,42	24,93	6,54
SITIO_APREHENSION	71,21	7,57	222,31	32,83	83,98	41,88	76,28	29,47	214,77	34,47
GRUPO_OPERATIVO	5,00	0,00	4,85	0,37	5,00	0,67	5,03	0,51	5,17	0,71

Fuente: Weka

En Tabla 45 se aprecia los valores correspondientes a cada uno de los atributos pertenecientes los clústers 5 al 9, resultantes de la aplicación del algoritmo *EM* al conjunto de datos correspondientes al PC1.

### Evaluación de resultados PC1

#### GRUPO

Para este atributo los clústers 0,8 y 9 presenta valores muy similares para el grupo “textiles”, con una desviación estándar de 1.40 para el clúster 0 y una desviación estándar medidas muy similares de 0.28 y 0.29 respectivamente.

#### SUBGRUPO

En el caso de este atributo en los clústers 8 y 9 se encuentra la categoría correspondiente a "textiles" mientras que el para el resto de clústers posee datos y con una desviación estándar igualmente variada.

#### **BODEGA**

Para este atributo los clústers 7 y 8 se encuentra la categoría correspondiente a “Destacamento Chacras” con desviaciones estándar similares, para el resto de clústers poseen datos variados del clúster 0 al 6 respectivamente Destacamento Tabacundo, Destacamento Catamayo, Dirección Distrital Esmeraldas, Destacamento El Empalme, Dirección Distrital Manabí, Destacamento Amazonas, Dirección Distrital Quito y para el último clúster la categoría Dirección Distrital Guayas.

#### **GRUPO\_OPERATIVO**

Este atributo en los clústers 0, 2, 3, 5, 7, 8, 9 se encuentra la categoría Cuerpo de Vigilancia Aduanera con desviaciones estándar bajas que quiere que decir que sus datos no son muy variados, para el clúster 1 y 6 se tiene como resultado la categoría Control Conjunto Interinstitucional de igual forma con una desviación estándar baja y finalmente el clúster 4 los datos corresponden a Fuerzas Armadas.

Con los datos obtenidos en la evaluación de resultados correspondiente al PC1 se puede identificar los siguientes registros desnormalizados de los clústers siendo estos los datos correspondientes para cada uno de los resultados obtenidos correspondiente a cada campo y clúster respectivo acorde al proceso de normalización de la información realizado en la sección 2.6.2.

Tabla 46 - Clústeres desnormalizados del algoritmo *EM* para el PC1 (Parte 1)

ATRIBUTO	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
GRUPO	TEXTILES	COSMETICOS	MADERA	EXPLOSIVOS	CIGARRILOS
SUBGRUPO	TELAS	PERFUMES	NARANJILLA	ENCENDEDORES Y SUS ACCESORIOS	DEPORTIVO
BODEGA	DESTACAMENTO TABACUNDO	DESTACAMENTO CATAMAYO	DIRECCION DISTRITAL GUAYAS	DESTACAMENTO MAÑOSCA	DIRECCION DISTRITAL MANABI
SITIO_APREHENSION	GUANO-ECUADOR	LA UNION - ECUADOR	NOBOL-ECUADOR	CATAMAYO ECUADOR	PINDAL-ECUADOR
GRUPO_OPERATIVO	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	FUERZAS ARMADAS	CUERPO DE VIGILANCIA ADUANERA	POLICIA NACIONAL

Fuente: Weka

Tabla 47 - Clústeres desnormalizados del algoritmo *EM* para el PC1 (Parte 2)

	CLUSTER 5	CLUSTER 6	CLUSTER 7	CLUSTER 8	CLUSTER 9
GRUPO	LICORES	CONFITERIA	CALZADO	TEXTILES	TEXTILES
SUBGRUPO	PIÑAS	GALLETAS	MATERIAL DIDACTICO KILOS	ALGODON/HILOS/ ELASTICOS/ENCAJES	ALGODON/HILOS/ ELASTICOS/ENCAJES
BODEGA	DESTACAMENTO CATAMAYO	DIRECCION DISTRITAL TULCAN	DESTACAMENTO EL EMPALME	DESTACAMENTO EL EMPALME	DIRECCION DISTRITAL LATACUNGA
SITIO_APREHENSION	CHACRAS-ECUADOR	RIO CONEJO - ECUADOR	COTACACHI- ECUADOR	CHAUCHIN-ECUADOR	PUYO - ECUADOR
GRUPO_OPERATIVO	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA

Fuente: Weka

Para el caso de PC2 se obtuvo los resultados que se detallan en la Tabla 48 y Tabla 49 que se muestran a continuación:

Tabla 48 - Resultado de algoritmo *EM* conjunto de datos PC2 (Parte 1)

ATRIBUTOS	CLUSTER									
	0		1		2		3		4	
	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar
CAT_TOTAL	1,48	0,63	2,16	0,47	1,81	0,69	1,51	0,62	1,37	0,49
ORIGEN_APREHENSION	18,04	0,34	2,76	0,44	4,75	3,19	8,00	3,83	11,10	2,76
GRUPO_OPERATIVO	4,03	0,18	6,30	1,00	5,60	1,46	5,00	0,00	4,72	0,46
ZONA	1,49	0,57	1,65	0,55	4,75	0,53	2,28	1,05	1,74	0,55

Fuente: Weka

En la Tabla 48 se aprecia los valores correspondientes a cada uno de los atributos pertenecientes los clústers 0 al 4, resultantes de la aplicación del algoritmo *EM* al conjunto de datos correspondientes al PC2.

Tabla 49 - Resultado de algoritmo *EM* conjunto de datos PC2 (Parte 2)

ATRIBUTOS	CLUSTER									
	5		6		7		8		9	
	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar
CAT_TOTAL	1,33	0,47	2,02	0,38	2,87	0,53	1,44	0,63	1,01	0,12
ORIGEN_APREHENSION	7,57	0,49	9,37	0,45	9,39	5,03	18,00	0,07	2,83	0,38
GRUPO_OPERATIVO	5,00	0,00	5,00	0,00	4,72	0,65	4,00	0,00	6,45	0,88
ZONA	4,60	0,50	1,81	0,47	3,42	1,47	4,86	0,35	1,71	0,54

Fuente: Weka

La Tabla 49 presenta los valores correspondientes a cada uno de los atributos pertenecientes a los clústers 5 al 9, resultantes de la aplicación del algoritmo *EM* al conjunto de datos correspondientes al PC2.

### Evaluación de resultados PC2

#### **CAT\_TOTAL**

Para este atributo los clústers 1,2,6 y 7 poseen la categoría Medio con una desviación estándar con valores bajos mientras que el resto de clústers presentan la categoría Bajo con desviaciones estándar con valores no muy altos.

#### **ORIGEN\_APREHENSION**

Para este atributo se identifica que los clústers 1 y 9 almacenan la categoría correspondiente a Acta de Entrega-Recepción con valores bajos para la desviación estándar, mientras que los clúster 6 y 7 corresponden a la categoría denominada como Control en Carretera con desviaciones estándar muy diferentes confirmación que los datos son variados, para los casos de los clústers 0 y 8 se tiene registros de valores muy bajos en desviación estándar para Operativo Conjunto, al contrario que el caso anterior confirmando de esta forma que los registros de ambos clúster son similares, finalmente para los clústers faltantes cada uno poseen atributos muy variados.

#### **GRUPO\_OPERATIVO**

En este atributo los clústers 0,4,7y 8 almacenan la categoría Control Conjunto Interinstitucional con valores muy bajos para la desviación estándar, mientras que los clústers 2,3,5 y 6 presentan la categoría Cuerpo de Vigilancia Aduanera con valores cero en su desviación estándar entendiéndose que los datos de estos clústers no son variados y finalmente los clústers 1 y 9 corresponden a la categoría de Fuerzas Armadas con un valor bajo para sus desviaciones estándar correspondiente significando que hay poca variación en sus datos.

#### **ZONA**

Para este atributo los clústers 0,1,3,4,6 y 9 tiene registrado en este atributo a la categoría Zona2 con valores bajos y con poca diferencia para su desviación estándar por lo que se entiende que sus datos no son muy variados, mientras que para los clústers 2, 5 y 8 corresponde a la categoría Zonas5 con valores similares en su desviación estándar para los dos primeros a diferencia del ultimo clúster mencionado.

Con los datos obtenidos en la evaluación de resultados correspondiente al PC2 se puede identificar los siguientes registros desnormalizados para cada uno de los clústers.

Tabla 50 - Clústeres desnormalizados del algoritmo *EM* para el PC2 (Parte 1)

ATRIBUTO	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
CAT_TOTAL	BAJO	MEDIO	BAJO	BAJO	BAJO
ORIGEN_APREHENSION	OPERATIVO CONJUNTO	ACTA DE ENTREGA - RECEPCION	CONTROL AEROPUERTO	CONTROL DE RUTINA	CONTROL MOVIL
GRUPO_OPERATIVO	CONTROL CONJUNTO INTERINSTITUCIONAL	FUERZAS ARMADAS	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CONTROL CONJUNTO INTERINSTITUCIONAL
ZONA	ZONA2	ZONA2	ZONA5	ZONA2	ZONA2

Fuente: Weka

Tabla 51 - Clústeres desnormalizados del algoritmo *EM* para el PC2 (Parte 2)

ATRIBUTO	CLUSTER 5	CLUSTER 6	CLUSTER 7	CLUSTER 8	CLUSTER 9
CAT_TOTAL	BAJO	MEDIO	MEDIO	BAJO	BAJO
ORIGEN_APREHENSION	CONTROL DE RUTINA	CONTROL EN CARRETERA	CONTROL EN CARRETERA	OPERATIVO CONJUNTO	ACTA DE ENTREGA - RECEPCION
GRUPO_OPERATIVO	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CONTROL CONJUNTO INTERINSTITUCIONAL	CONTROL CONJUNTO INTERINSTITUCIONAL	FUERZAS ARMADAS
ZONA	ZONA5	ZONA2	ZONA3	ZONA5	ZONA2

Fuente: Weka

Para el caso de **PC3** se obtuvo los resultados que se detallan en la Tabla 52 y Tabla 53 que se muestran a continuación:

Tabla 52 -Resultado de algoritmo *EM* conjunto de datos PC3 (Parte 1)

ATRIBUTOS	CLUSTER									
	0		1		2		3		4	
	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar
GRUPO	10,53	1,60	26,11	0,44	13,86	1,87	24,21	3,32	24,06	3,39
SUBGRUPO	51,93	18,10	189,87	4,48	125,66	18,17	183,17	12,84	182,65	12,57
ORIGEN_APREHENSION	7,78	2,11	7,89	1,89	7,99	1,03	7,90	2,74	15,44	3,95
SITIO_APREHENSION	74,62	30,28	77,44	30,20	73,88	19,12	209,82	41,26	114,79	82,53
GRUPO_OPERATIVO	4,97	0,54	5,00	0,00	4,99	0,12	5,00	0,00	4,11	0,49

Fuente: Weka

La Tabla 52 presenta los valores correspondientes a cada uno de los atributos de los primeros clústers (0-4) resultantes de la aplicación del algoritmo *EM* al conjunto de datos correspondientes al PC3.

Tabla 53 - Resultado de algoritmo *EM* conjunto de datos PC3 (Parte 2)

ATRIBUTOS	CLUSTER									
	5		6		7		8		9	
	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar	Media	Desviación Estándar
GRUPO	3,91	1,05	7,91	3,22	8,08	3,83	24,11	3,44	7,80	3,30
SUBGRUPO	19,37	5,80	40,80	29,34	47,16	33,18	183,31	13,37	40,37	28,60
ORIGEN_APREHENSION	7,31	2,54	8,02	2,79	2,99	0,13	2,82	0,42	18,04	0,33
SITIO_APREHENSION	85,14	37,75	220,65	30,87	166,14	65,31	150,41	68,36	114,70	93,42
GRUPO_OPERATIVO	4,96	0,29	4,98	0,16	6,90	0,38	6,39	0,98	4,03	0,16

Fuente: Weka

En la Tabla 53 se aprecia los valores correspondientes a cada uno de los atributos pertenecientes los clústers 5 al 9, resultantes de la aplicación del algoritmo *EM* al conjunto de datos correspondientes al PC2.

### **Evaluación de resultados PC3**

#### **GRUPO**

Para este atributo los clústers 6 y 9 poseen la categoría Cosméticos con una desviación estándar alta pero muy similar por lo que se entiende que los datos son variados, mientras que para los clústers 3,4 y 8 registran la categoría Repuestos y Accesorios igualmente como en caso anterior su desviación estándar es muy similar pero considerable para entender que los datos son muy variados. Para el resto de clústers los datos son muy variados.

#### **SUBGRUPO**

En este atributo los clústers 6 y 9 presentan la categoría Reproductores De Sonido con los valores de desviación estándar muy similares pero altos representando una gran variación de los registros, mientras que para los clústers 3,4 y 8 almacenan la categoría correspondiente a Repuestos Y Accesorios De Motos con una desviación estándar superior a 12 entendiéndose que los datos de este atributo son muy variados. En el caso del resto de clústers se encuentra almacenado diferentes subgrupos y con altos valores en su desviación estándar por lo que los registros son muy variados.

#### **ORIGEN\_APREHENSION**

Para este atributo los clústers 0,1,2,3 y 6 almacenan la categoría corresponde a Control de Rutina con los valores de desviación estándar parecidas pero mayores a 1 entendiéndose que se presenta cierta variación en lo datos registrados, para los clústers 7 y 8 se registran la categoría Acta de Entrega – Recepción con una desviación estándar por debajo de 1 dando a entender que los datos no se los aprecia con una gran variación. Para los clústers restantes las categorías presentadas corresponden a Control de Ingreso de Divisas, Inspección de Mercancía, Operativo conjunto respectivamente.

## 🚦 SITIO\_APREHENSION

En este caso los clústers 4 y 9 presentan la categoría Gonzalo Pizarro-Ecuador con los valores de desviación estándar sumamente altos por lo que se entiende que los datos de este atributo son altamente variados, mientras que para el resto de los clústers los datos almacenados son muy variados.

## 🚦 GRUPO\_OPERATIVO

Para este atributo los clústers 0,1,3,2,5,6 y almacenan la categoría de Cuerpo de Vigilancia Aduanera acorde a los valores de desviación estándar que se encuentran desde 0 hasta por encima de 0.50 se entiende que los datos de cada clúster en este atributo no son muy variados, mientras que para los clústers 4 y 9 la categoría corresponde a Control Conjunto Interinstitucional y los restantes a Fuerzas Armadas

Con los datos obtenidos en la evaluación de resultados correspondiente al PC3 se puede identificar los siguientes registros desnormalizados para cada uno de los clústers apreciados en la Tabla 54 y Tabla 55.

Tabla 54 - Clústeres desnormalizados del algoritmo EM para el PC3 (Parte 1)

	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
GRUPO	EXPLOSIVOS	TEXTILES	LICORES	REPUESTOS Y ACCESORIOS	REPUESTOS Y ACCESORIOS
SUBGRUPO	ENCENDEDORES Y SUS ACCESORIOS	ALGODON/HILOS/ ELASTICOS/ENCAJES	BRANDY	DE MOTOS	DE MOTOS
ORIGEN_APREHENSION	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE RUTINA	INSPECCION DE MERCANCIA
SITIO_APREHENSION	CHASQUI- ECUADOR	CHITAN DE NAVARRTE-ECUADOR	CHAPUEL- ECUADOR	PUERTO MARITIMO- ECUADOR	GONZANAMA- ECUADOR
GRUPO_OPERATIVO	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CUERPO DE VIGILANCIA ADUANERA	CONTROL CONJUNTO INTERINSTITUCIONAL

Fuente: Weka

Tabla 55 - Desnormalizados del algoritmo EM para el PC3 (Parte 1)

	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
GRUPO	EXPLOSIVOS	TEXTILES	LICORES	REPUESTOS Y ACCESORIOS	REPUESTOS Y ACCESORIOS
SUBGRUPO	ENCENDEDORES Y SUS ACCESORIOS	ALGODON/HILOS/ ELASTICOS/ENCAJES	BRANDY	DE MOTOS	DE MOTOS
ORIGEN_APREHENSION	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE RUTINA	CONTROL DE RUTINA	INSPECCION DE MERCANCIA
SITIO_APREHENSION	CHASQUI- ECUADOR	CHITAN DE NAVARRTE-ECUADOR	CHAPUEL- ECUADOR	PUERTO MARITIMO- ECUADOR	GONZANAMA- ECUADOR



	CUERPO DE VIGILANCIA	CUERPO DE VIGILANCIA	CUERPO DE VIGILANCIA	CUERPO DE VIGILANCIA	CONTROL CONJUNTO INTERINSTITUCIONAL
GRUPO_OPERATIVO	ADUANERA	ADUANERA	ADUANERA	ADUANERA	

Fuente: Weka

### 3.1.2. Evaluación, análisis e interpretación de tareas de Asociación

#### Algoritmo Apriori

Aplicación del algoritmo *Apriori* en el conjunto de datos correspondiente a 17 variables y 102667 instancias.

Figura 26 - Resultados algoritmo *apriori* para el conjunto de datos total

```

=== Run information ===

Scheme: weka.associations.Apriori -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: DATAWARE_HOUSEv09-weka.filters.unsupervised.attribute.Remove-R7-
weka.filters.unsupervised.attribute.NumericToNominal-R1-3-weka.filters.unsupervised.attribute.Remove-R10,13
Instances: 102667
Attributes: 17
  >¿ANIO
  MES
  DIA
  CAT_CANTIDAD
  UNIDADES
  CAT_PRECIO
  CAT_TOTAL
  GRUPO
  SUBGRUPO
  PROCEDENCIA
  BODEGA
  STATUS
  ORIGEN_APREHENSION
  SITIO_APREHENSION
  GRUPO_OPERATIVO
  DISTRITO
  ZONA

=== Associator model (full training set) ===
Apriori
=====
Minimum support: 0.6 (61600 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 8

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6
Size of set of large itemsets L(2): 11
Size of set of large itemsets L(3): 7
Size of set of large itemsets L(4): 1

Best rules found:

1. UNIDADES=UNIDAD CAT_PRECIO=BAJO 65827 ==> STATUS=BUENO 65827 <conf:(1)> lift:(1) lev:(0) [1] conv:(1.28)
2. CAT_PRECIO=BAJO 94947 ==> STATUS=BUENO 94946 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.92)
3. CAT_PRECIO=BAJO PROCEDENCIA=EXTRANJERA 82387 ==> STATUS=BUENO 82386 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.8)
4. UNIDADES=UNIDAD 73092 ==> STATUS=BUENO 73091 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.71)
5. CAT_PRECIO=BAJO GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 72018 ==> STATUS=BUENO 72017
<conf:(1)> lift:(1) lev:(0) [0] conv:(0.7)
6. UNIDADES=UNIDAD PROCEDENCIA=EXTRANJERA 63594 ==> STATUS=BUENO 63593 <conf:(1)> lift:(1) lev:(0) [0]
conv:(0.62)
7. CAT_PRECIO=BAJO PROCEDENCIA=EXTRANJERA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 61810
==> STATUS=BUENO 61809 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.6)
8. PROCEDENCIA=EXTRANJERA 88801 ==> STATUS=BUENO 88799 <conf:(1)> lift:(1) lev:(-0) [0] conv:(0.58)

```

9. GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 78485 ==> STATUS=BUENO 78483 <conf:(1)> lift:(1) lev:(-0) [0] conv:(0.51)  
 10. PROCEDENCIA=EXTRANJERA GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 67196 ==> STATUS=BUENO 67194 <conf:(1)> lift:(1) lev:(-0) [0] conv:(0.44)  
 11. ORIGEN\_APREHENSION=CONTROL DE RUTINA 63523 ==> STATUS=BUENO 63521 <conf:(1)> lift:(1) lev:(-0) [0] conv:(0.41)  
 12. ORIGEN\_APREHENSION=CONTROL DE RUTINA GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 62317 ==> STATUS=BUENO 62315 <conf:(1)> lift:(1) lev:(-0) [0] conv:(0.4)  
 13. ORIGEN\_APREHENSION=CONTROL DE RUTINA 63523 ==> GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 62317 <conf:(0.98)> lift:(1.28) lev:(0.13) [13756] conv:(12.4)  
 14. STATUS=BUENO ORIGEN\_APREHENSION=CONTROL DE RUTINA 63521 ==> GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 62315 <conf:(0.98)> lift:(1.28) lev:(0.13) [13755] conv:(12.4)  
 15. ORIGEN\_APREHENSION=CONTROL DE RUTINA 63523 ==> STATUS=BUENO GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 62315 <conf:(0.98)> lift:(1.28) lev:(0.13) [13755] conv:(12.38)  
 16. PROCEDENCIA=EXTRANJERA STATUS=BUENO 88799 ==> CAT\_PRECIO=BAJO 82386 <conf:(0.93)> lift:(1) lev:(0) [264] conv:(1.04)  
 17. PROCEDENCIA=EXTRANJERA 88801 ==> CAT\_PRECIO=BAJO 82387 <conf:(0.93)> lift:(1) lev:(0) [263] conv:(1.04)  
 18. PROCEDENCIA=EXTRANJERA 88801 ==> CAT\_PRECIO=BAJO STATUS=BUENO 82386 <conf:(0.93)> lift:(1) lev:(0) [263] conv:(1.04)  
 19. STATUS=BUENO 102665 ==> CAT\_PRECIO=BAJO 94946 <conf:(0.92)> lift:(1) lev:(0) [0] conv:(1)  
 20. PROCEDENCIA=EXTRANJERA STATUS=BUENO GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 67194 ==> CAT\_PRECIO=BAJO 61809 <conf:(0.92)> lift:(0.99) lev:(-0) [-332] conv:(0.94)

Fuente: Weka

### ✓ Evaluación resultados de algoritmo apriori conjunto de datos totales

Para el caso de conjunto de datos completos se aprecian las mejores 20 reglas de asociación resultado de la aplicación del algoritmo *apriori* en la Tabla 56. Para el cálculo de los de soporte y confianza se debe tener en cuenta su fórmula que es:

$$confianza(x \Rightarrow y) = \frac{soporte(x)}{soporte(y)} \quad Ec.3$$

Tabla 56 - Reglas de asociación, soporte y confianza algoritmo *Apriori* conjunto de datos completos

N.REGLA	ÍTEMS EN X	SOPORTE (x)	ÍTEMS EN Y	SOPORTE (x∩y)	CONFIANZA
1	UNIDADES=UNIDAD CAT_PRECIO=BAJO	65827	STATUS=BUENO	65827	1,0000
2	CAT_PRECIO=BAJO	94947	STATUS=BUENO	94946	1,0000
3	CAT_PRECIO=BAJO PROCEDENCIA=EXTRANJERA	82387	STATUS=BUENO	82386	1,0000
4	UNIDADES=UNIDAD	73092	STATUS=BUENO	73091	1,0000
5	CAT_PRECIO=BAJO GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	72018	STATUS=BUENO	72017	1,0000
6	UNIDADES=UNIDAD PROCEDENCIA=EXTRANJERA	63594	STATUS=BUENO	63593	1,0000
7	CAT_PRECIO=BAJO PROCEDENCIA=EXTRANJERA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	61810	STATUS=BUENO	61809	1,0000
8	PROCEDENCIA=EXTRANJERA	88801	STATUS=BUENO	88799	1,0000
9	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	78485	STATUS=BUENO	78483	1,0000
10	PROCEDENCIA=EXTRANJERA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	67196	STATUS=BUENO	67194	1,0000
11	ORIGEN_APREHENSION=CONTROL DE RUTINA	63523	STATUS=BUENO	63521	1,0000

12	ORIGEN_APREHENSION=CONTROL DE RUTINA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	62317	STATUS=BUENO	62315	1,0000
13	ORIGEN_APREHENSION=CONTROL DE RUTINA	63523	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	62317	0,9810
14	STATUS=BUENO ORIGEN_APREHENSION=CONTROL DE RUTINA	63521	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	62315	0,9810
15	ORIGEN_APREHENSION=CONTROL DE RUTINA	63523	STATUS=BUENO GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	62315	0,9810
16	PROCEDENCIA=EXTRANJERA STATUS=BUENO	88799	CAT_PRECIO=BAJO	82386	0,9278
17	PROCEDENCIA=EXTRANJERA	88801	CAT_PRECIO=BAJO	82387	0,9278
18	PROCEDENCIA=EXTRANJERA	88801	CAT_PRECIO=BAJO STATUS=BUENO	82386	0,9278
19	STATUS=BUENO	102665	CAT_PRECIO=BAJO	94946	0,9248
20	PROCEDENCIA=EXTRANJERA STATUS=BUENO GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	67194	CAT_PRECIO=BAJO	61809	0,9199

Fuente: Weka

Con el propósito de obtener resultados más específicos se procedió a la aplicación del algoritmo *apriori* a los conjuntos de datos sugeridos por el algoritmo PCA sección 2.7.4. En las figuras 27, 28 y 29 se presentan los resultados obtenidos para PC1, PC2 y PC3 respectivamente.

Figura 27 – Resultado algoritmo *Apriori* conjunto de datos PC1

```

=== Run information ===

Scheme:   weka.associations.Apriori -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: DATAWARE_HOUSEv09-weka.filters.unsupervised.attribute.Remove-R1-8,11-12,14-16,19-20
Instances: 102667
Attributes: 5
          GRUPO
          SUBGRUPO
          BODEGA
          SITIO_APREHENSION
          GRUPO_OPERATIVO
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (10267 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11
Size of set of large itemsets L(2): 10
Size of set of large itemsets L(3): 4

Best rules found:

1. SUBGRUPO=PRENDAS DE VESTIR/NUEVAS 21377 ==> GRUPO=TEXTILES 21377 <conf:(1)> lift:(4.13) lev:(0.16) [16201]
   conv:(16201.56)
2. SUBGRUPO=PRENDAS DE VESTIR/NUEVAS GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 17907 ==>
   GRUPO=TEXTILES 17907 <conf:(1)> lift:(4.13) lev:(0.13) [13571] conv:(13571.66)
3. GRUPO=LICORES SITIO_APREHENSION=CHACRAS-ECUADOR 10997 ==> GRUPO_OPERATIVO=CUERPO DE
   VIGILANCIA ADUANERA 10475 <conf:(0.95)> lift:(1.25) lev:(0.02) [2068] conv:(4.95)

```

4.	BODEGA=DESTACAMENTO CHACRAS GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 26009 ==> SITIO_APREHENSION=CHACRAS-ECUADOR 24268 <conf:(0.93)> lift:(2.36) lev:(0.14) [13965] conv:(9.02)
5.	BODEGA=BODEGA HUAQUILLAS SITIO_APREHENSION=CHACRAS-ECUADOR 12319 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 11431 <conf:(0.93)> lift:(1.21) lev:(0.02) [2013] conv:(3.26)
6.	BODEGA=BODEGA HUAQUILLAS GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 12320 ==> SITIO_APREHENSION=CHACRAS-ECUADOR 11431 <conf:(0.93)> lift:(2.34) lev:(0.06) [6550] conv:(8.36)
7.	SITIO_APREHENSION=CHACRAS-ECUADOR 40668 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 37274 <conf:(0.92)> lift:(1.2) lev:(0.06) [6184] conv:(2.82)
8.	BODEGA=DESTACAMENTO CHACRAS SITIO_APREHENSION=CHACRAS-ECUADOR 26677 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 24268 <conf:(0.91)> lift:(1.19) lev:(0.04) [3874] conv:(2.61)
9.	BODEGA=DESTACAMENTO CHACRAS 29389 ==> SITIO_APREHENSION=CHACRAS-ECUADOR 26677 <conf:(0.91)> lift:(2.29) lev:(0.15) [15035] conv:(6.54)

Fuente: Weka

Figura 28 - Resultado algoritmo *Apriori* conjunto de datos PC2

```

=== Run information ===
Scheme: weka.associations.Apriori -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: DATAWARE_HOUSEv09-weka.filters.unsupervised.attribute.Remove-R1-7,9-15,17
Instances: 102667
Attributes: 5
    CAT_TOTAL
    ORIGEN_APREHENSION
    GRUPO_OPERATIVO
    DISTRITO
    ZONA
=== Associator model (full training set) ===
Apriori
=====
Minimum support: 0.2 (20533 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 12

Size of set of large itemsets L(3): 11

Size of set of large itemsets L(4): 5

Size of set of large itemsets L(5): 1

Best rules found:

1. DISTRITO=IXDH 44147 ==> ZONA=ZONA2 44147 <conf:(1)> lift:(1.79) lev:(0.19) [19534] conv:(19534.98)
2. GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH 39263 ==> ZONA=ZONA2 39263 <conf:(1)> lift:(1.79) lev:(0.17) [17373] conv:(17373.82)
3. ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH 36438 ==> ZONA=ZONA2 36438 <conf:(1)> lift:(1.79) lev:(0.16) [16123] conv:(16123.76)
4. ORIGEN_APREHENSION=CONTROL DE RUTINA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH 36136 ==> ZONA=ZONA2 36136 <conf:(1)> lift:(1.79) lev:(0.16) [15990] conv:(15990.13)
5. CAT_TOTAL=BAJO DISTRITO=IXDH 27460 ==> ZONA=ZONA2 27460 <conf:(1)> lift:(1.79) lev:(0.12) [12151] conv:(12151.01)
6. CAT_TOTAL=BAJO GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH 24874 ==> ZONA=ZONA2 24874 <conf:(1)> lift:(1.79) lev:(0.11) [11006] conv:(11006.71)
7. CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH 23009 ==> ZONA=ZONA2 23009 <conf:(1)> lift:(1.79) lev:(0.1) [10181] conv:(10181.45)
8. CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH 22882 ==> ZONA=ZONA2 22882 <conf:(1)> lift:(1.79) lev:(0.1) [10125] conv:(10125.25)
9. CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA ZONA=ZONA2 26299 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 26162 <conf:(0.99)> lift:(1.3) lev:(0.06) [6057] conv:(44.89)
10. CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH 23009 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 22882 <conf:(0.99)> lift:(1.3) lev:(0.05) [5292] conv:(42.34)
11. CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH ZONA=ZONA2 23009 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 22882 <conf:(0.99)> lift:(1.3) lev:(0.05) [5292] conv:(42.34)
12. CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH 23009 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA ZONA=ZONA2 22882 <conf:(0.99)> lift:(2.12) lev:(0.12) [12081] conv:(95.38)
13. ORIGEN_APREHENSION=CONTROL DE RUTINA ZONA=ZONA2 42317 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 41974 <conf:(0.99)> lift:(1.3) lev:(0.09) [9624] conv:(28.97)

```

14.	ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH 36438 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 36136 <conf:(0.99)> lift:(1.3) lev:(0.08) [8280] conv:(28.33)
15.	ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH ZONA=ZONA2 36438 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 36136 <conf:(0.99)> lift:(1.3) lev:(0.08) [8280] conv:(28.33)
16.	ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH 36438 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA ZONA=ZONA2 36136 <conf:(0.99)> lift:(2.11) lev:(0.19) [19031] conv:(63.81)
17.	CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA 34931 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 34387 <conf:(0.98)> lift:(1.29) lev:(0.07) [7683] conv:(15.1)
18.	ORIGEN_APREHENSION=CONTROL DE RUTINA 63523 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 62317 <conf:(0.98)> lift:(1.28) lev:(0.13) [13756] conv:(12.4)
19.	CAT_TOTAL=MEDIO ORIGEN_APREHENSION=CONTROL DE RUTINA 24548 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 24033 <conf:(0.98)> lift:(1.28) lev:(0.05) [5266] conv:(11.21)
20.	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH 39263 ==> ORIGEN_APREHENSION=CONTROL DE RUTINA 36136 <conf:(0.92)> lift:(1.49) lev:(0.12) [11842] conv:(4.79)

Fuente: Weka

Figura 29 - Resultado algoritmo *Apriori* conjunto de datos PC3

```

=== Run information ===
Scheme: weka.associations.Apriori -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: DATAWARE_HOUSEv09-weka.filters.unsupervised.attribute.Remove-R1-8,11-15,19-20
Instances: 102667
Attributes: 5
    GRUPO
    SUBGRUPO
    ORIGEN_APREHENSION
    SITIO_APREHENSION
    GRUPO_OPERATIVO
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (10267 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10
Size of set of large itemsets L(2): 12
Size of set of large itemsets L(3): 7
Size of set of large itemsets L(4): 1

Best rules found:
1. SUBGRUPO=PRENDAS DE VESTIR/NUEVAS 21377 ==> GRUPO=TEXTILES 21377 <conf:(1)> lift:(4.13) lev:(0.16) [16201]
   conv:(16201.56)
2. SUBGRUPO=PRENDAS DE VESTIR/NUEVAS GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 17907 ==>
   GRUPO=TEXTILES 17907 <conf:(1)> lift:(4.13) lev:(0.13) [13571] conv:(13571.66)
3. SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA 13847 ==>
   GRUPO=TEXTILES 13847 <conf:(1)> lift:(4.13) lev:(0.1) [10494] conv:(10494.6)
4. SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA
   GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 13710 ==> GRUPO=TEXTILES 13710 <conf:(1)> lift:(4.13)
   lev:(0.1) [10390] conv:(10390.77)
5. ORIGEN_APREHENSION=CONTROL DE RUTINA SITIO_APREHENSION=CHACRAS-ECUADOR 34925 ==>
   GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 34872 <conf:(1)> lift:(1.31) lev:(0.08) [8173] conv:(152.34)
6. GRUPO_OPERATIVO=POLICIA NACIONAL 12974 ==> ORIGEN_APREHENSION=ACTA DE ENTREGA - RECEPCION 12893
   <conf:(0.99)> lift:(6.72) lev:(0.11) [10974] conv:(134.83)
7. GRUPO=LICORES ORIGEN_APREHENSION=CONTROL DE RUTINA 11737 ==> GRUPO_OPERATIVO=CUERPO DE
   VIGILANCIA ADUANERA 11657 <conf:(0.99)> lift:(1.3) lev:(0.03) [2684] conv:(34.13)
8. SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA 13847 ==>
   GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 13710 <conf:(0.99)> lift:(1.3) lev:(0.03) [3124] conv:(23.63)
9. GRUPO=TEXTILES SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA 13847
   ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 13710 <conf:(0.99)> lift:(1.3) lev:(0.03) [3124] conv:(23.63)
10. SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA 13847 ==>
   GRUPO=TEXTILES GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 13710 <conf:(0.99)> lift:(4.9) lev:(0.11)
   [10912] conv:(80.07)
11. GRUPO=TEXTILES ORIGEN_APREHENSION=CONTROL DE RUTINA 15783 ==> GRUPO_OPERATIVO=CUERPO DE
   VIGILANCIA ADUANERA 15616 <conf:(0.99)> lift:(1.29) lev:(0.03) [3550] conv:(22.13)

```

12.	ORIGEN_APREHENSION=CONTROL DE RUTINA 63523 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 62317 <conf:(0.98)> lift:(1.28) lev:(0.13) [13756] conv:(12.4)
13.	GRUPO=LICORES SITIO_APREHENSION=CHACRAS-ECUADOR 10997 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 10475 <conf:(0.95)> lift:(1.25) lev:(0.02) [2068] conv:(4.95)
14.	SITIO_APREHENSION=CHACRAS-ECUADOR GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 37274 ==> ORIGEN_APREHENSION=CONTROL DE RUTINA 34872 <conf:(0.94)> lift:(1.51) lev:(0.12) [11809] conv:(5.91)
15.	GRUPO=LICORES GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 12718 ==> ORIGEN_APREHENSION=CONTROL DE RUTINA 11657 <conf:(0.92)> lift:(1.48) lev:(0.04) [3788] conv:(4.57)
16.	SITIO_APREHENSION=CHACRAS-ECUADOR 40668 ==> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA 37274 <conf:(0.92)> lift:(1.2) lev:(0.06) [6184] conv:(2.82)

Fuente: Weka

En las Tablas 57, 58, y 59 a continuación se presenta las reglas de asociación con los valores correspondientes a soporte y confianza para cada uno de los conjuntos de datos reducidos resultado de la aplicación del algoritmo PCA (sección 2.7.4).

Tabla 57 - Reglas de asociación, soporte y confianza resultado del algoritmo *Apriori* PC1

N. REGLA	ÍTEMS EN X	SOPORTE (x)	ÍTEMS EN Y	SOPORTE (x∩y)	CONFIANZA
1	SUBGRUPO=PRENDAS DE VESTIR/NUEVAS	21377	GRUPO=TEXTILES	21377	1,0000
2	SUBGRUPO=PRENDAS DE VESTIR/NUEVAS GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	17907	GRUPO=TEXTILES	17907	1,0000
3	GRUPO=LICORES SITIO_APREHENSION=CHACRAS-ECUADOR	10997	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	10475	0,9525
4	BODEGA=DESTACAMENTO CHACRAS GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	26009	SITIO_APREHENSION=CHACRAS-ECUADOR	24268	0,9331
5	BODEGA=BODEGA HUAQUILLAS SITIO_APREHENSION=CHACRAS-ECUADOR	12319	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	11431	0,9279
6	BODEGA=BODEGA HUAQUILLAS GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	12320	SITIO_APREHENSION=CHACRAS-ECUADOR	11431	0,9278
7	SITIO_APREHENSION=CHACRAS-ECUADOR	40668	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	37274	0,9165
8	BODEGA=DESTACAMENTO CHACRAS SITIO_APREHENSION=CHACRAS-ECUADOR	26677	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	24268	0,9097
9	BODEGA=DESTACAMENTO CHACRAS	29389	SITIO_APREHENSION=CHACRAS-ECUADOR	26677	0,9077

Fuente: Weka

En el caso particular de la información con la que se trabaja se evidencia que los datos no se encuentran balanceados, por lo que varios de los ítems con categorías específicas se presentan de manera predominante como para el caso de las reglas resultado del PC1 se tiene: GRUPO\_OPERATIVO con su categoría “CUERPO DE VIGILANCIA ADUANERA”.

Tabla 58 - Reglas de asociación, soporte y confianza resultado del algoritmo *Apriori* PC2

N. REGLA	ÍTEMS EN X	SOPORTE (x)	ÍTEMS EN Y	SOPORTE (x∩y)	CONFIANZA
1	DISTRITO=IXDH	44147	ZONA=ZONA2	44147	1,0000
2	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH	39263	ZONA=ZONA2	39263	1,0000

3	ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH	36438	ZONA=ZONA2	36438	1,0000
4	ORIGEN_APREHENSION=CONTROL DE RUTINA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH	36136	ZONA=ZONA2	36136	1,0000
5	CAT_TOTAL=BAJO DISTRITO=IXDH	27460	ZONA=ZONA2	27460	1,0000
6	CAT_TOTAL=BAJO GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH	24874	ZONA=ZONA2	24874	1,0000
7	CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH	23009	ZONA=ZONA2	23009	1,0000
8	CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH	22882	ZONA=ZONA2	22882	1,0000
9	CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA ZONA=ZONA2	26299	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	26162	0,9948
10	CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH	23009	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	22882	0,9945
11	CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH ZONA=ZONA2	23009	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	22882	0,9945
12	CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH	23009	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA ZONA=ZONA2	22882	0,9945
13	ORIGEN_APREHENSION=CONTROL DE RUTINA ZONA=ZONA2	42317	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	41974	0,9919
14	ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH	36438	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	36136	0,9917
15	ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH ZONA=ZONA2	36438	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	36136	0,9917
16	ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH	36438	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA ZONA=ZONA2	36136	0,9917
17	CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA	34931	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	34387	0,9844
18	ORIGEN_APREHENSION=CONTROL DE RUTINA	63523	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	62317	0,9810
19	CAT_TOTAL=MEDIO ORIGEN_APREHENSION=CONTROL DE RUTINA	24548	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	24033	0,9790
20	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH	39263	ORIGEN_APREHENSION=CONTROL DE RUTINA	36136	0,9204

Fuente: Weka

Al igual que en caso de la tabla anterior que puede constatar que el balance de la información influye en los resultados para el PC2 por lo que ORIGEN\_APREHENSION con su categoría “CONTROL DE RUTINA”, es la que predomina como también en ZONA con su categoría “ZONA2” y GRUPO\_OPERATIVO con su categoría “CUERPO DE VIGILANCIA ADUANERA” entre otros.

Tabla 59 - Reglas de asociación, soporte y confianza resultado del algoritmo *Apriori* PC3

N.REGLA	ÍTEMS EN X	SOPORTE (x)	ÍTEMS EN Y	SOPORTE (x∩y)	CONFIANZA
1	SUBGRUPO=PRENDAS DE VESTIR/NUEVAS	21377	GRUPO=TEXTILES	21377	1,0000
2	SUBGRUPO=PRENDAS DE VESTIR/NUEVAS GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	17907	GRUPO=TEXTILES	17907	1,0000
3	SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA	13847	GRUPO=TEXTILES	13847	1,0000
4	SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	13710	GRUPO=TEXTILES	13710	1,0000
5	ORIGEN_APREHENSION=CONTROL DE RUTINA SITIO_APREHENSION=CHACRAS-ECUADOR	34925	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	34872	0,9985
6	GRUPO_OPERATIVO=POLICIA NACIONAL	12974	ORIGEN_APREHENSION=ACTA DE ENTREGA - RECEPCION	12893	0,9938
7	GRUPO=LICORES ORIGEN_APREHENSION=CONTROL DE RUTINA	11737	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	11657	0,9932
8	SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA	13847	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	13710	0,9901
9	GRUPO=TEXTILES SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA	13847	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	13710	0,9901
10	SUBGRUPO=PRENDAS DE VESTIR/NUEVAS ORIGEN_APREHENSION=CONTROL DE RUTINA	13847	GRUPO=TEXTILES GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	13710	0,9901
11	GRUPO=TEXTILES ORIGEN_APREHENSION=CONTROL DE RUTINA	15783	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	15616	0,9894
12	ORIGEN_APREHENSION=CONTROL DE RUTINA	63523	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	62317	0,9810
13	GRUPO=LICORES SITIO_APREHENSION=CHACRAS-ECUADOR	10997	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	10475	0,9525
14	SITIO_APREHENSION=CHACRAS-ECUADOR GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	37274	ORIGEN_APREHENSION=CONTROL DE RUTINA	34872	0,9356
15	GRUPO=LICORES GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	12718	ORIGEN_APREHENSION=CONTROL DE RUTINA	11657	0,9166
16	SITIO_APREHENSION=CHACRAS-ECUADOR	40668	GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	37274	0,9165

Fuente: Weka

De igual forma en la tabla anterior se constata que no hay un balance de la información en los resultados para el PC3 por lo que GRUPO con su categoría “TEXTILES”, es la que predomina como también en GRUPO\_OPERATIVO con su categoría “CUERPO DE VIGILANCIA ADUANERA” y entre otros.



Con los resultados obtenidos en las Tablas 57, 58 y 59 se seleccionaron las reglas más relevantes evitando redundancia de las reglas dando como resultado las siguientes reglas de la Tabla 60 como las de mayor interés:

Tabla 60 – Listado de mejores reglas de asociación con el algoritmo *Apriori*

ÍTEMS EN X	ÍTEMS EN Y
SUBGRUPO=PRENDAS DE VESTIR/NUEVAS GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	=> GRUPO=TEXTILES
GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH	=> ZONA=ZONA2
ORIGEN_APREHENSION=CONTROL DE RUTINA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH	=> ZONA=ZONA2
CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA DISTRITO=IXDH	=> ZONA=ZONA2
SUBGRUPO=PRENDAS DE VESTIR/NUEVAS	=> GRUPO=TEXTILES
ORIGEN_APREHENSION=CONTROL DE RUTINA GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
ORIGEN_APREHENSION=CONTROL DE RUTINA SITIO_APREHENSION=CHACRAS-ECUADOR	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA ZONA=ZONA2	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH ZONA=ZONA2	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
GRUPO_OPERATIVO=POLICIA NACIONAL	=> ORIGEN_APREHENSION=ACTA DE ENTREGA - RECEPCION
GRUPO=LICORES ORIGEN_APREHENSION=CONTROL DE RUTINA	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
ORIGEN_APREHENSION=CONTROL DE RUTINA DISTRITO=IXDH ZONA=ZONA2	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
GRUPO=TEXTILES ORIGEN_APREHENSION=CONTROL DE RUTINA	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
CAT_TOTAL=BAJO ORIGEN_APREHENSION=CONTROL DE RUTINA	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
ORIGEN_APREHENSION=CONTROL DE RUTINA	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
CAT_TOTAL=MEDIO ORIGEN_APREHENSION=CONTROL DE RUTINA	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA
GRUPO=LICORES SITIO_APREHENSION=CHACRAS-ECUADOR	=> GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA

Fuente: Adaptación de resultados de Weka

## Algoritmo FP Growth

Aplicación del algoritmo *FP Growth* en el conjunto de datos correspondiente a 17 variables binarizadas con un total de 744 atributos y 102667 instancias.

Figura 30 - Resultados algoritmo FP Growth para el conjunto de datos total

```

=== Run information ===

Scheme:   weka.associations.FPGrowth -P 2 -I -1 -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation: DATAWARE_HOUSEv09-weka.filters.unsupervised.attribute.Remove-R11,14-weka.filters.unsupervised.attribute.Remove-R7-
weka.filters.unsupervised.attribute.{0,1}ToNominal-R1-3-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6-
weka.filters.unsupervised.attribute.NominalToBinary-A-Rfirst-last-weka.filters.unsupervised.attribute.NominalToBinary-A-Rfirst-last-
weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-weka.filters.unsupervised.attribute.NumericToBinary-Rfirst-last
Instances: 102667
Attributes: 744

```

[list of attributes omitted]  
 === Associator model (full training set) ===

FPGrowth found 28 rules (displaying top 20)

1. [CAT\_PRECIO=BAJO=1\_binarized=1, UNIDADES=UNIDAD=1\_binarized=1]: 65827 ==> [STATUS=BUENO=1\_binarized=1]: 65827 <conf:(1)> lift:(1) lev:(0) conv:(1.28)
2. [CAT\_PRECIO=BAJO=1\_binarized=1]: 94947 ==> [STATUS=BUENO=1\_binarized=1]: 94946 <conf:(1)> lift:(1) lev:(0) conv:(0.92)
3. [CAT\_PRECIO=BAJO=1\_binarized=1, PROCEDENCIA=EXTRANJERA=1\_binarized=1]: 82387 ==> [STATUS=BUENO=1\_binarized=1]: 82386 <conf:(1)> lift:(1) lev:(0) conv:(0.8)
4. [UNIDADES=UNIDAD=1\_binarized=1]: 73092 ==> [STATUS=BUENO=1\_binarized=1]: 73091 <conf:(1)> lift:(1) lev:(0) conv:(0.71)
5. [CAT\_PRECIO=BAJO=1\_binarized=1, GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA=1\_binarized=1]: 72018 ==> [STATUS=BUENO=1\_binarized=1]: 72017 <conf:(1)> lift:(1) lev:(0) conv:(0.7)
6. [PROCEDENCIA=EXTRANJERA=1\_binarized=1, UNIDADES=UNIDAD=1\_binarized=1]: 63594 ==> [STATUS=BUENO=1\_binarized=1]: 63593 <conf:(1)> lift:(1) lev:(0) conv:(0.62)
7. [CAT\_PRECIO=BAJO=1\_binarized=1, PROCEDENCIA=EXTRANJERA=1\_binarized=1, GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA=1\_binarized=1]: 61810 ==> [STATUS=BUENO=1\_binarized=1]: 61809 <conf:(1)> lift:(1) lev:(0) conv:(0.6)
8. [PROCEDENCIA=EXTRANJERA=1\_binarized=1]: 88801 ==> [STATUS=BUENO=1\_binarized=1]: 88799 <conf:(1)> lift:(1) lev:(-0) conv:(0.58)
9. [GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA=1\_binarized=1]: 78485 ==> [STATUS=BUENO=1\_binarized=1]: 78483 <conf:(1)> lift:(1) lev:(-0) conv:(0.51)
10. [PROCEDENCIA=EXTRANJERA=1\_binarized=1, GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA=1\_binarized=1]: 67196 ==> [STATUS=BUENO=1\_binarized=1]: 67194 <conf:(1)> lift:(1) lev:(-0) conv:(0.44)
11. [ORIGEN\_APREHENSION=CONTROL DE RUTINA=1\_binarized=1]: 63523 ==> [STATUS=BUENO=1\_binarized=1]: 63521 <conf:(1)> lift:(1) lev:(-0) conv:(0.41)
12. [GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA=1\_binarized=1, ORIGEN\_APREHENSION=CONTROL DE RUTINA=1\_binarized=1]: 62317 ==> [STATUS=BUENO=1\_binarized=1]: 62315 <conf:(1)> lift:(1) lev:(-0) conv:(0.4)
13. [ORIGEN\_APREHENSION=CONTROL DE RUTINA=1\_binarized=1]: 63523 ==> [GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA=1\_binarized=1]: 62317 <conf:(0.98)> lift:(1.28) lev:(0.13) conv:(12.4)
14. [STATUS=BUENO=1\_binarized=1, ORIGEN\_APREHENSION=CONTROL DE RUTINA=1\_binarized=1]: 63521 ==> [GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA=1\_binarized=1]: 62315 <conf:(0.98)> lift:(1.28) lev:(0.13) conv:(12.4)
15. [ORIGEN\_APREHENSION=CONTROL DE RUTINA=1\_binarized=1]: 63523 ==> [STATUS=BUENO=1\_binarized=1, GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA=1\_binarized=1]: 62315 <conf:(0.98)> lift:(1.28) lev:(0.13) conv:(12.38)
16. [STATUS=BUENO=1\_binarized=1, PROCEDENCIA=EXTRANJERA=1\_binarized=1]: 88799 ==> [CAT\_PRECIO=BAJO=1\_binarized=1]: 82386 <conf:(0.93)> lift:(1) lev:(0) conv:(1.04)
17. [PROCEDENCIA=EXTRANJERA=1\_binarized=1]: 88801 ==> [CAT\_PRECIO=BAJO=1\_binarized=1]: 82387 <conf:(0.93)> lift:(1) lev:(0) conv:(1.04)
18. [PROCEDENCIA=EXTRANJERA=1\_binarized=1]: 88801 ==> [STATUS=BUENO=1\_binarized=1, CAT\_PRECIO=BAJO=1\_binarized=1]: 82386 <conf:(0.93)> lift:(1) lev:(0) conv:(1.04)
19. [STATUS=BUENO=1\_binarized=1]: 102665 ==> [CAT\_PRECIO=BAJO=1\_binarized=1]: 94946 <conf:(0.92)> lift:(1) lev:(0) conv:(1)
20. [STATUS=BUENO=1\_binarized=1, PROCEDENCIA=EXTRANJERA=1\_binarized=1, GRUPO\_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA=1\_binarized=1]: 67194 ==> [CAT\_PRECIO=BAJO=1\_binarized=1]: 61809 <conf:(0.92)> lift:(0.99) lev:(-0) conv:(0.94)

Fuente: Weka

En las tablas a continuación se presenta las reglas de asociación obtenidas mediante el algoritmo *FP Growth* con los valores correspondientes a soporte y confianza.

Tabla 61 - Reglas de asociación, soporte y confianza resultado del algoritmo *FP Growth* conjunto de datos completos

N. REGLA	ÍTEMS EN X	SOPORTE (x)	ÍTEMS EN Y	SOPORTE (x∩y)	CONFIANZA A
1	[CAT_PRECIO=BAJO, UNIDADES=UNIDAD]	65827	[STATUS=BUENO]	65827	1,0000
2	[CAT_PRECIO=BAJO]	94947	[STATUS=BUENO]	94946	1,0000
3	[CAT_PRECIO=BAJO, PROCEDENCIA=EXTRANJERA]	82387	[STATUS=BUENO]	82386	1,0000
4	[UNIDADES=UNIDAD]	73092	[STATUS=BUENO]	73091	1,0000
5	[CAT_PRECIO=BAJO, GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	72018	[STATUS=BUENO]	72017	1,0000
6	[PROCEDENCIA=EXTRANJERA, UNIDADES=UNIDAD]	63594	[STATUS=BUENO]	63593	1,0000
7	[CAT_PRECIO=BAJO, PROCEDENCIA=EXTRANJERA, GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	61810	[STATUS=BUENO]	61809	1,0000

8	[PROCEDENCIA=EXTRANJERA]	88801	[STATUS=BUENO]	88799	1,0000
9	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	78485	[STATUS=BUENO]	78483	1,0000
10	[PROCEDENCIA=EXTRANJERA, GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	67196	[STATUS=BUENO]	67194	1,0000
11	[ORIGEN_APREHENSION=CONTROL DE RUTINA]	63523	[STATUS=BUENO]	63521	1,0000
12	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, ORIGEN_APREHENSION=CONTROL DE RUTINA]	62317	[STATUS=BUENO]	62315	1,0000
13	[ORIGEN_APREHENSION=CONTROL DE RUTINA]	63523	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	62317	0,9810
14	[STATUS=BUENO, ORIGEN_APREHENSION=CONTROL DE RUTINA]	63521	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	62315	0,9810
15	[ORIGEN_APREHENSION=CONTROL DE RUTINA]	63523	[STATUS=BUENO, GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	62315	0,9810
16	[STATUS=BUENO, PROCEDENCIA=EXTRANJERA]	88799	[CAT_PRECIO=BAJO]	82386	0,9278
17	[PROCEDENCIA=EXTRANJERA]	88801	[CAT_PRECIO=BAJO]	82387	0,9278
18	[PROCEDENCIA=EXTRANJERA]	88801	[STATUS=BUENO, CAT_PRECIO=BAJO]	82386	0,9278
19	[STATUS=BUENO]	102665	[CAT_PRECIO=BAJO]	94946	0,9248
20	[STATUS=BUENO, PROCEDENCIA=EXTRANJERA, GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	67194	[CAT_PRECIO=BAJO]	61809	0,9199

Fuente: Weka

Al igual que con el algoritmo anterior con el propósito de obtener resultados más específicos se procedió a la aplicación del algoritmo *FP Growth* a los conjuntos de datos sugeridos por el algoritmo PCA 2.7.4. En las figuras 28, 29 y 30 se presentan los resultados obtenidos para PC1, PC2 y PC3 respectivamente.

En conjunto de datos correspondiente al PC1 se obtuvo un total de 575 atributos binarizados e igualmente se trabajó con las 102667 instancias disponibles como se visualiza en la Fig. 31.

Figura 31 - Resultado algoritmo *FP Growth* conjunto de datos PC1

```

=== Run information ===
Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation: DATAWARE_HOUSEv09-weka.filters.unsupervised.attribute.Remove-R1-8,11-12,14-16,19-20-
weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-
weka.filters.unsupervised.attribute.NumericToBinary-Rfirst-last-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-
weka.filters.unsupervised.attribute.NumericToBinary-Rfirst-last-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-
weka.filters.unsupervised.attribute.NumericToBinary-Rfirst-last
Instances: 102667
Attributes: 575
[list of attributes omitted]
=== Associator model (full training set) ===

FPGrowth found 9 rules (displaying top 9)

1. [SUBGRUPO=PRENDAS DE VESTIR/NUEVAS=1_binarized=1_binarized=1_binarized=1]: 21377 ==>
[GRUPO=TEXTILES=1_binarized=1_binarized=1_binarized=1]: 21377 <conf:(1)> lift:(4.13) lev:(0.16) conv:(16201.56)
2. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, SUBGRUPO=PRENDAS DE
VESTIR/NUEVAS=1_binarized=1_binarized=1_binarized=1]: 17907 ==> [GRUPO=TEXTILES=1_binarized=1_binarized=1_binarized=1]:
17907 <conf:(1)> lift:(4.13) lev:(0.13) conv:(13571.66)
3. [SITIO_APREHENSION=CHACRAS-ECUADOR=1_binarized=1_binarized=1_binarized=1,
GRUPO=LICORES=1_binarized=1_binarized=1_binarized=1]: 10997 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA
ADUANERA_binarized=1]: 10475 <conf:(0.95)> lift:(1.25) lev:(0.02) conv:(4.95)

```

4.	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, BODEGA=DESTACAMENTO CHACRAS=1_binarized=1_binarized=1_binarized=1]: 26009 ==> [SITIO_APREHENSION=CHACRAS-ECUADOR=1_binarized=1_binarized=1_binarized=1]: 24268 <conf:(0.93)> lift:(2.36) lev:(0.14) conv:(9.02)
5.	[SITIO_APREHENSION=CHACRAS-ECUADOR=1_binarized=1_binarized=1_binarized=1, BODEGA=BODEGA HUAQUILLAS=1_binarized=1_binarized=1_binarized=1]: 12319 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 11431 <conf:(0.93)> lift:(1.21) lev:(0.02) conv:(3.26)
6.	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, BODEGA=BODEGA HUAQUILLAS=1_binarized=1_binarized=1_binarized=1]: 12320 ==> [SITIO_APREHENSION=CHACRAS-ECUADOR=1_binarized=1_binarized=1_binarized=1]: 11431 <conf:(0.93)> lift:(2.34) lev:(0.06) conv:(8.36)
7.	[SITIO_APREHENSION=CHACRAS-ECUADOR=1_binarized=1_binarized=1_binarized=1]: 40668 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 37274 <conf:(0.92)> lift:(1.2) lev:(0.06) conv:(2.82)
8.	[SITIO_APREHENSION=CHACRAS-ECUADOR=1_binarized=1_binarized=1_binarized=1, BODEGA=DESTACAMENTO CHACRAS=1_binarized=1_binarized=1_binarized=1]: 26677 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 24268 <conf:(0.91)> lift:(1.19) lev:(0.04) conv:(2.61)
9.	[BODEGA=DESTACAMENTO CHACRAS=1_binarized=1_binarized=1_binarized=1]: 29389 ==> [SITIO_APREHENSION=CHACRAS-ECUADOR=1_binarized=1_binarized=1_binarized=1]: 26677 <conf:(0.91)> lift:(2.29) lev:(0.15) conv:(6.54)

Fuente: Weka

Para el conjunto de datos correspondiente al PC2 se obtuvo un total de 45 atributos binarizados e igualmente se empleó 102667 instancias disponibles que se muestra en al Fig. 32.

Figura 32- Resultado algoritmo *FP Growth* conjunto de datos PC2

```

=== Run information ===
Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation: DATAWARE_HOUSEv09-weka.filters.unsupervised.attribute.Remove-R1-3-weka.filters.unsupervised.attribute.Remove-R4-weka.filters.unsupervised.attribute.Remove-R1-3,5-11,13-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-weka.filters.unsupervised.attribute.NumericToBinary-Rfirst-last
Instances: 102667
Attributes: 45
=== Associator model (full training set) ===

FPGrowth found 28 rules (displaying top 20)

1. [DISTRITO=IXDH_binarized=1]: 44147 ==> [ZONA=ZONA2_binarized=1]: 44147 <conf:(1)> lift:(1.79) lev:(0.19) conv:(19534.98)
2. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, DISTRITO=IXDH_binarized=1]: 39263 ==> [ZONA=ZONA2_binarized=1]: 39263 <conf:(1)> lift:(1.79) lev:(0.17) conv:(17373.82)
3. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, DISTRITO=IXDH_binarized=1]: 36438 ==> [ZONA=ZONA2_binarized=1]: 36438 <conf:(1)> lift:(1.79) lev:(0.16) conv:(16123.76)
4. [CAT_TOTAL=BAJO_binarized=1, DISTRITO=IXDH_binarized=1]: 27460 ==> [ZONA=ZONA2_binarized=1]: 27460 <conf:(1)> lift:(1.79) lev:(0.12) conv:(12151.01)
5. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, DISTRITO=IXDH_binarized=1]: 36136 ==> [ZONA=ZONA2_binarized=1]: 36136 <conf:(1)> lift:(1.79) lev:(0.16) conv:(15990.13)
6. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, CAT_TOTAL=BAJO_binarized=1, DISTRITO=IXDH_binarized=1]: 24874 ==> [ZONA=ZONA2_binarized=1]: 24874 <conf:(1)> lift:(1.79) lev:(0.11) conv:(11006.71)
7. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, CAT_TOTAL=BAJO_binarized=1, DISTRITO=IXDH_binarized=1]: 23009 ==> [ZONA=ZONA2_binarized=1]: 23009 <conf:(1)> lift:(1.79) lev:(0.1) conv:(10181.45)
8. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, CAT_TOTAL=BAJO_binarized=1, DISTRITO=IXDH_binarized=1]: 22882 ==> [ZONA=ZONA2_binarized=1]: 22882 <conf:(1)> lift:(1.79) lev:(0.1) conv:(10125.25)
9. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, ZONA=ZONA2_binarized=1, CAT_TOTAL=BAJO_binarized=1]: 26299 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 26162 <conf:(0.99)> lift:(1.3) lev:(0.06) conv:(44.89)
10. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, CAT_TOTAL=BAJO_binarized=1, DISTRITO=IXDH_binarized=1]: 23009 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 22882 <conf:(0.99)> lift:(1.3) lev:(0.05) conv:(42.34)
11. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, CAT_TOTAL=BAJO_binarized=1, DISTRITO=IXDH_binarized=1]: 23009 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, ZONA=ZONA2_binarized=1]: 22882 <conf:(0.99)> lift:(2.12) lev:(0.12) conv:(95.38)
12. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, ZONA=ZONA2_binarized=1, CAT_TOTAL=BAJO_binarized=1, DISTRITO=IXDH_binarized=1]: 23009 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 22882 <conf:(0.99)> lift:(1.3) lev:(0.05) conv:(42.34)
13. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, ZONA=ZONA2_binarized=1]: 42317 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 41974 <conf:(0.99)> lift:(1.3) lev:(0.09) conv:(28.97)
14. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, DISTRITO=IXDH_binarized=1]: 36438 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 36136 <conf:(0.99)> lift:(1.3) lev:(0.08) conv:(28.33)
15. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, DISTRITO=IXDH_binarized=1]: 36438 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, ZONA=ZONA2_binarized=1]: 36136 <conf:(0.99)> lift:(2.11) lev:(0.19) conv:(63.81)

```

```

16. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, ZONA=ZONA2_binarized=1, DISTRITO=IXDH_binarized=1]:
36438 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 36136 <conf:(0.99)> lift:(1.3) lev:(0.08)
conv:(28.33)
17. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, CAT_TOTAL=BAJO_binarized=1]: 34931 ==>
[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 34387 <conf:(0.98)> lift:(1.29) lev:(0.07) conv:(15.1)
18. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1]: 63523 ==> [GRUPO_OPERATIVO=CUERPO DE
VIGILANCIA ADUANERA_binarized=1]: 62317 <conf:(0.98)> lift:(1.28) lev:(0.13) conv:(12.4)
19. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, CAT_TOTAL=MEDIO_binarized=1]: 24548 ==>
[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 24033 <conf:(0.98)> lift:(1.28) lev:(0.05) conv:(11.21)
20. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, DISTRITO=IXDH_binarized=1]: 39263 ==>
[ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1]: 36136 <conf:(0.92)> lift:(1.49) lev:(0.12) conv:(4.79)

```

Fuente: Weka

Para los datos correspondiente al PC3 se obtuvo un total de 557 atributos binarizados e igualmente se empleó 102667 instancias disponibles como se aprecia en la Fig. 33.

Figura 33 - Resultado algoritmo *FP Growth* conjunto de datos PC3

```

=== Run information ===
Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation: DATAWARE_HOUSEv09-weka.filters.unsupervised.attribute.Remove-R1-8,11-15,19-20-
weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-weka.filters.unsupervised.attribute.NumericToBinary-Rfirst-last
Instances: 102667
Attributes: 557
[list of attributes omitted]
=== Associator model (full training set) ===

FPGrowth found 16 rules (displaying top 16)

1. [SUBGRUPO=PRENDAS DE VESTIR/NUEVAS_binarized=1]: 21377 ==> [GRUPO=TEXTILES_binarized=1]: 21377 <conf:(1)>
lift:(4.13) lev:(0.16) conv:(16201.56)
2. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, SUBGRUPO=PRENDAS DE
VESTIR/NUEVAS_binarized=1]: 17907 ==> [GRUPO=TEXTILES_binarized=1]: 17907 <conf:(1)> lift:(4.13) lev:(0.13) conv:(13571.66)
3. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS_binarized=1]:
13847 ==> [GRUPO=TEXTILES_binarized=1]: 13847 <conf:(1)> lift:(4.13) lev:(0.1) conv:(10494.6)
4. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, ORIGEN_APREHENSION=CONTROL DE
RUTINA_binarized=1, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS_binarized=1]: 13710 ==> [GRUPO=TEXTILES_binarized=1]:
13710 <conf:(1)> lift:(4.13) lev:(0.1) conv:(10390.77)
5. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, SITIO_APREHENSION=CHACRAS-ECUADOR_binarized=1]:
34925 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 34872 <conf:(1)> lift:(1.31) lev:(0.08)
conv:(152.34)
6. [GRUPO_OPERATIVO=POLICIA NACIONAL_binarized=1]: 12974 ==> [ORIGEN_APREHENSION=ACTA DE ENTREGA -
RECEPCION_binarized=1]: 12893 <conf:(0.99)> lift:(6.72) lev:(0.11) conv:(134.83)
7. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, GRUPO=LICORES_binarized=1]: 11737 ==>
[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 11657 <conf:(0.99)> lift:(1.3) lev:(0.03) conv:(34.13)
8. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS_binarized=1]:
13847 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 13710 <conf:(0.99)> lift:(1.3) lev:(0.03)
conv:(23.63)
9. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS_binarized=1]:
13847 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, GRUPO=TEXTILES_binarized=1]: 13710
<conf:(0.99)> lift:(4.9) lev:(0.11) conv:(80.07)
10. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, GRUPO=TEXTILES_binarized=1, SUBGRUPO=PRENDAS DE
VESTIR/NUEVAS_binarized=1]: 13847 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 13710
<conf:(0.99)> lift:(1.3) lev:(0.03) conv:(23.63)
11. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1, GRUPO=TEXTILES_binarized=1]: 15783 ==>
[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 15616 <conf:(0.99)> lift:(1.29) lev:(0.03) conv:(22.13)
12. [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1]: 63523 ==> [GRUPO_OPERATIVO=CUERPO DE
VIGILANCIA ADUANERA_binarized=1]: 62317 <conf:(0.98)> lift:(1.28) lev:(0.13) conv:(12.4)
13. [SITIO_APREHENSION=CHACRAS-ECUADOR_binarized=1, GRUPO=LICORES_binarized=1]: 10997 ==>
[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1]: 10475 <conf:(0.95)> lift:(1.25) lev:(0.02) conv:(4.95)
14. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, SITIO_APREHENSION=CHACRAS-
ECUADOR_binarized=1]: 37274 ==> [ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1]: 34872 <conf:(0.94)> lift:(1.51)
lev:(0.12) conv:(5.91)
15. [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA_binarized=1, GRUPO=LICORES_binarized=1]: 12718 ==>
[ORIGEN_APREHENSION=CONTROL DE RUTINA_binarized=1]: 11657 <conf:(0.92)> lift:(1.48) lev:(0.04) conv:(4.57)
16. [SITIO_APREHENSION=CHACRAS-ECUADOR_binarized=1]: 40668 ==> [GRUPO_OPERATIVO=CUERPO DE VIGILANCIA
ADUANERA_binarized=1]: 37274 <conf:(0.92)> lift:(1.2) lev:(0.06) conv:(2.82)

```

Fuente: Weka

En las tablas 62, 63 y 64 continuación se presenta las reglas de asociación con los valores correspondientes a soporte y confianza para cada uno de los conjuntos de datos reducidos resultado de la aplicación del algoritmo *FP Growth*.

Tabla 62 - Reglas de asociación, soporte y confianza resultado del algoritmo *FP Growth* PC1

N. REGLA	ÍTEMS EN X	SOPORTE (x)	ÍTEMS EN Y	SOPORTE (x∩y)	CONFIANZA
1	[SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	21377	[GRUPO=TEXTILES]	21377	1,0000
2	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	17907	[GRUPO=TEXTILES]	17907	1,0000
3	[SITIO_APREHENSION=CHACRAS-ECUADOR, GRUPO=LICORES]	10997	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	10475	0,9525
4	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, BODEGA=DESTACAMENTO CHACRAS]	26009	[SITIO_APREHENSION=CHACRAS-ECUADOR]	24268	0,9331
5	[SITIO_APREHENSION=CHACRAS-ECUADOR, BODEGA=BODEGA HUAQUILLAS]	12319	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	11431	0,9279
6	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, BODEGA=BODEGA HUAQUILLAS]	12320	[SITIO_APREHENSION=CHACRAS-ECUADOR]	11431	0,9278
7	[SITIO_APREHENSION=CHACRAS-ECUADOR]	40668	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	37274	0,9165
8	[SITIO_APREHENSION=CHACRAS-ECUADOR, BODEGA=DESTACAMENTO CHACRAS]	26677	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	24268	0,9097
9	[BODEGA=DESTACAMENTO CHACRAS]	29389	[SITIO_APREHENSION=CHACRAS-ECUADOR]	26677	0,9077

Fuente: Weka

Con la que se trabaja se evidencia que los datos no se encuentran balanceados por lo que varios de los ítems con categorías específicas se repiten como, por ejemplo: GRUPO con su categoría "TEXTILES" o GRUPO\_OPERATIVO con su categoría "CUERPO DE VIGILANCIA ADUANERA" y no muestran información relevante.

Tabla 63 - Reglas de asociación, soporte y confianza resultado del algoritmo *FP Growth* PC2

N. REGLA	ÍTEMS EN X	SOPORTE (x)	ÍTEMS EN Y	SOPORTE (x∩y)	CONFIANZA
1	[DISTRITO=IXDH]	44147	[ZONA=ZONA2]	44147	1,0000
2	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, DISTRITO=IXDH]	39263	[ZONA=ZONA2]	39263	1,0000
3	[ORIGEN_APREHENSION=CONTROL DE RUTINA, DISTRITO=IXDH]	36438	[ZONA=ZONA2]	36438	1,0000
4	[CAT_TOTAL=BAJO, DISTRITO=IXDH]	27460	[ZONA=ZONA2]	27460	1,0000
5	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, ORIGEN_APREHENSION=CONTROL DE RUTINA, DISTRITO=IXDH]	36136	[ZONA=ZONA2]	36136	1,0000
6	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, CAT_TOTAL=BAJO, DISTRITO=IXDH]	24874	[ZONA=ZONA2]	24874	1,0000
7	[ORIGEN_APREHENSION=CONTROL DE RUTINA, CAT_TOTAL=BAJO, DISTRITO=IXDH]	23009	[ZONA=ZONA2]	23009	1,0000

8	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, ORIGEN_APREHENSION=CONTROL DE RUTINA, CAT_TOTAL=BAJO, DISTRITO=IXDH]	22882	[ZONA=ZONA2]	22882	1,0000
9	[ORIGEN_APREHENSION=CONTROL DE RUTINA, ZONA=ZONA2, CAT_TOTAL=BAJO]	26299	[GRUPO_OPERATIVO=CUERPO DE ADUANERA]	26162	0,9948
10	[ORIGEN_APREHENSION=CONTROL DE RUTINA, CAT_TOTAL=BAJO, DISTRITO=IXDH]	23009	[GRUPO_OPERATIVO=CUERPO DE ADUANERA]	22882	0,9945
11	[ORIGEN_APREHENSION=CONTROL DE RUTINA, CAT_TOTAL=BAJO, DISTRITO=IXDH]	23009	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	22882	0,9945
12	[ORIGEN_APREHENSION=CONTROL DE RUTINA, ZONA=ZONA2, CAT_TOTAL=BAJO, DISTRITO=IXDH]	23009	[GRUPO_OPERATIVO=CUERPO DE ADUANERA]	22882	0,9945
13	[ORIGEN_APREHENSION=CONTROL DE RUTINA, ZONA=ZONA2]	42317	[GRUPO_OPERATIVO=CUERPO DE ADUANERA]	41974	0,9919
14	[ORIGEN_APREHENSION=CONTROL DE RUTINA, DISTRITO=IXDH]	36438	[GRUPO_OPERATIVO=CUERPO DE ADUANERA]	36136	0,9917
15	[ORIGEN_APREHENSION=CONTROL DE RUTINA, DISTRITO=IXDH]	36438	[GRUPO_OPERATIVO=CUERPO DE ADUANERA, ZONA=ZONA2]	36136	0,9917
16	[ORIGEN_APREHENSION=CONTROL DE RUTINA, ZONA=ZONA2, DISTRITO=IXDH]	36438	[GRUPO_OPERATIVO=CUERPO DE ADUANERA]	36136	0,9917
17	[ORIGEN_APREHENSION=CONTROL DE RUTINA, CAT_TOTAL=BAJO]	34931	[GRUPO_OPERATIVO=CUERPO DE ADUANERA]	34387	0,9844
18	[ORIGEN_APREHENSION=CONTROL DE RUTINA]	63523	[GRUPO_OPERATIVO=CUERPO DE ADUANERA]	62317	0,9810
19	[ORIGEN_APREHENSION=CONTROL DE RUTINA, CAT_TOTAL=MEDIO]	24548	[GRUPO_OPERATIVO=CUERPO DE ADUANERA]	24033	0,9790
20	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, DISTRITO=IXDH]	39263	[ORIGEN_APREHENSION=CONTROL DE RUTINA]	36136	0,9204

Fuente: Weka

Como en el caso anterior, en la Tabla 63 que puede constatar que el balance de la información influye en los resultados para el PC2 por lo que ZONAS con su categoría “ZONA2”, es la que predomina al igual que GRUPO\_OPERATIVO con su categoría “CUERPO DE VIGILANCIA ADUANERA”.

Tabla 64 - Reglas de asociación, soporte y confianza resultado del algoritmo *FP Growth* PC3

N. REGLA	ÍTEMS EN X	SOPORTE (x)	ÍTEMS EN Y	SOPORTE (x∩y)	CONFIANZA
1	[SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	21377	[GRUPO=TEXTILES]	21377	1,0000
2	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	17907	[GRUPO=TEXTILES]	17907	1,0000
3	[ORIGEN_APREHENSION=CONTROL DE RUTINA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	13847	[GRUPO=TEXTILES]	13847	1,0000
4	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, ORIGEN_APREHENSION=CONTROL DE RUTINA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	13710	[GRUPO=TEXTILES]	13710	1,0000
5	[ORIGEN_APREHENSION=CONTROL DE RUTINA, SITIO_APREHENSION=CHACRAS-ECUADOR]	34925	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	34872	0,9985
6	[GRUPO_OPERATIVO=POLICIA NACIONAL]	12974	[ORIGEN_APREHENSION=ACTA DE ENTREGA - RECEPCION]	12893	0,9938
7	[ORIGEN_APREHENSION=CONTROL DE RUTINA, GRUPO=LICORES]	11737	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	11657	0,9932

8	[ORIGEN_APREHENSION=CONTROL DE RUTINA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	13847	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	13710	0,9901
9	[ORIGEN_APREHENSION=CONTROL DE RUTINA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	13847	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, GRUPO=TEXTILES]	13710	0,9901
10	[ORIGEN_APREHENSION=CONTROL DE RUTINA, GRUPO=TEXTILES, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	13847	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	13710	0,9901
11	[ORIGEN_APREHENSION=CONTROL DE RUTINA, GRUPO=TEXTILES]	15783	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	15616	0,9894
12	[ORIGEN_APREHENSION=CONTROL DE RUTINA]	63523	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	62317	0,9810
13	[SITIO_APREHENSION=CHACRAS-ECUADOR, GRUPO=LICORES]	10997	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	10475	0,9525
14	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, SITIO_APREHENSION=CHACRAS-ECUADOR]	37274	[ORIGEN_APREHENSION=CONTROL DE RUTINA]	34872	0,9356
15	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, GRUPO=LICORES]	12718	[ORIGEN_APREHENSION=CONTROL DE RUTINA]	11657	0,9166
16	[SITIO_APREHENSION=CHACRAS-ECUADOR]	40668	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA]	37274	0,9165

Fuente: Weka

En la Tabla 64 se verifica que no hay un balance de la información en los resultados para el PC3 por lo que GRUPO\_OPERATIVO con su categoría “CUERPO DE VIGILANCIA ADUANERA”, es la que predomina como también en GRUPO con su categoría “TEXTILES”.

Como resultado de la aplicación del algoritmo *FP Growth* se pudo obtener las siguientes reglas como las más importantes las que se muestra en la Tabla 65 a continuación:

Tabla 65- Listado de mejores reglas de asociación según el algoritmo *FP Growth*

N. REGLA	ÍTEMS EN X	SOPORTE (x)	ÍTEMS EN Y	SOPORTE (x∩y)	CONFIANZA
1	[DISTRITO=IXDH]	44147	[ZONA=ZONA2]	44147	1,0000
2	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, DISTRITO=IXDH]	39263	[ZONA=ZONA2]	39263	1,0000
3	[ORIGEN_APREHENSION=CONTROL DE RUTINA, DISTRITO=IXDH]	36438	[ZONA=ZONA2]	36438	1,0000
4	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, ORIGEN_APREHENSION=CONTROL DE RUTINA, DISTRITO=IXDH]	36136	[ZONA=ZONA2]	36136	1,0000
5	[CAT_TOTAL=BAJO, DISTRITO=IXDH]	27460	[ZONA=ZONA2]	27460	1,0000
6	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, CAT_TOTAL=BAJO, DISTRITO=IXDH]	24874	[ZONA=ZONA2]	24874	1,0000
7	[ORIGEN_APREHENSION=CONTROL DE RUTINA, CAT_TOTAL=BAJO, DISTRITO=IXDH]	23009	[ZONA=ZONA2]	23009	1,0000
8	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, ORIGEN_APREHENSION=CONTROL DE RUTINA, CAT_TOTAL=BAJO, DISTRITO=IXDH]	22882	[ZONA=ZONA2]	22882	1,0000
9	[SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	21377	[GRUPO=TEXTILES]	21377	1,0000
10	[SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	21377	[GRUPO=TEXTILES]	21377	1,0000



11	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	17907	[GRUPO=TEXTILES]	17907	1,0000
12	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	17907	[GRUPO=TEXTILES]	17907	1,0000
13	[ORIGEN_APREHENSION=CONTROL DE RUTINA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	13847	[GRUPO=TEXTILES]	13847	1,0000
14	[GRUPO_OPERATIVO=CUERPO DE VIGILANCIA ADUANERA, ORIGEN_APREHENSION=CONTROL DE RUTINA, SUBGRUPO=PRENDAS DE VESTIR/NUEVAS]	13710	[GRUPO=TEXTILES]	13710	1,0000

Fuente: Weka

### 3.2. Atípicos

El cálculo de los valores que se los consideran atípicos se los realizó tomando en cuenta los valores categorizados acorde a los criterios definidos en el capítulo anterior, debido que el proceso del cálculo para la identificación de los valores requeridos como son rango intercuartil y datos atípicos se requiere ordenarlos de mayor a menor. Como requisito previo a los cálculos anterior se necesita conocer los cuartiles, siendo los que fraccionan la distribución de datos en 4 partes iguales; el primer cuartil  $Q_1$  deja por debajo el 25% de los datos, el segundo cuartil  $Q_2$  el 50% por debajo y por último el tercer cuartil  $Q_3$  deja por debajo el 75% de los datos (Posada Hernández, 2016).

Tabla 66 - Formulas de cálculo por cuartil

CUARTIL	% DE CASOS	FÓRMULA POR CUARTIL
$Q_1$	25%	$Q_1 = (n + 1) \frac{25}{100}$
$Q_2$	50% (MEDIANA)	$Q_2 = (n + 1) \frac{50}{100}$
$Q_3$	75%	$Q_3 = (n + 1) \frac{75}{100}$
$Q_4$	100%	

Fuente: Adaptación de Posada Hernández, 2016

La definición de valores atípicos se lo realiza en base a la definición de los límites que se los obtiene con las siguientes formulas:

#### ▫ Evaluación

Para la evaluación y determinación de datos atípicos se realizó el cálculo de los valores de cuartiles, mediana, rango intercuartil y limites obteniendo como resultado los datos registrados en la Tabla 67 a continuación:

Tabla 67 – Calculo de cuartiles, mediana y límites

ATRIBUTO	Q1	Q2 MEDIANA	Q3	Q4	RI (Q3-Q1)	Q3 + 1.5 (RI)	Q1 - 1.5 (RI)
AÑO	2	3	4	6	2	7	-1
MES	4	7	9	12	5	16,5	-3,5
DIA	13	21	28	31	15	50,5	-9,5
CAT_CANTIDAD	1	2	2	5	1	3,5	-0,5
CAT_CANTIDAD	1	2	2	5	1	3,5	-0,5
UNIDADES	50	59	59	59	9	72,5	36,5
CAT_PRECIO	1	1	1	4	0	1	1
CAT_TOTAL	1	1	2	4	1	3,5	-0,5
GRUPO	9	13	26	27	17	51,5	-16,5
SUBGRUPO	32	118	188	217	156	422	-202
PROCEDENCIA	9	9	9	12	0	9	9
BODEGA	16	16	28	39	12	46	-2
STATUS	1	1	1	3	0	1	1
ORIGEN_APREHENSION	8	8	8	20	0	8	8
SITIO_APREHENSION	71	71	207	287	136	411	-133
GRUPO_OPERATIVO	5	5	5	8	0	5	5
DISTRITO	5	8	9	10	4	15	-1
ZONA	2	2	3	5	1	4,5	0,5
TOTAL	85	300	1068	2000000	983	2542,5	-1389,5

Fuente: Elaboración Propia

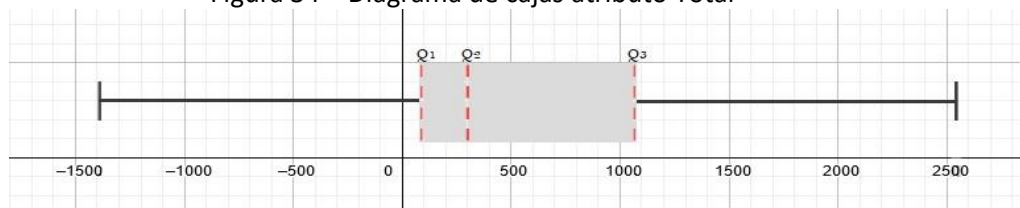
Se identifica como atributos no balanceados a CAT\_PRECIO, PROCEDENCIA, STATUS, ORGIEN\_APREHENSION, GRUPO\_OPERATIVO por poseer un valor predominante para Q1, Q2(media) y Q3. En el caso de los atributos de CAT\_CANTIDAD, UNIDADES, el valor correspondiente a la mediana (Q2) y Q3 son iguales por lo que poder determinar la presencia a atípicos es sumamente difícil.

En base a la Tabal 67 se identifica datos atípicos de las variables cuantitativas y cualitativas usando diagramas de cajas, respectivamente.

## TOTAL

Los datos atípicos para este atributo son aquellos menores -1389,5 y mayores a 2542,5 en la base de datos se encontraron **12130** registros con valores mayores a 2542,5 en la figura continuación se representa la distribución de los datos atípicos de este atributo

Figura 34 – Diagrama de cajas atributo Total



Fuente: Elaboración Propia

## **AÑO**

En este atributo almacena el año en el que se registró la aprehensión que después de la categorización y cálculos se muestra que no puede ser menor a -1 o mayor a 7 para ser considerados atípicos, por lo que se identifica que este atributo no tiene datos atípicos ya que los datos categorizados se los almacenó del 1 al 6 respectivamente.

## **MES**

Para que un dato del atributo mes sea considerado como atípico debe ser menor a -3,5 o mayor a 16,5 teniendo en cuenta que para este atributo se categorizo los meses de año con datos almacenados del 1 al 12 respectivamente, por este motivo en este atributo no se identifica ningún valor atípico.

## **DIA**

El atributo día al igual que los dos anteriores atributos presenta una categorización en base a la fecha desde el 1 al 31 respectivamente, para ese caso los valores atípicos se los considera aquellos que son menores a -9,5 o mayores a 50,5 por lo no se identifica ningún valor a considerarse como atípico.

## **CAT\_CANTIDAD**

Los datos considerados atípicos para el atributo Cat\_Cantidad son aquellos menores -0,5 y mayores a 3,5 por lo que al estar categorizados con valores del 1 al 4 no se puede identificar ningún dato atípico.

## **UNIDADES**

Para el atributo unidades se identifica como atípico aquellos valores que sean menores a 36,5 y mayores a 72,5 al estar categorizados respectivamente del 1 al 59 no se puede identificar valores atípicos.

## **CAT\_TOTAL**

El atributo Cat\_Total se encuentra relacionado con el atributo Total por lo que identificar atípicos con la categorización aplicada es del 1 al 4 por lo que no es posible identificar atípicos ya que deberían ser valores menores a -0,5 o mayores a 3,5.

## **GRUPO**

Los datos atípicos para el atributo grupo son identificados si son menores a -16,5 o mayores a 51,5 por lo que en este caso no se puede identificar valores atípicos al estar los valores almacenados del 1 al 33 respectivamente.

## **SUBGRUPO**

Para el atributo subgrupo los valores a ser considerados atípicos deben ser los registros con valores menores a -202 y mayores a 422 por lo que no se puede identificar aquellos atípicos por lo que se los categorizo del 1 al 217.

## **BODEGA**

El atributo bodega presenta una categorización en base a la fecha desde el 1 al 39 respectivamente, para ese caso los valores atípicos se los considera aquellos que son menores a -2 o mayores a 46 por lo no se identifica ningún valor a considerarse como atípico.

## **SITIO\_APREHENSION**

Para el atributo Sitio\_Aprehension los valores a ser considerados atípicos deben ser los registros con valores menores a -133 y mayores a 411 por lo que no se puede identificar aquellos atípicos por lo que se los categorizo del 1 al 287.

## **DISTRITO**

Los datos atípicos para el atributo distrito son identificados si son menores a -1 o mayores a 15 por lo que en este caso no se identificó valores atípicos al estar los valores almacenados del 1 al 10 respectivamente.

## **ZONA**

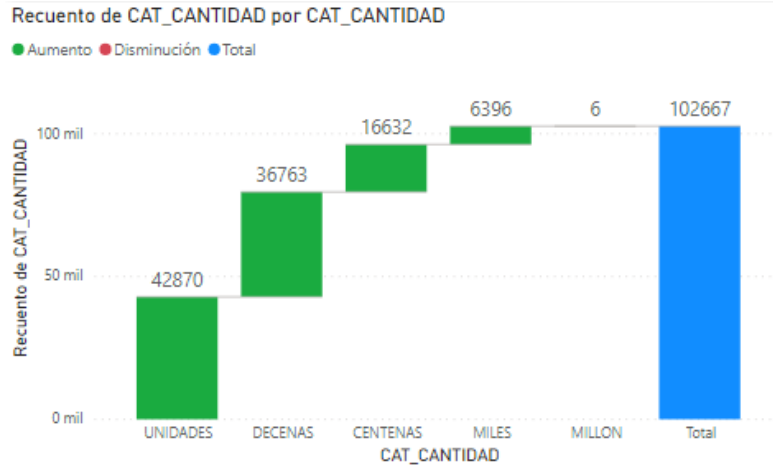
Para el atributo zona se identifica como atípico aquellos valores que sean menores a 0,5 y mayores a 4,5 al estar categorizados respectivamente del 1 al 5 no se puede identificar valores atípicos.

Al disponer de información compuesta de atributos no balanceados, identificación de valores atípicos cercanos a la realidad es complicado, es este tipo de información la identificación del valor predominante permite identificar los valores típicos para cada atributo como son los siguientes:

## CAT\_CANTIDAD

En el atributo CAT\_CANTIDAD se observa que predomina la opción de unidades, seguido de los grupos decenas y centenas, mientras que la opción de millón no tiene un valor representativo considerándolo, como atípico respecto a total.

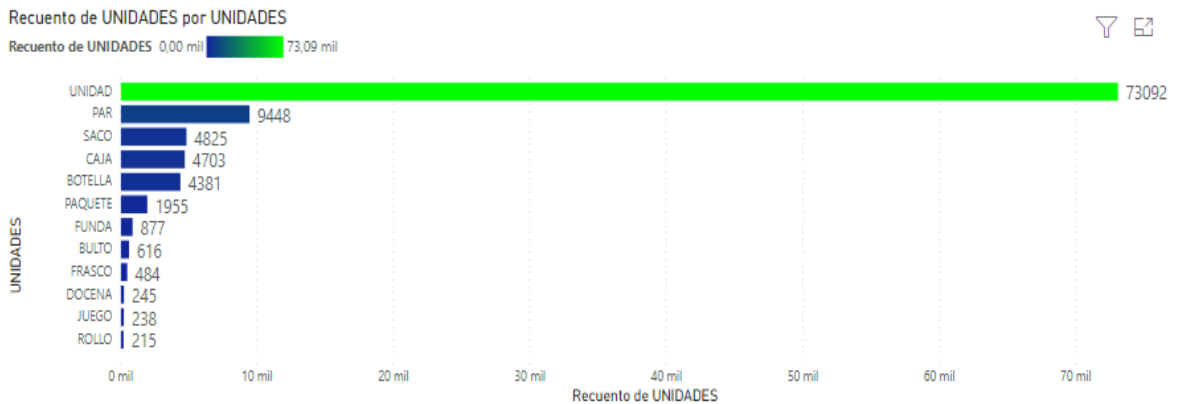
Figura 35 - Estadística del atributo CAT\_CANTIDAD



## UNIDADES

Para el atributo UNIDADES se observa notablemente que predomina la opción de unidades con un 71,9% de la totalidad de los registros, seguido de las unidades par, saco, caja y botella de tal manera que se suma el 93,94% entre todas las unidades antes mencionadas, por lo que se puede concluir que los valores atípicos de este atributo se encuentran presentes en todas las opciones restantes que conforma una lista de 54 tipos de unidades.

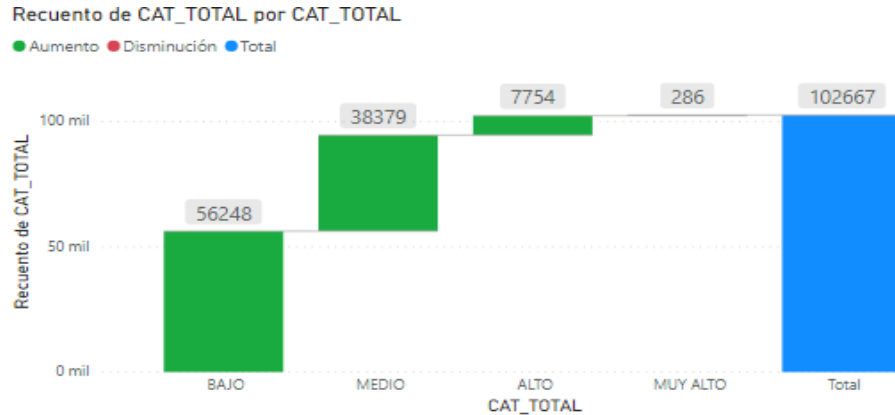
Figura 36 - Estadísticas del atributo UNIDADES



## CAT\_TOTAL

En el atributo CAT\_TOTAL se observa que predomina categorización de valores correspondiente a bajo, seguido de la opción medio, mientras que la opción categorizada como muy alto no tiene un valor representativo considerándolo, como dato atípico.

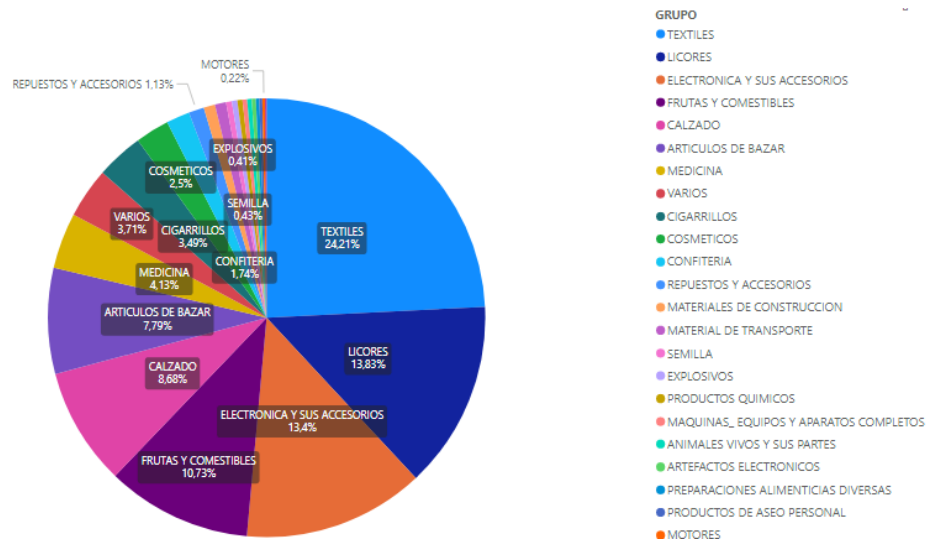
Figura 37 - Estadística del atributo CAT\_TOTAL



## GRUPO

En el atributo GRUPO se observa notablemente que predomina el grupo textiles, seguido de los grupos licores, electrónica y sus accesorios, frutas y comestibles, calzado y artículos de bazar, mientras que los grupos de courier, perlas finas piedras preciosas y metales preciosos no tiene un valor representativo además embarcaciones con 0,01%, madera con 0,02% y productos DEL MAR con 0,06% son considerados como atípicos respecto a total.

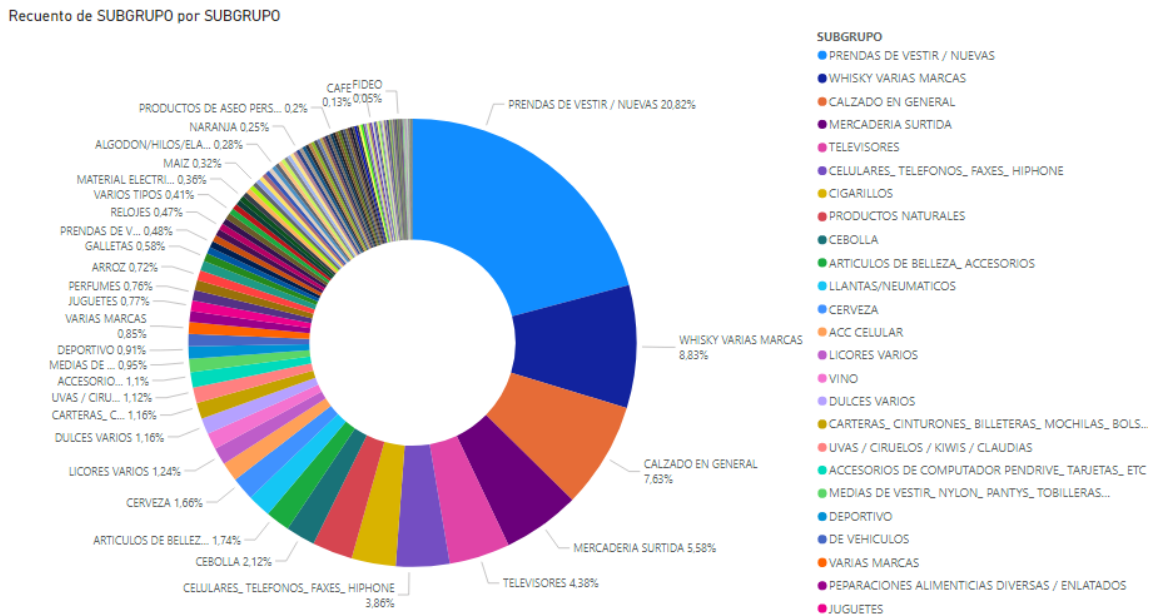
Figura 38 – Estadísticas del atributo GRUPO



## SUBGRUPO

Para el atributo SUBGRUPO se observa notablemente que predomina la opción de prendas de vestir/nuevas con un 20,82% de la totalidad de los registros, seguido de los subgrupos: whisky varias marcas, calzado en general, mercadería surtida y televisores, identificando como atípicos todos aquellos subgrupos que se encuentra por debajo del 1,0% de la totalidad.

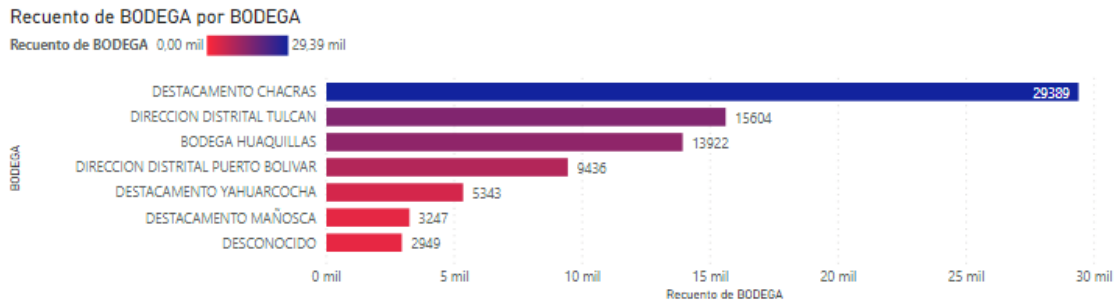
Figura 39 – Estadística del atributo SUBGRUPO



## BODEGA

En el atributo BODEGA se observa notablemente que predomina la bodega del Destacamento Chacras, seguido de las bodegas de Dirección Distrital Tulcán, Bodega Huaquillas, Dirección Distrital Puerto Bolívar mientras que las bodegas denominadas como: de Bodega de SAI-SENAE Aeropuerto JJO GYE, Destacamento Saucillo, Unidad de Lavado de Activos de la PPNN, Almacén Temporal Bolivariana, Aretina Y Bodega DE SAI-SENAE Aeropuerto MS UIO no presentan un valor representativo además Bodega 22 DE LA CAE Puerto Marítimo con 0,01% y Bodega de Aduana - Pto. De Manta con 0,07% son considerados como atípicos respecto a total.

Figura 40 - Estadísticas del atributo BODEGA



### SITIO APREHENSION

En el atributo BODEGA se observa notablemente que predomina la bodega del Destacamento Chacras, seguido de las bodegas de Dirección Distrital Tulcán, Bodega Huaquillas, Dirección Distrital Puerto Bolívar mientras que las bodegas denominadas como: de Bodega de SAI-SENAE Aeropuerto JJO GYE, Destacamento Saucillo, Unidad de Lavado de Activos de la PPNN, Almacén Temporal Bolivariana, Aretina Y Bodega DE SAI-SENAE Aeropuerto MS UIO no presentan un valor representativo además Bodega 22 DE LA CAE Puerto Marítimo con 0,01% y Bodega de Aduana - Pto. De Manta con 0,07% son considerados como atípicos respecto a total.

Figura 41 - Estadísticas del atributo SITIO\_ APREHENSION

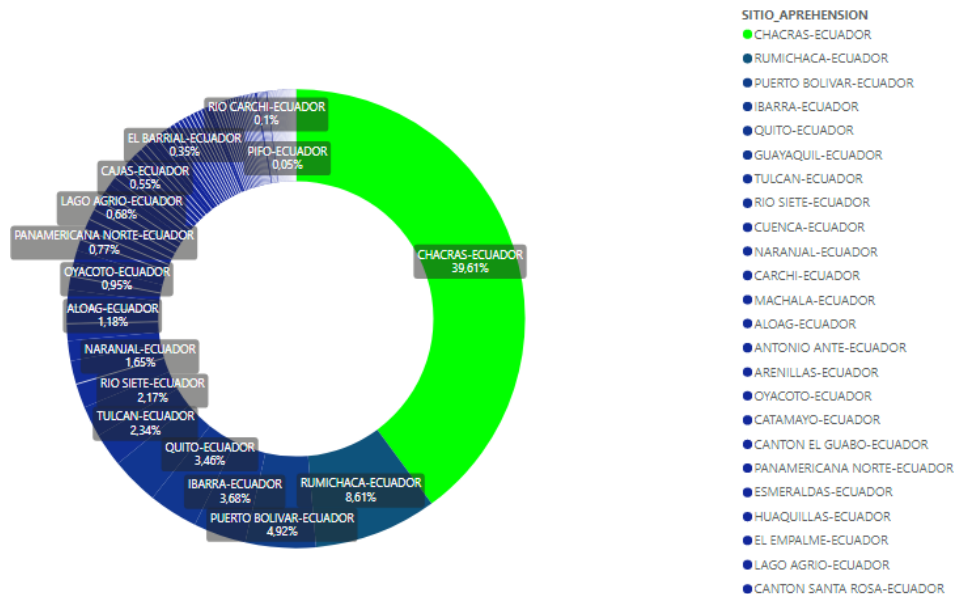
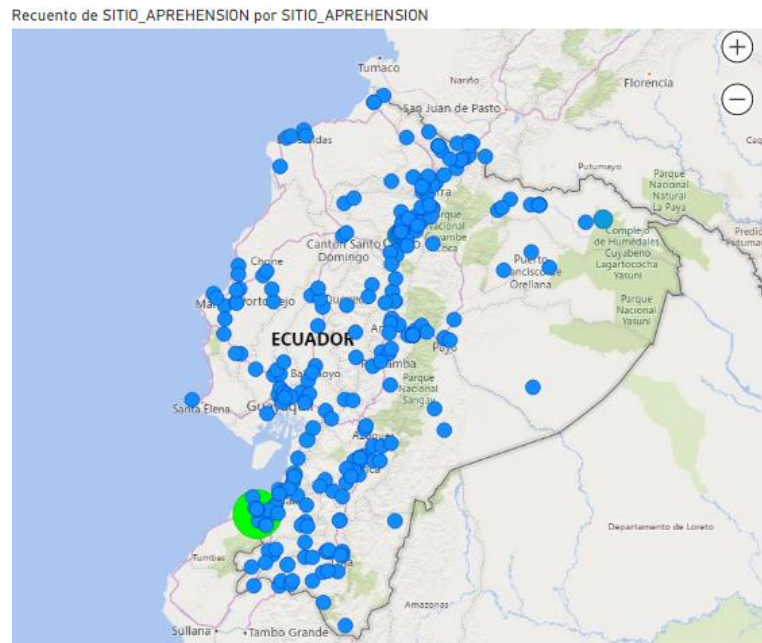




Figura 42 – Mapa de ubicaciones geográficas de Sitios de Aprehensión



**GRUPO OPERATIVO**

En el atributo GRUPO\_OPERATIVO se observa que predomina el Cuerpo De Vigilancia Aduanera sobre el resto de las opciones, mientras que para Agrocalidad se registra un 0,01% de total de igual manera para la ARCSA e Inspectoría de Pesca no tiene un valor representativo, considerando de esta manera las categorías mencionadas anteriormente como datos atípicos de este atributo.

Figura 43 - Estadísticas del atributo GRUPO\_OPERATIVO

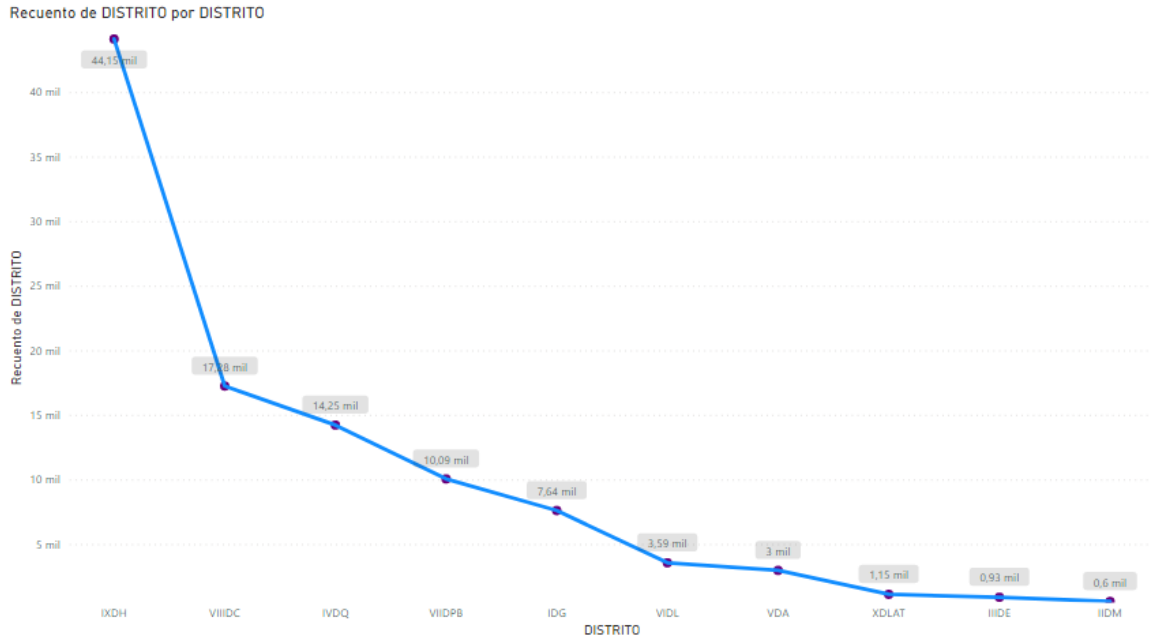


**DISTRITO**

En el atributo DISTRITO se observa que predomina el IX Distrito (Huaquillas), seguido por el VIII Distrito (Carchi) y IV Distrito (Quito), mientras que el III Distrito (Esmeraldas) y el II

Distrito (Manta) representan 0,90% y 0,58% del total considerándose los últimos dos distritos como atípicos.

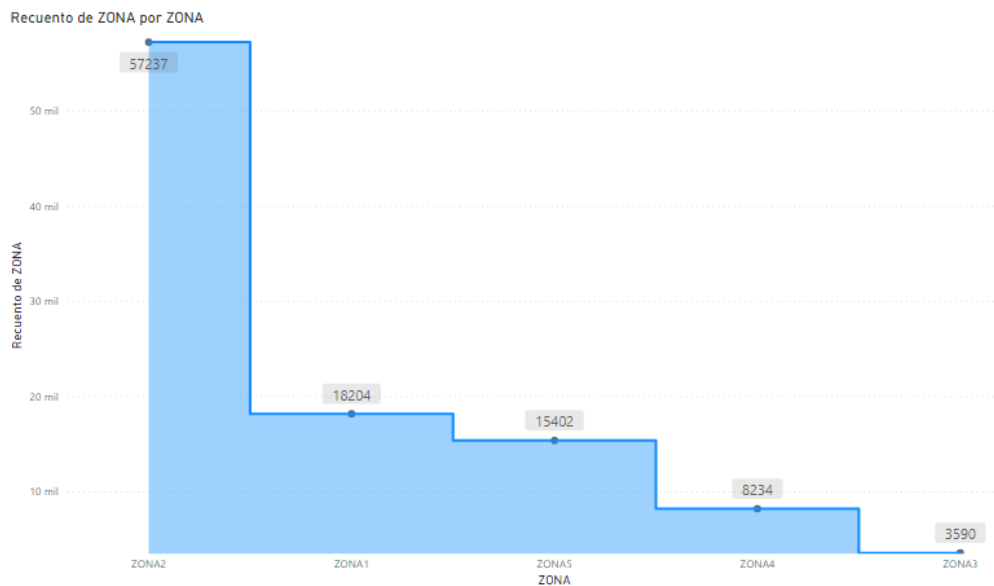
Figura 44 - Estadísticas del atributo DISTRITO



## ZONA

En el atributo ZONA se observa que predomina la Zona 2 con las del 50% de la totalidad de la información, seguido por la Zona 1 y 5 mientras que la Zona 3 con un 3,50% de total se la consideraría como atípico para este atributo.

Figura 45 - Estadísticas del atributo ZONA



### 3.3. Obtención del conocimiento

#### Asociación

Las reglas obtenidas en el proceso de interpretación de tareas de asociación reflejan en gran coincidencia con los valores estadísticos ya conocidos previamente en los informes presentados por la entidad de control (anónima), sin presentar conocimiento relevante o de importancia para los expertos. Las reglas obtenidas se aprecian en la Tabla 58 para el algoritmo *Apriori* y para el algoritmo *FP Growth* en la Tabla 63 ambas correspondiente a la sección 3.1.2.

#### Agrupamiento

##### *Algoritmo Kmeans*

#### Conocimiento obtenido del dataset completo:

**Clúster 0:** aprehensiones realizadas en el mes de febrero, de prendas de vestir/nuevas del grupo textiles. Identificadas por unidades de cantidad tipo decena, categorizado en un total bajo de procedencia extranjera. Resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en los Chacras – Ecuador.

**Clúster 1:** aprehensiones realizadas en el mes de marzo, de prendas de vestir/nuevas del grupo textiles. Identificadas por unidades de cantidad tipo unidades, categorizado en un total bajo de procedencia extranjera. Resultado de Acta de Entrega Recepción realizado por la Policía Nacional en Esmeraldas – Ecuador.

**Clúster 2:** aprehensiones realizadas en el mes de agosto, de whisky varias marcas identificadas por unidades del grupo Licores, de cantidad tipo decena categorizado en un total bajo de procedencia extranjera. Resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en los Chacras – Ecuador.

**Clúster 3:** aprehensiones realizadas en el mes de agosto, de medias de vestir, nylon, pantys o tobilleras por pares, pertenecientes al grupo Textiles, de cantidad tipo centenas categorizado en un total bajo, de procedencia extranjera. Resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en los Chacras – Ecuador.

**Clúster 4:** aprehensiones realizadas en el mes de enero, de prendas de vestir/nuevas por unidades del grupo Textiles, de cantidad tipo decenas, categorizado en un total bajo de procedencia extranjera. Resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en los Chacras – Ecuador.

**Clúster 5:** aprehensiones realizadas en el mes de abril, de arroz por sacos del grupo Frutas y Comestibles, de cantidad tipo unidades, categorizado en un total bajo de procedencia extranjera. Resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en los Chacras – Ecuador.

**Clúster 6:** aprehensiones realizadas en el mes de junio, del grupo Máquinas, equipos y aparatos de laboratorio y sus accesorios, de cantidad tipo unidades, categorizado en un total medio de procedencia extranjera. Resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en Río Siete – Ecuador.

**Clúster 7:** aprehensiones realizadas en el mes de julio, de calzado deportivo por pares del grupo Calzado, de cantidad tipo decenas, categorizado en un total medio de procedencia extranjera. Resultado de un Allanamiento realizado por el Cuerpo de Vigilancia Aduanera en Guayaquil – Ecuador.

**Clúster 8:** aprehensiones realizadas en el mes de noviembre, de preparaciones alimenticias diversas/enlatados por cajas del grupo Frutas y Comestibles, de cantidad tipo decenas, categorizado en un total medio, de procedencia extranjera. Resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en Tulcán – Ecuador.

**Clúster 9:** aprehensiones realizadas en el mes de febrero, de medias de vestir, nylon, pantys o tobilleras por pares del grupo Textiles, de cantidad tipo unidades, categorizado en un total bajo de procedencia extranjera. Resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en los Rumichaca – Ecuador.

#### **Conocimiento obtenido del PC1**

**Clúster 0 y 4 (IX Distrito Huaquillas):** aprehensiones de prendas de vestir / nuevas del grupo Textiles en Chacras – Ecuador realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 1 (VII Distrito Puerto Bolívar):** aprehensiones de grupo de frutas y comestibles y también de cigarrillos para el caso del específico de subgrupo, en Puerto Bolívar – Ecuador realizadas por Policía Nacional.

**Clúster 2 (V Distrito Azuay):** aprehensiones de whisky varias marcas pertenecientes al grupo Licores en Chacras – Ecuador realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 3 (VI Distrito Loja):** aprehensiones del perteneciente al grupo de Frutas y Comestibles y también de neumáticos para el caso específico de subgrupo, en Chacras – Ecuador realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 5 (No Class):** aprehensiones de cebollas perteneciente al grupo Frutas y Comestibles en Chacras - Ecuador realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 6 (X Distrito Latacunga):** aprehensiones de mercadería surtida perteneciente al grupo de Artículos de bazar en Chacras - Ecuador realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 7 (I Distrito Guayas):** aprehensiones de calzado en general perteneciente al grupo de Calzado en Guayaquil – Ecuador realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 8 (IV Distrito Quito):** aprehensiones de teléfonos, faxes o iPhone pertenecientes al grupo de Electrónica y sus accesorios en Tulcán – Ecuador realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 9 (VIII Distrito Carchi):** aprehensiones de televisores perteneciente al grupo de Electrónica y sus Accesorios en Rumichaca – Ecuador realizadas por el Cuerpo de Vigilancia Aduanera.

#### **Conocimiento obtenido del PC2**

**Clúster 0 (IX Distrito Huaquillas):** aprehensiones categorizadas con un monto total Bajo en la Zona2 resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 1 (VII Distrito Puerto Bolívar):** aprehensiones categorizadas con un monto total Bajo en la Zona2 resultado de una Acta Entrega Recepción realizado por el Policía Nacional.

**Clúster 2 (VI Distrito Loja):** aprehensiones categorizadas con un monto total Medio en la Zona2 resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 3 (I Distrito Guayas):** aprehensiones categorizadas con un monto total Medio en la Zona4 resultado de un Allanamiento realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 4 (VIII Distrito Carchi):** aprehensiones categorizadas con un monto total Medio en la Zona1 resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 5 (III Distrito Esmeraldas):** aprehensiones categorizadas con un monto total Bajo en la Zona1 resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 6 (IV Distrito Quito) :** aprehensiones categorizadas con un monto total Bajo en la Zona5 resultado de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 7 (X Distrito Latacunga):** aprehensiones categorizadas con un monto total Alto en la Zona1 resultado de Acciones Inteligencia y Protección DIP realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 8 (II Distrito Manta):** aprehensiones categorizadas con un monto total Bajo en la Zona2 resultado de Acciones Inteligencia y Protección DIP realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 9 (V Distrito Azuay):** aprehensiones categorizadas con un monto total Bajo en la Zona1 resultado de Control Conjunto Inter Institucional realizado por el Cuerpo de Vigilancia Aduanera.

### **Conocimiento obtenido del PC3**

**Clúster 0 (IX Distrito Huaquillas):** aprehensiones de prendas de vestir / nuevas del grupo Textiles en Chacras – Ecuador resultado de Control de Rutina realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 1 (VII Distrito Puerto Bolívar):** aprehensiones del grupo de frutas y comestibles y además de prendas de vestir / nuevas en Puerto Bolívar resultado de Acta de Entrega Recepción realizado por la Policía Nacional.

**Clúster 2 (II Distrito Manta):** aprehensiones de whisky varias marcas del grupo Licores en Chacras – Ecuador resultado de Control de Rutina realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 3 (III Distrito Esmeraldas):** aprehensiones de medias de vestir, nylon, pantys o tobilleras del grupo Textiles en Chacras – Ecuador resultado de Control de Rutina realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 4 (VI Distrito Loja):** aprehensiones de cebollas del grupo Frutas y Comestibles en Chacras – Ecuador resultado de Control de Rutina realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 5 (V Distrito Azuay):** aprehensiones de calzado en general del grupo Calzado en Chacras – Ecuador resultado de Control de Rutina realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 6 (I Distrito Guayas):** aprehensiones de calzado en general del grupo Calzado en Guayaquil – Ecuador resultado de Allanamiento realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 7 (IV Distrito Quito):** aprehensiones de celulares, teléfonos, faxes o iPhone del grupo Electrónica y sus Accesorios en Tulcán – Ecuador resultado de Control de Rutina realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 8 (VII Distrito Carchi):** aprehensiones de Televisores del grupo Electrónica y sus Accesorios en Tulcán – Ecuador resultado de Control de Rutina realizadas por el Cuerpo de Vigilancia Aduanera.

**Clúster 9 (No Class):** aprehensiones de prendas de vestir / nuevas del grupo Textiles en el Cantón El Guabo – Ecuador resultado de Control de Rutina realizadas por el Cuerpo de Vigilancia Aduanera.

### ***Algoritmo EM***

#### **Conocimiento obtenido del dataset completo:**

**Clúster 0 (X Distrito Latacunga):** aprehensiones realizadas en el mes de junio valoras con un precio bajo y un total medio correspondientes al grupo de repuestos y accesorios de vehículos,

con procedencia española almacenados en el Destacamento Saucillo resultado de un control de rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 1 (Distrito Quito IV):** aprehensiones realizadas en el mes de julio valoras con un precio bajo y un total bajo correspondientes a Acc. Celular perteneciente grupo de electrónica y sus accesorios, con procedencia extranjera almacenados en el Destacamento Saucillo resultado de un control en carretera realizado en un control conjunto interinstitucional.

**Clúster 2 (I Distrito Guayas):** aprehensiones realizadas en el mes de junio valoras con un precio bajo y un total medio correspondientes a confitería perteneciente grupo de dulces varios, con procedencia extranjera almacenados en el Destacamento El Empalme como resultado de un control de rutina realizado por el cuerpo de vigilancia aduanera.

**Clúster 3 (VI Distrito Loja):** aprehensiones realizadas en el mes de junio valoras con un precio bajo y un total bajo correspondientes al grupo Licores, con procedencia extranjera almacenados en el Destacamento Chacras como resultado de un control de rutina realizado por el cuerpo de vigilancia aduanera.

**Clúster 4 (IX Distrito Huaquillas):** aprehensiones realizadas en el mes de abril valoras con un precio bajo y un total bajo correspondientes al grupo Electrónica y sus Accesorios, con procedencia española almacenados en el Destacamento Catamayo como resultado de un control de rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 5 (V Distrito Azuay):** aprehensiones realizadas en el mes de agosto valoras con un precio bajo y un total medio correspondientes a medias de vestir, nylon, pantys, tobilleras pertenecientes al grupo textiles, con procedencia china almacenados en el Destacamento Chacras como resultado de un control de rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 6 (II Distrito Manta):** aprehensiones realizadas en el mes de octubre valoras con un precio bajo y un total bajo correspondientes al grupo frutas y comestibles, con procedencia extranjera almacenados en el Destacamento Amazonas como resultado de un control de rutina realizado por el Cuerpo de Vigilancia Aduanera.



**Clúster 7 (VII Distrito Puerto Bolívar):** aprehensiones realizadas en el mes de mayo valoras con un precio bajo y un total bajo correspondientes a algodón, hilos, elásticos, encajes pertenecientes al grupo Textiles, con procedencia extranjera almacenados en el Destacamento Tabacundo como resultado de un control de rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 8 (III Distrito Esmeraldas):** aprehensiones realizadas en el mes de julio valoras con un precio bajo y un total medio correspondientes al grupo Electrónica y sus accesorios, con procedencia extranjera almacenados en el Destacamento Chacras como resultado de un control de rutina realizado por el Cuerpo de Vigilancia Aduanera.

**Clúster 9 (VIII Distrito Carchi):** aprehensiones realizadas en el mes de junio valoras con un precio bajo y un total medio correspondientes al grupo Electrónica y sus accesorios, con procedencia extranjera almacenados en la Dirección Distrital Quito como resultado de un Control de Ingreso de Divisas

#### **Conocimiento obtenido del PC1**

**Clúster 0 (X Distrito Latacunga):** aprehensiones de telas del grupo Textiles realizadas por el Cuerpo de Vigilancia Aduanera en Guano – Ecuador.

**Clúster 1 (Distrito Quito IV):** aprehensiones de perfumes del grupo Cosméticos realizadas por el Cuerpo de Vigilancia Aduanera en La Unión – Ecuador.

**Clúster 2 (VII Distrito Puerto Bolívar):** aprehensiones de naranjilla y de productos de grupo Madera realizadas por la Fuerzas Armadas en Nobol – Ecuador.

**Clúster 3 (VI Distrito Loja):** aprehensiones de encendedores y sus accesorios del grupo Explosivos realizadas por el Cuerpo de Vigilancia Aduanera en Catamayo – Ecuador.

**Clúster 4 (III Distrito Esmeraldas):** aprehensiones de calzado Deportivo y de productos del grupo Cigarrillos realizadas por la Policía Nacional en Pindal – Ecuador.

**Clúster 5 (IX Distrito Huaquillas):** aprehensiones de Piñas y productos del grupo Licores realizadas por el Cuerpo de Vigilancia Aduanera en Chacras – Ecuador.

**Clúster 6 (VIII Distrito Carchi):** aprehensiones de galletas del grupo Confitería realizadas por el Cuerpo de Vigilancia Aduanera en el Río Conejo – Ecuador.

**Clúster 7 (I Distrito Guayas):** aprehensiones de material didáctico kilos y productos del grupo Calzado realizadas por el Cuerpo de Vigilancia Aduanera en Cotacachi – Ecuador.

**Clúster 8 (V Distrito Azuay):** aprehensiones de algodón, hilos, elásticos o encajes del grupo Textiles realizadas por el Cuerpo de Vigilancia Aduanera en Chauchin – Ecuador.

**Clúster 9 (II Distrito Manta):** aprehensiones de algodón, hilos, elásticos o encajes del grupo Textiles realizadas por el Cuerpo de Vigilancia Aduanera en Puyo – Ecuador.

### **Conocimiento obtenido del PC2**

**Clúster 0 (V Distrito Azuay):** aprehensiones con un total de categoría bajo origen de un Operativo Conjunto realizado en un Control Conjunto Interinstitucional en la Zona 2.

**Clúster 1 (VIII Distrito Carchi):** aprehensiones con un total de categoría medio origen de Acta de Entrega – Recepción realizado por las Fuerzas Armadas en la Zona 2.

**Clúster 2 (I Distrito Guayas):** aprehensiones con un total de categoría bajo origen de un Control Aeropuerto realizado por el Cuerpo de Vigilancia Aduanera en la Zona 5.

**Clúster 3 (IX Distrito Huaquillas):** aprehensiones con un total de categoría bajo origen de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en la Zona 2.

**Clúster 4 (VI Distrito Loja):** aprehensiones con un total de categoría bajo origen de un Control Móvil realizado por el Cuerpo de Vigilancia Aduanera en la Zona 2.

**Clúster 5 (No Class):** aprehensiones con un total de categoría bajo origen de un Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en la Zona 5.

**Clúster 6 (No Class):** aprehensiones con un total de categoría medio origen de un Control en Carretera realizado por el Cuerpo de Vigilancia Aduanera en la Zona 2.

**Clúster 7 (III Distrito Esmeraldas):** aprehensiones con un total de categoría medio origen de un Control en Carretera realizado en un Control Conjunto Interinstitucional en la Zona 3.

**Clúster 8 (IV Distrito Quito):** aprehensiones con un total de categoría bajo origen de un Operativo Conjunto realizado en un Control Conjunto Interinstitucional en la Zona 5.

**Clúster 9 (VII Distrito Puerto Bolívar):** aprehensiones con un total de categoría bajo origen de Acta de Entrega – Recepción realizado por las Fuerzas Armadas en la Zona 2.

### **Conocimiento obtenido del PC3**

**Clúster 0 (VI Distrito Loja):** aprehensiones de Encendedores y sus accesorios del grupo Explosivos origen de Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en el Chasqui – Ecuador.

**Clúster 1 (V Distrito Azuay):** aprehensiones de algodón, hilos, elásticos o encajes del grupo Textiles origen de Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en Chitan de Navarrete – Ecuador.

**Clúster 2 (IX Distrito Huaquillas):** aprehensiones de brandy del grupo Licores origen de Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en Chapuel – Ecuador.

**Clúster 3 (IV Distrito Quito):** aprehensiones de repuestos y accesorios de motos origen de Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en el Puerto Marítimo – Ecuador.

**Clúster 4 (II Distrito Manta):** aprehensiones de repuestos y accesorios de motos origen de una Inspección de Mercadería realizado en un Control Conjunto Interinstitucional en Gonzamana – Ecuador.

**Clúster 5 (I Distrito Guayas):** aprehensiones de material didáctico kilos y productos del grupo Calzado origen de Control de Ingreso de Divisas realizado por el Cuerpo de Vigilancia Aduanera en Cotopaxi – Ecuador.

**Clúster 6 (VIII Distrito Carchi):** aprehensiones de reproductores de sonidos y productos del grupo Cosméticos origen de Control de Rutina realizado por el Cuerpo de Vigilancia Aduanera en el Río Siete – Ecuador.






**Clúster 7 (VII Distrito Puerto Bolívar):** aprehensiones de secador, alisador, rasurados, cortador, de pelo y productos del grupo Courier origen de Acta de Entrega – Recepción realizado por las Fuerzas Armadas en el Río Siete – Ecuador.

**Clúster 8 (III Distrito Esmeraldas):** aprehensiones de repuestos y accesorios de motos origen de una Acta de Entrega – Recepción realizado por las Fuerzas Armadas en Latacunga – Ecuador.

**Clúster 9 (X Distrito Latacunga):** aprehensiones de reproductores de sonidos y productos del grupo Cosméticos origen de Operativo Conjunto realizado en un Control Conjunto Interinstitucional en Gonzamana – Ecuador.

### ***Atípicos***

La identificación de las características poco frecuentes en las aprehensiones y retenciones que realiza la entidad de control a nivel nacional se realiza inicialmente con la identificación de variables cuantitativas y cualitativas para proceder con la aplicación de diferentes técnicas para el análisis de la información como el análisis de porcentajes para las variables cualitativas y el diagrama de cajas en el caso de las cuantitativas. Se obtuvo como resultado principal de datos atípicos:

-  Aprehensiones con cantidades correspondientes millones.
-  Aprehensiones que registren un valor total estimado superior a los \$60mil, categorizado como "Muy Alto"
-  Aprehensiones y retenciones de productos correspondiente a los grupos de courier, perlas finas piedras y metales preciosos, embarcaciones, madera y productos del mar.
-  Aprehensiones correspondientes a los siguientes subgrupos: productos naturales, oro en bruto, frutas (naranja, maracuyá, mamoncillo, higos, guaba, toronja, piñas, ajonjolí, zapotes), estuches para cd, carbón, canoas, botes, borregos, ultrasonido, tricar, pulpo, mariscos varios, tractores o retroexcavadoras, bandas de cuero, vehículos menaje de casa y registros de datos data logger.
-  Aprehensiones y retenciones que tengan como destino las bodegas Bodega de SAI-SENAE Aeropuerto JJO GYE, Destacamento Saucillo, Unidad de Lavado de Activos de la PPNN, Almacén Temporal Bolivariana, Aretina, Bodega DE SAI-SENAE Aeropuerto MS UIO, Bodega 22 DE LA CAE Puerto Marítimo y Bodega de Aduana-Pto. De Manta

- 🚧 Aprehensiones realizadas por Agrocalidad, ARCSA e Inspectoría de Pesca
- 🚧 Aprehensiones realizadas en los distritos de Esmeraldas y Manta.
- 🚧 Aprehensiones en la Zona 3 con respecto al resto de zonas.

### **3.4. Resumen Ejecutivo del conocimiento obtenido**

Los resultados obtenidos de la aplicación de las técnica de asociación con la aplicación de los algoritmos Apriori y FP Growth reflejan como patrones principales las aprehensiones correspondientes a la Zona 2 para el Distrito Huaquillas ubicado en sur del territorio nacional, en los cuales la mercadería que se intenta ingresar por contrabando pertenece al grupo Textiles con subgrupo Prendas de Vestir nuevas y Licores, mismas aprehensiones corresponden a la categoría Bajo con respecto a su total, por lo que se puede apreciar que la información no es muy representativa.

Los patrones obtenidos para el caso de los algoritmos KMeans y EM correspondientes a la técnica de agrupamiento permiten identificar los puntos geográficos donde se realizan aprehensiones y retenciones con para cada distrito como lo es Chacras-Ecuador, Guayaquil-Ecuador, Rumichaca-Ecuador, Chitan de Navarrete-Ecuador, Chasqui-Ecuador ,Puerto Marítimo-Ecuador, Gonzamana-Ecuador, Cotopaxi-Ecuador, Rio Siete-Ecuador, Chapuel-Ecuador y Latacunga-Ecuador permitiendo identificar puntos estratégicos para futuros controles y gestión logística del personal encargado, de igual manera se identificó que los meses de julio y junio son representativos para los distritos X , IV ,I , III Y VIII por el ingreso de Accesorios de Vehículos , Accesorios de celular, Confitería, y mercadería correspondiente a Electrónica y sus Accesorios respectivamente; mientras que para los meses de abril, mayo, agosto y octubre los ingresos en los distritos IX, VII, V y II corresponden a Electrónica y sus accesorios, algodón, hilos, elásticos y encajes del grupo Textiles, medias de vestir, nylon, pantys y tobilleras y Frutas y Comestibles respectivamente.

Al haber aplicado dos algoritmos en subconjuntos del conjunto de datos total de información permitió identificar otros patrones que no difieren en su totalidad de los antes mencionados pero agregan detalles relevantes para la identificación de nuevos posibles patrones de contrabando

como por ejemplo para los distritos IV , II, y III el ingreso de mercadería correspondiente a Repuestos y Accesorios de Motos, para el caso de los distritos VIII y X el ingreso de mercadería representativa corresponde a Reproductores de sonido y Cosméticos mientras que para los distritos VI,V, IX, I, VII las aprehensiones corresponder a Encendedores y sus accesorios, Algodón, hilos, elásticos o encajes ,Brandy del grupo licores, Material didáctico y Secador, alisador, rasurados, cortador, etc. de pelo respectivamente.

Además, se pudo identificar que las aprehensiones y retenciones son realizadas por el Cuerpo de Vigilancia Aduanera en su mayor porcentaje y el restante corresponde a acciones de Fuerzas Armadas y Policía Nacional.

Finalmente con la técnica de detección de atípicos mediante la aplicación de un análisis estadístico se pudo identificar que las aprehensiones y retenciones poco frecuentes dentro del conjunto de datos siendo aquellas que su total corresponden valores superiores a \$60 mil mismo que encaja en la categoría de Muy Alto, además las aprehensiones donde la mercadería ingresada corresponder a los grupos courier, perlas finas piedras y metales preciosos, embarcaciones, madera y productos del mar y con ello los siguientes subgrupos: productos naturales, oro en bruto, frutas (naranja, maracuyá, mamoncillo, higos, guaba, toronja, piñas, ajonjolí, zapotes), estuches para cd, carbón, canoas, botes, borregos, ultrasonido, tricar, pulpo, mariscos varios, tractores o retroexcavadoras, bandas de cuero, vehículos menaje de casa y registros de datos data logger. Igualmente se pudo identificar que las aprehensiones a cargo de Agrocalidad, ARCSA e Inspectoría de Pesca son poco frecuentes asimismo como las acciones realizadas en los distritos Esmeraldas y Manta.

### **3.5. Análisis de impacto**

Con los resultados que las técnicas descriptivas de minería de datos (clustering, asociación y atípicos) permiten identificar se puede realizar un siguiente proceso el cual permite que todo el conocimiento obtenido sea de utilidad para los directivos y personal de la entidad de control para llevar a cabo una mejor organización de sus recursos para así poder combatir el delito de contrabando que aqueja a la sociedad en la cual nos desarrollamos actualmente.

En análisis de impacto que se propone en la presente sección se lo realizo en base la matriz de impactos misma que permite la identificación de aspectos positivos como negativos en el proyecto en la ejecución de las tareas de vigilancia y control de ingreso de mercadería por contrabando. En la Tabla 68 a continuación se aprecia los niveles de impacto considerados.

Tabla 68 - Niveles de Impacto

NIVELES DE IMPACTO	PONDERACIÓN
Impacto Alto Positivo	3
Impacto Medio Positivo	2
Impacto Bajo Positivo	1
Punto de Indiferencia	0
Impacto Bajo Negativo	-1
Impacto Medio Negativo	-2
Impacto Alto Negativo	-3

Fuente: Posso,2013

Para el análisis en este caso se toma en cuenta los ámbitos, sociocultural, económico y tecnológico que se muestran en las Tablas 69, 70 y 71 a continuación seguido del impacto general del proyecto.

Tabla 69 - Impacto Sociocultural

Indicador	-3	-2	-1	0	1	2	3
Calidad de vida y seguridad ciudadana							x
Cumplimiento de normativa						x	
Empleo							x
TOTAL						2	6

$$Nivel\ de\ impacto = \frac{\Sigma}{N\acute{u}mero\ de\ indicadores}$$

$$Nivel\ de\ impacto = \frac{8}{3} = 2.667$$

Nivel de Impacto Sociocultural = Alto Positivo

Fuente: Posso, 2013

El impacto sociocultural se considera alto positivo, puesto que la calidad de vida de los habitantes de las fronteras mejora al confiar en que la seguridad está presente con mayor control por parte de la entidad de control de los puntos críticos y estratégicos a nivel nacional.

En cuanto a la normativa vigente para este delito se podrá contribuir con el cumplimiento de esta por parte de las autoridades logrando estimar las posibilidades de ingreso de mercadería por contrabando de una manera más certera y acorde a los valores que los algoritmos presentan.

El desempleo dentro del país genera que los miembros de la sociedad busquen una forma de ingresos incurriendo en delitos como el del contrabando, por ello con la identificación de los puntos críticos generará mayores posibilidades de trabajo, con el apoyo de las entidades que deben involucrarse en ello.

Tabla 70 - Impacto Económico

Indicador	-3	-2	-1	0	1	2	3
Productividad por parte de la entidad							x
Uso de recurso económico							x
Uso de recurso humano							x
TOTAL							9

$$\text{Nivel de impacto} = \frac{\Sigma}{\text{Número de indicadores}}$$

$$\text{Nivel de impacto} = \frac{9}{3} = 3$$

Nivel de Impacto Sociocultural = Alto Positivo

Fuente: Posso,2013

En el ámbito económico se considera que tendrá un impacto alto positivo puesto que la productividad de los sectores comerciales ecuatorianos aumentará gracias al control y mitigación del ingreso de mercadería de forma ilegal, además con la correcta gestión y manejo de la información el análisis de los datos permitirá aumentar la productividad de los componentes que conforman la entidad.

Con respecto al uso del recurso económico designando para la entidad de control se tiene un impacto alto positivo puesto que la identificación de los puntos críticos permitirá que recursos de movilización, gestión, controles y demás actividades se los organice bajo una mejor organización permitiendo que se trate de disminuir la gran incurrancia en el delito del contrabando en el Ecuador.

Con lo que respecta al uso de los recursos humanos disponibles en la entidad de control se tiene un impacto alto positivo puesto que permitirá poner en marcha una logística que permitirá un control más específico de los puntos críticos con el propósito de evitar que en ingreso de mercadería por contrabando sea controlado desde los puntos invisibles hasta lo más significativo presente.



Tabla 71 - Impacto Tecnológico

Indicador	-3	-2	-1	0	1	2	3
Automatización de procesos							x
Innovación							x
Desempeño del personal							x
<b>TOTAL</b>							<b>9</b>

$$\text{Nivel de impacto} = \frac{\Sigma}{\text{Número de indicadores}}$$

$$\text{Nivel de impacto} = \frac{9}{3} = 3$$

Nivel de Impacto Sociocultural = Alto Positivo

Fuente: Posso,2013

El impacto tecnológico se lo considera como alto positivo gracias a las ventajas de proveer un sistema en línea para la realización de los procesos correspondientes a aprehensiones y retenciones con un registro exitoso de la información, además el desempeño del personal aumentara gracias a las facilidades que brinda un software especializado en su labor.

Impacto General							
Indicador	-3	-2	-1	0	1	2	3
Impacto Sociocultural							x
Impacto Económico							x
Impacto Tecnológico							x
<b>TOTAL</b>							<b>9</b>

$$\text{Nivel de impacto} = \frac{\Sigma}{\text{Número de indicadores}}$$

$$\text{Nivel de impacto} = \frac{9}{3} = 3$$

Nivel de Impacto Tecnológico = Alto Positivo

Fuente: Posso,2013

El impacto general es alto positivo generando expectativas para promover la toma de decisiones en base al conocimiento generando donde el control de ingreso de mercadería por contrabando beneficie toda la sociedad ecuatoriana, especialmente a los productores ecuatorianos quienes en su mayoría suelen verse afectados por el ingreso de mercadería extranjera que les quita su mercado y comercialización normal del producto nacional.

## LIMITACIONES

- La información base proporcionada por la entidad de control es de carácter confidencial debido a que es un proyecto de cooperación entre instituciones públicas.
- El periodo de tiempo de recolección de la información a analizar comprende los años 2014 - 2019.
- El conjunto de datos entregado por la entidad de control no tiene un número significativo de variables para la aplicación de técnicas de minería de datos a profundidad.
- Los datos almacenados en el documento de Excel proporcionado por la entidad de control no presentaban ningún tipo de normalización de la información, además no se controla el uso de mayúsculas o minúsculas además de los caracteres especiales por lo que crean gran cantidad de redundancia en la información por errores de tipeo.
- La información no incluyó elementos claves para la obtención de conocimiento de mayor relevancia, en el caso de este tipo de información se consideraría como necesario registrar los datos del individuo que comete el delito, especificaciones del vehículo en el que se moviliza, al igual que la información específica del sitio de la aprehensión para su ubicación geográfica posterior.

## CONCLUSIONES

La aplicación de técnicas descriptivas de minería de datos con información de aprehensiones y retenciones a causa del delito de contrabando en todo el territorio ecuatoriano permitió identificar un conjunto de 10 clústers que agrupan la información de dichas aprehensiones, cada uno de ellos corresponde respectivamente a los Distritos definidos por la entidad de control para la gestión de sus actividades.

En el desarrollo del proyecto se pudo constatar la nula existencia de trabajos relacionados con el área de minería de datos y delitos de contrabando que refuercen la base teórica y aplicación práctica, porque se buscó trabajos con temáticas similares y centrados uso de la minería de datos para la identificación de patrones para: seguridad ciudadana, gestión bancaria, técnicas de aprendizaje, entre otros.

La gestión de los datos entregados correspondientes a aprehensiones y retenciones por parte de la entidad de control no refleja una estructura robusta de la información por la poca especificidad que esta posee, de esta manera se puede constatar que un conjunto de datos de un volumen (instancias) considerable sin un grupo numeroso y estratégico de variables no refleja resultados específicos al momento de la aplicación de técnicas de minería de datos.

Los productos resultantes del presente trabajo son el set de datos resultado de la aplicación de las fases de selección, limpieza y transformación, además del conjunto de patrones obtenidos mediante las técnicas descriptivas de agrupamiento, asociación y atípicos. Finalmente se tiene los productos correspondientes a la forma de ingreso de datos en la herramienta Microsoft Excel que aborda los puntos estratégicos de la información a considerar para futuros trabajos de minería de datos en este campo y los tableros elaborados con la herramienta Microsoft Power BI.

Los resultados obtenidos de las técnicas de asociación con la aplicación de los algoritmos Apriori y FP Growth reflejan patrones donde se evidencia que las principales aprehensiones son correspondientes a la Zona 2 para el Distrito Huaquillas ubicado en sur del territorio nacional, en los cuales la mercadería que se intenta ingresar por contrabando pertenece al grupo Textiles y Licores mismas que se categorizan como Bajas con respecto a su total.

Los patrones obtenidos para el caso de los algoritmos KMeans y EM correspondientes a la técnica de agrupamiento permitieron identificar los puntos geográficos estratégicos donde se realizan aprehensiones y retenciones, además se identificó que los meses de julio y junio son representativos para los distritos I, III, IV, VIII, y X mientras que para los meses de abril, mayo, agosto y octubre los distritos IX, VII, V y II respectivamente presentan ingresos de mercadería por contrabando representativo.

Además, se pudo identificar que las aprehensiones y retenciones son realizadas por el Cuerpo de Vigilancia Aduanera en su mayor porcentaje y el restante corresponde a acciones de Fuerzas Armadas y Policía Nacional.

Por otro lado la técnica de detección de atípicos mediante el análisis estadístico permitió identificar las aprehensiones y retenciones poco frecuentes dentro del conjunto de datos, siendo aquellas que su total corresponden a valores de alrededor USD 60 000 mismo que encaja en la

categoría de Muy Alto, además las aprehensiones donde la mercadería ingresada corresponder a los grupos courier, perlas finas piedras y metales preciosos, embarcaciones, madera y productos del mar y con ello los varios subgrupos.

Finalmente se pudo identificar que las aprehensiones a cargo de Agrocalidad, ARCSA e Inspectoría de Pesca son poco frecuentes asimismo como las acciones realizadas en los distritos Esmeraldas y Manta.

## RECOMENDACIONES

Dentro del desarrollo de un proyecto de minería de datos la información con la cual se trabaja debe contener la suficiente especificidad y atributos para poder aplicar las diferentes técnicas descriptivas propuestas y obtener patrones más relevantes, por lo que se recomienda gestionar la información mediante un proceso automático con el uso de un formulario de ingreso o un sistema que permita evitar errores digitación, interpretación o mal uso de caracteres.

Con respecto a los atributos que se registra al momento de reportar las aprehensiones se recomienda registrar la *información del individuo* implicado en el delito de contrabando lo que corresponde a género, etnia, edad, nacionalidad, estado civil, discapacidad y antecedentes, además se recomienda el registro de la información del vehículo en el que se transporta la mercadería por contrabando siendo los sugeridos los atributos tipo, modelo, número de placa (letra inicial), numero de ejes, tipo de combustibles, color y capacidad de carga (TM).

Por otro lado, con lo que respecta al registro de atributos es esencia registrar la información de geolocalización del sitio donde ocurrió la aprehensión para poder sustraer la información a nivel más específico con los respecta a zona, provincia, y cantón; con todo lo antes mencionado los alcances de los resultados de la minería de datos podrán ser más profundos y de mayor relevancia para la entidad de control.

Se recomienda que la entidad de control se encargue de la identificación de los puntos críticos de su labor para que puedan sugerir a las autoridades encargadas el uso recursos tecnológicos con el propósito de gestionar y optimizar los recursos tanto humanos como económicos en cada una de las aprehensiones y retenciones para que de esta manera tanto los vigilantes y responsables encargados puedan ser más productivos en cada tarea encomendada.

El análisis del recurso de datos históricos de la información base debe ser considerado como punto clave para la planificación de las actividades zonales como distritales para que se pueda generar un modelo posible de las circunstancias a las cuales cada punto crítico debe estar preparado.

## ANEXOS

### Anexo 1

#### Carta de confidencialidad entre las partes

	<b>Acuerdo de Confidencialidad (Información de Propiedad del Cliente y de la Firma)</b>	Código: CC01
		Versión: 0.1
		Fecha de vigencia: 2019-2020

### ACUERDO DE CONFIDENCIALIDAD

Intervienen en la celebración del presente Acuerdo de Confidencialidad, por una parte, el Ing. MARCELO VILLACRÉS, como representante de la Coordinación del Centro de Formación de Vigilancia Aduanera (CFVA) y por otra parte: la Srta. DIANA CAROLINA ROSERO REA y Sr. TOMMY BRYAN MANCERO MENOSCAL, tesistas de la Carrera de Sistemas Computacionales de la Universidad Técnica del Norte (UTN) y el PhD. IVÁN GARCÍA SANTILLÁN, Docente-Investigador de la Carrera de Software de la UTN, quienes intervienen como el equipo responsable en la elaboración y entrega del **Proyecto de análisis de datos** para el Centro de Formación de Vigilancia Aduanera.

#### PRIMERA.- ANTECEDENTES:


1.1.-La Constitución de la República, en el artículo 227 prescribe que la administración pública constituye un servicio a la colectividad que se rige, entre otros, por el principio de transparencia.

1.2.- La Coordinación del Coordinación del Centro de Formación de Vigilancia Aduanera, en base a lo determinado en el Reglamento Interno para el CFVA, específicamente en el Art. 20 establece las atribuciones del Coordinador del Centro, concretamente en el numeral 15).- *"Propender a la vinculación y buenas relaciones con instituciones de educación de nivel superior y Escuelas de Formación de la Fuerzas Armadas y Policía Nacional, mediante el desarrollo de actividades académicas, culturales, sociales y deportivas";*

1.3.- Con Oficio N° CVA-CFVA-JAC-2019-026, suscrito por el Coordinador del Centro de Formación de Vigilancia Aduanera, que en su parte pertinente manifiesta: *"(...)Además en base a conversación mantenida con su autoridad referente a mantener una coordinación y desarrollo de actividades académicas de vinculación, específicamente al solicitarle y proponerle que estudiantes de su facultad puedan desarrollar un esquema de base de datos y visualización de resultados de las capturas y aprehensiones realizadas por el personal del Cuerpo de Vigilancia Aduanera, (...).*

#### SEGUNDA.- OBJETIVO DEL ACUERDO:

El presente acuerdo se refiere a conservar la confidencialidad de la información y los papeles de trabajo del cliente (CFVA) y protegerlos del acceso y divulgación no autorizados, en el entendido de que ninguna de las partes tome ventaja o se aproveche al recibir la información considerada como clasificada de acuerdo a los términos de este documento, con el objetivo de colaborar el proyecto denominado Proyecto de análisis de datos en el Centro de Formación de Vigilancia Aduanera (CFVA) a través de trabajos de titulación de grado.

	<b>Acuerdo de Confidencialidad (Información de Propiedad del Cliente y de la Firma)</b>	Código: CC01
		Versión: 0.1
		Fecha de vigencia: 2019-2020

### TERCERA.- USUARIOS:

#### **Cliente:**

- Ing. Marcelo Villacrés (CFVA)

#### **Responsables:**

- Ph.D. Iván García Santillán (UTN)
- Srta. Diana Carolina Rosero Rea (UTN)
- Sr. Tommy Bryan Mancero Menoscal (UTN)

### CUARTA.- CONFIDENCIALIDAD DE LA INFORMACIÓN Y

#### RESPONSABILIDADES:

La carta de confidencialidad es un acuerdo entre los usuarios del proyecto, es decir, el cliente (CFVA) y los tesistas de la Carrera de Sistemas Computacionales de la UTN, teniendo como fin defender ante eventuales amenazas, referentes al mal uso de la información que se entrega dentro del proceso de análisis de datos y sus consecuencias en caso de ser divulgada o entregadas a terceros sin la autorización del cliente y los responsables del proyecto.

Los papeles de trabajo, informes, archivos y documentos elaborados por los tesistas son confidenciales, por lo cual deben protegerse de accesos no autorizados, tanto los archivos que reposan donde el cliente, como la información que se maneja dentro del proceso del encargo, por lo cual, los equipos deben mantener la seguridad que indique el cliente, como claves y copias de seguridad de la información.

Los papeles de trabajo no deben entregarse a terceros, a menos que los responsables y el cliente lo autoricen por escrito o verbal, lo exijan dentro de un proceso legal o judicial o alguna regulación o norma exija su revelación. En los casos legales, no deben ser entregados hasta tanto el abogado del cliente no dé su aceptación por escrito.

Los responsables del proyecto podrán difundir los resultados obtenidos en el estudio, únicamente, con fines académicos y de investigación, conservando el anonimato de los datos procesados.



	<b>Acuerdo de Confidencialidad (Información de Propiedad del Cliente y de la Firma)</b>	Código: CC01
		Versión: 0.1
		Fecha de vigencia: 2019-2020

El presente Acuerdo entrará en vigor en el momento de la firma del mismo por ambas partes, extendiéndose su vigencia hasta un plazo de 2 años después de finalizada la relación entre las partes o, en su caso, la prestación del servicio.

Para constancia, las partes firman el presente Acuerdo de Confidencialidad en 3 ejemplares, en la ciudad de Ibarra a los cinco días del mes de diciembre de 2019.

*Diana Rosero*

Srta. Diana Rosero  
CI:1003560222  
Tesisista UTN

*Tommy Mancero*

Sr. Tommy Mancero  
CI:1003973813  
Tesisista UTN

*Iván García*

PhD. Iván García  
CI:1002292603  
Docente Tutor UTN



*Marcelo Villacrés*

Ing. Marcelo Villacrés  
CI: 1708862774  
Representante CFVA





Anexo 2  
Preguntas del negocio para análisis exploratorio (BI)

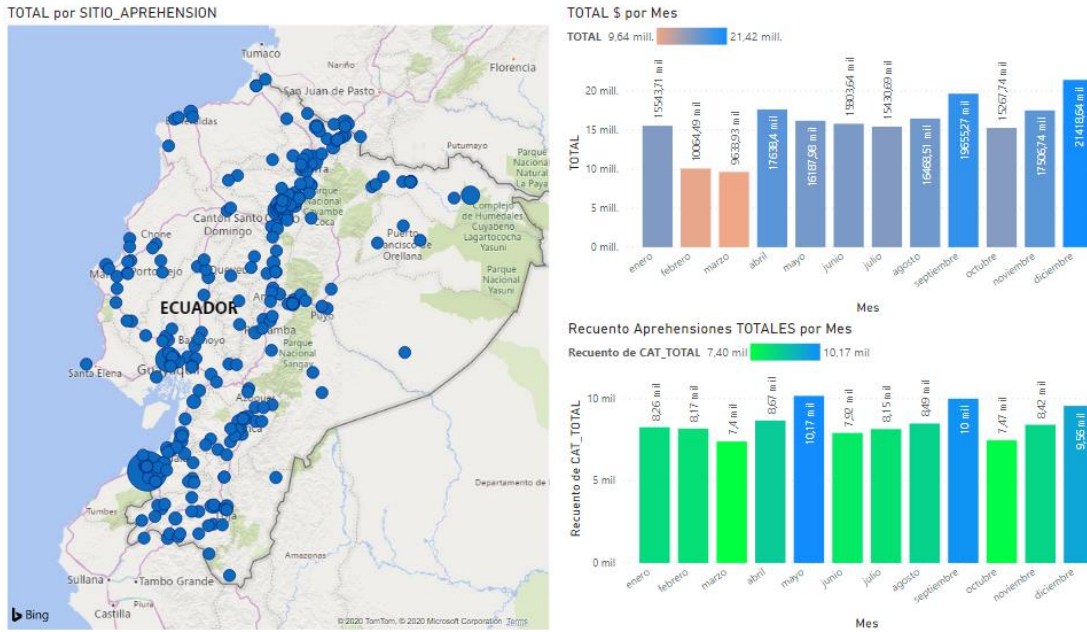
- 1) ¿Cuáles son los meses en los que se encuentra mayor y menor cantidad de aprehensiones y monto económico reunido?
- 2) ¿Cuál es el año cuenta con más aprehensiones en registros?
- 3) ¿Cuáles son las cantidades más altas por grupo que se ha aprehendido, por año, mes, zona?
- 4) ¿Cuáles son las unidades que se presentan con más frecuencia en las aprehensiones en cada zona, distrito?
- 5) ¿Cuáles son los precios máximos y mínimos que se han presentado en las aprehensiones?
- 6) ¿Cuál es el monto total aprehendido anual, semestral, mensual correspondiente a cada zona, distrito y destacamento?
- 7) ¿Cuáles son los grupos que se presentan con mayor frecuencia en las aprehensiones por zona, distrito según registros anuales, trimestrales y mensuales?
- 8) ¿Cuáles son los subgrupos que se presentan con mayor frecuencia en las aprehensiones por zona, distrito según registros anuales, trimestrales y mensuales?
- 9) ¿Cuáles son los países de procedencia con más registros de aprehensiones por zona, distrito y destacamento según registros anuales, trimestrales y mensuales?
- 10) Marcas más reconocidas en el contrabando
- 11) ¿Cuáles son los orígenes de aprehensiones más frecuentes por zona, distrito y destacamento según registros anuales, semestrales y mensuales?
- 12) ¿Cuáles son los sitios de aprehensión con mayores registros y con monto económico aprehendido?
- 13) ¿Cuáles son los grupos operativos que realizan con mayor frecuencia las aprehensiones por zona, distrito y destacamento según registros anuales, trimestrales y mensuales?

### Anexo 3

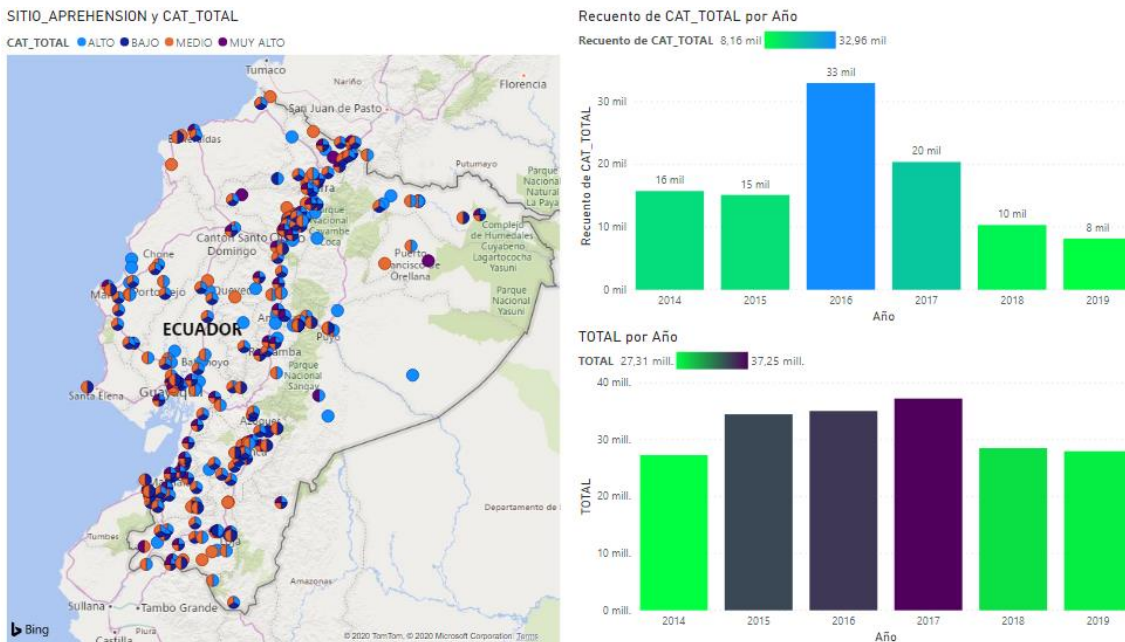
#### Dashboards para responder preguntas del negocio propuesta en el Anexo2

El conjunto de imágenes que se muestran a continuación se los elaboro con la herramienta de Bussines Intelligence Microsoft Power BI especializada para el análisis de la información del negocio y la toma de decisiones.

#### Pregunta 1:



#### Pregunta 2:





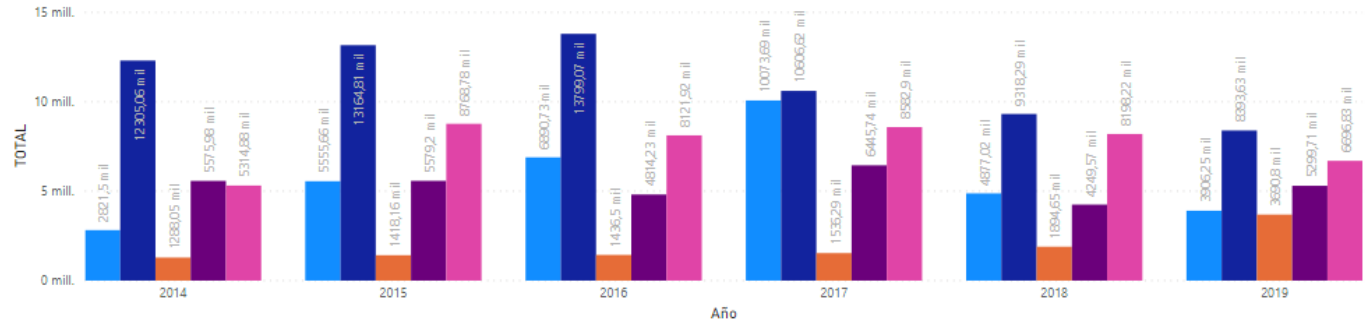
### Pregunta 5:



### Pregunta 6:

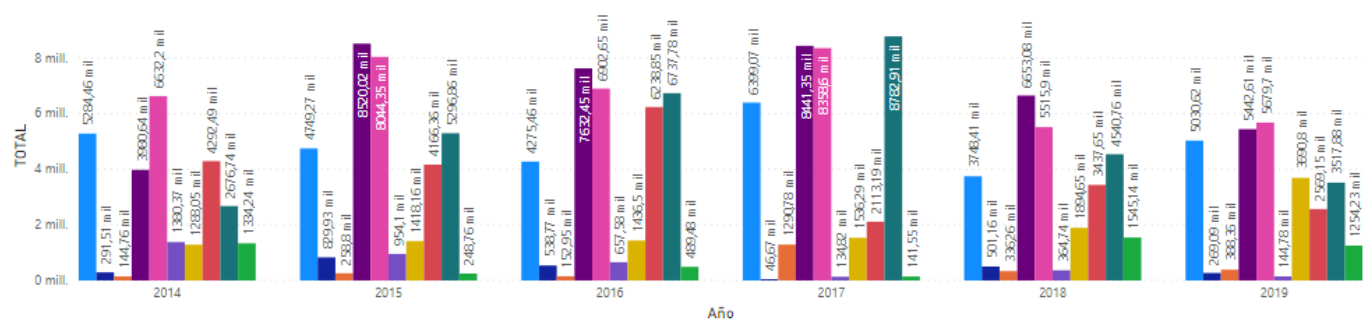
TOTAL por Año y ZONA

ZONA ● ZONA1 ● ZONA2 ● ZONA3 ● ZONA4 ● ZONA5



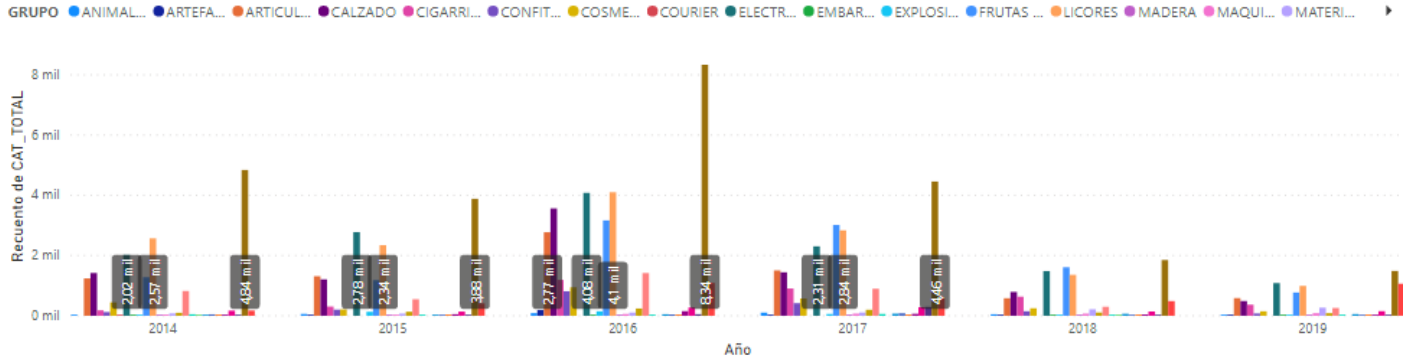
TOTAL por Año y DISTRITO

DISTRITO ● IDG ● IIDM ● IIIE ● IVDQ ● IXDH ● VDA ● WDL ● VIIDP8 ● VIIICD ● XDLAT

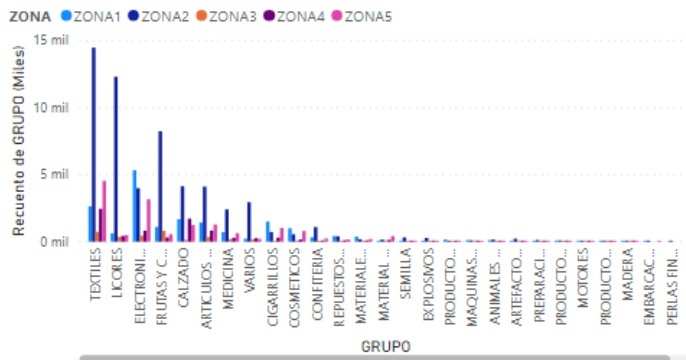


### Pregunta 7:

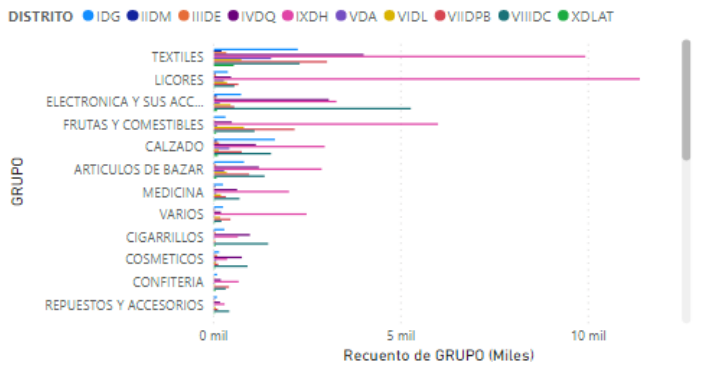
Recuento de CAT\_TOTAL por Año y GRUPO



Recuento de GRUPO por GRUPO y ZONA

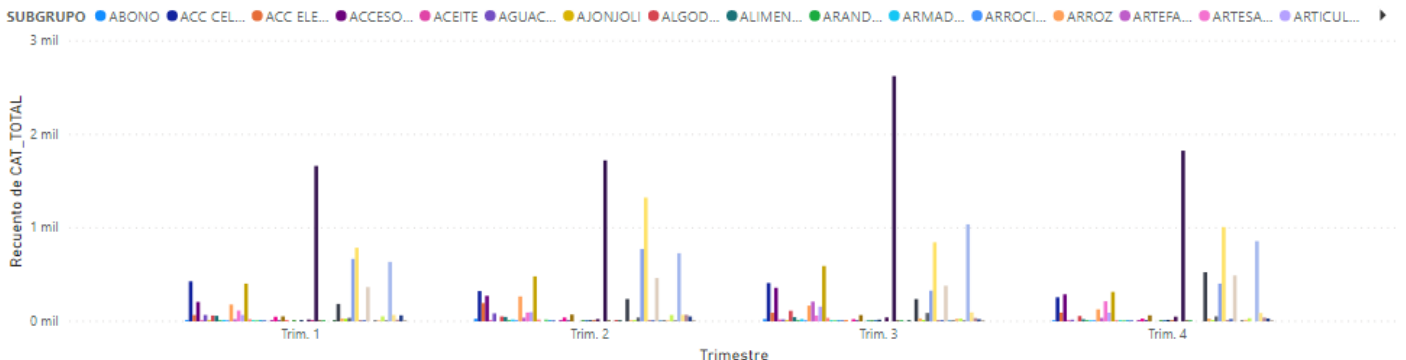


Recuento de GRUPO por GRUPO y DISTRITO

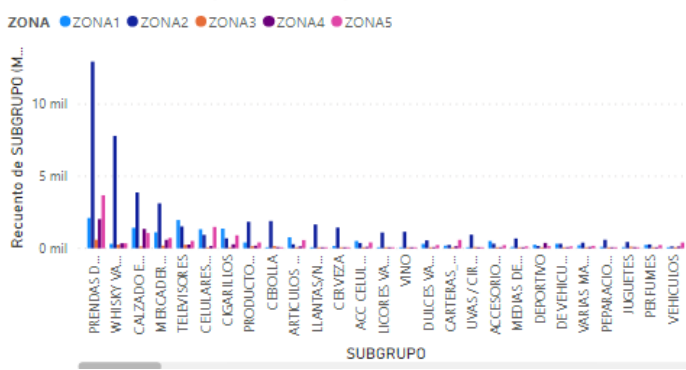


### Pregunta 8:

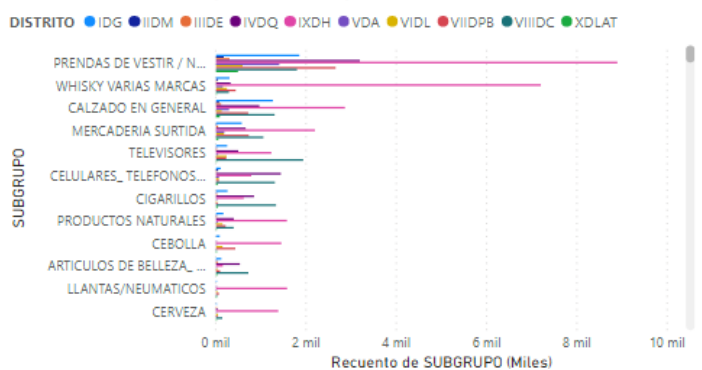
Recuento de CAT\_TOTAL por Trimestre y SUBGRUPO



Recuento de SUBGRUPO por SUBGRUPO y ZONA

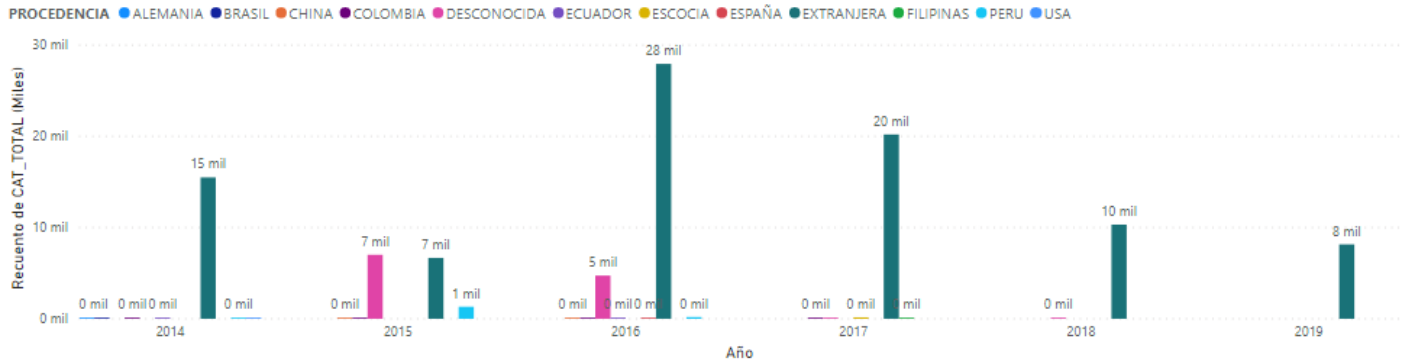


Recuento de SUBGRUPO por SUBGRUPO y DISTRITO

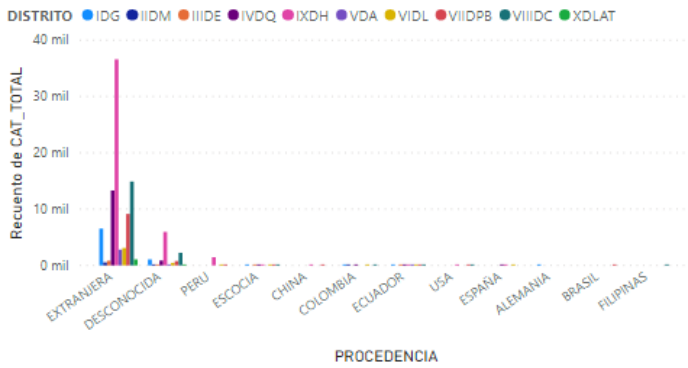


### Pregunta 9:

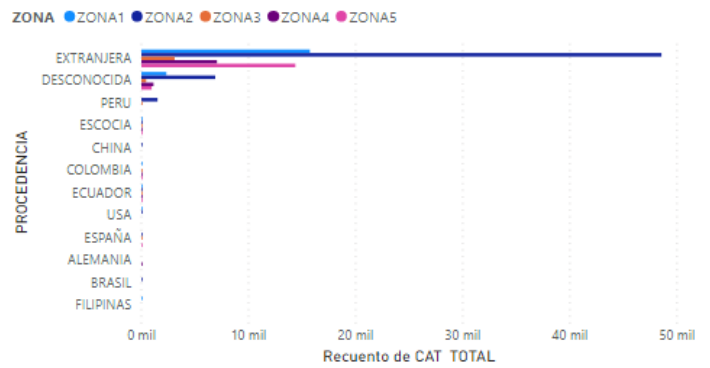
Recuento de CAT\_TOTAL por Año y PROCEDENCIA



Recuento de CAT\_TOTAL por PROCEDENCIA y DISTRITO

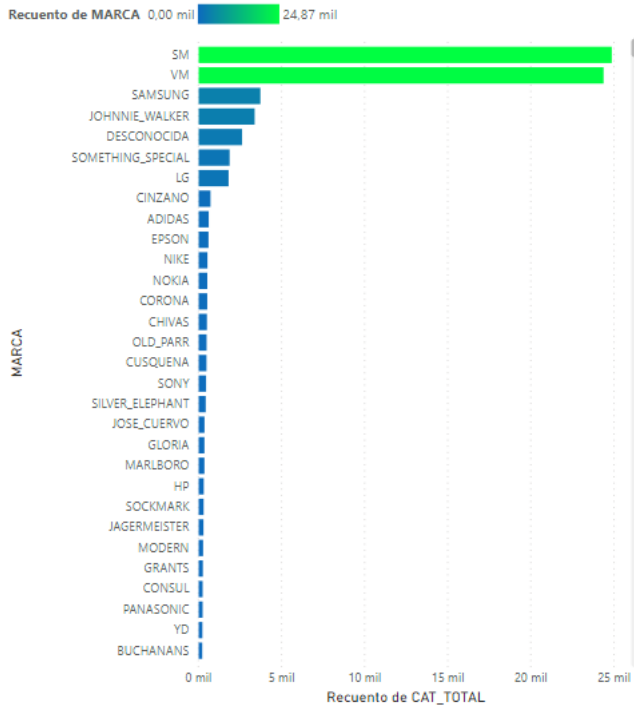


Recuento de CAT\_TOTAL por PROCEDENCIA y ZONA

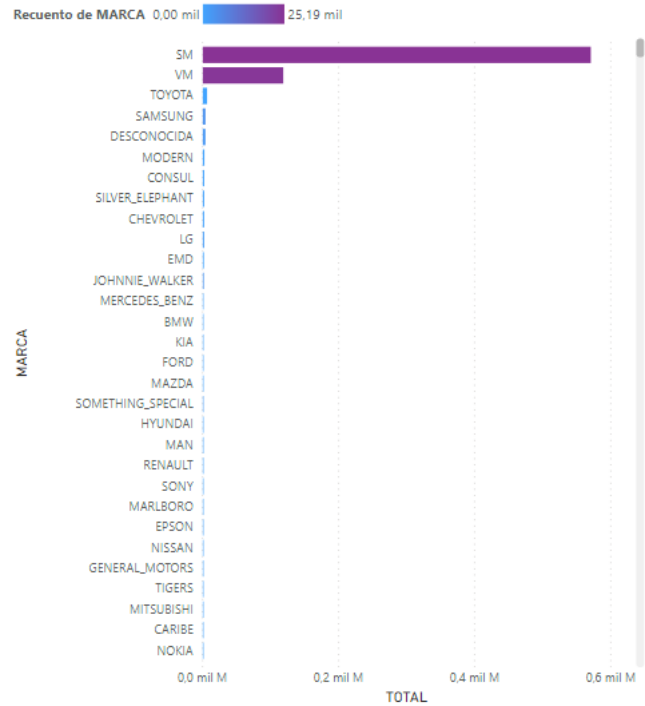


### Pregunta 10:

Recuento de CAT\_TOTAL por MARCA



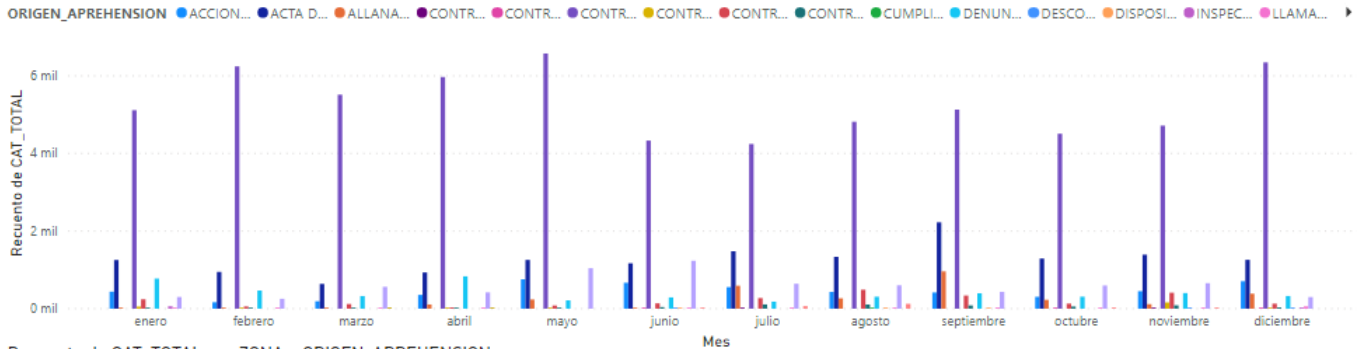
TOTAL por MARCA



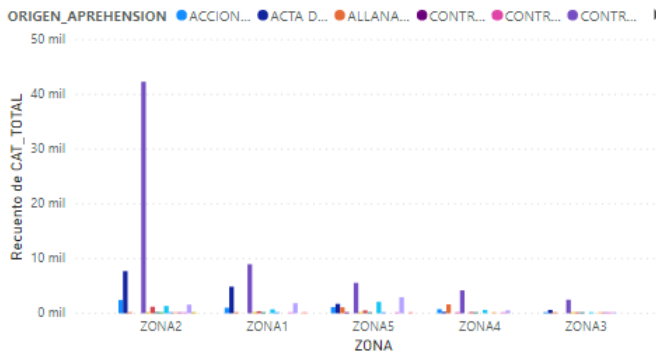


### Pregunta 11:

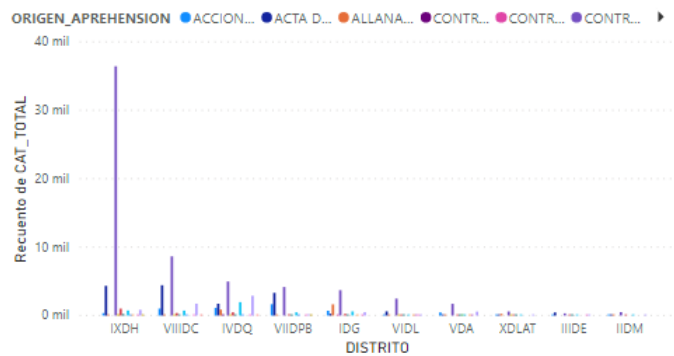
Recuento de CAT\_TOTAL por Mes y ORIGEN\_APREHENSION



Recuento de CAT\_TOTAL por ZONA y ORIGEN\_APREHENSION

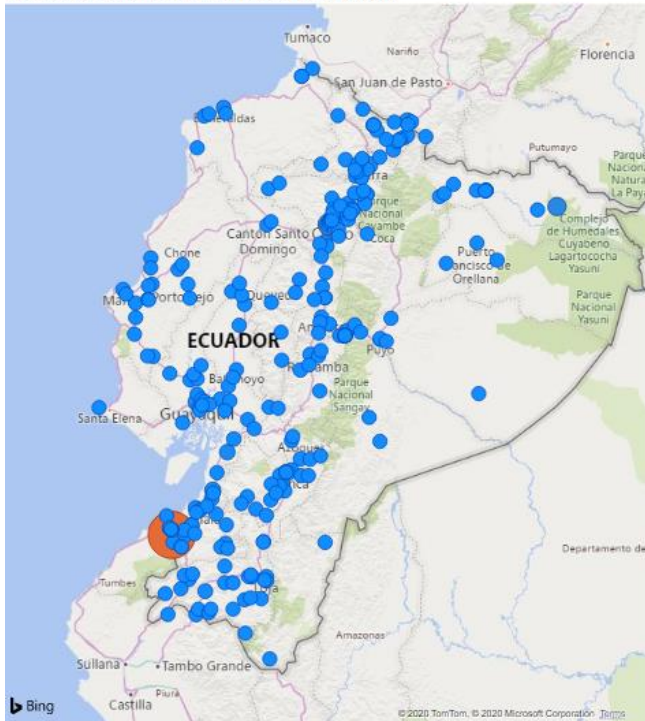


Recuento de CAT\_TOTAL por DISTRITO y ORIGEN\_APREHENSION

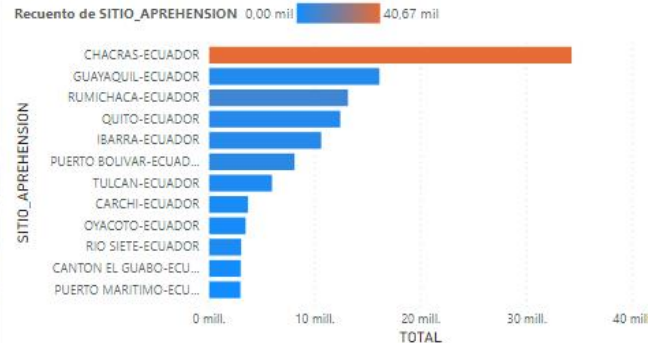


### Pregunta 12:

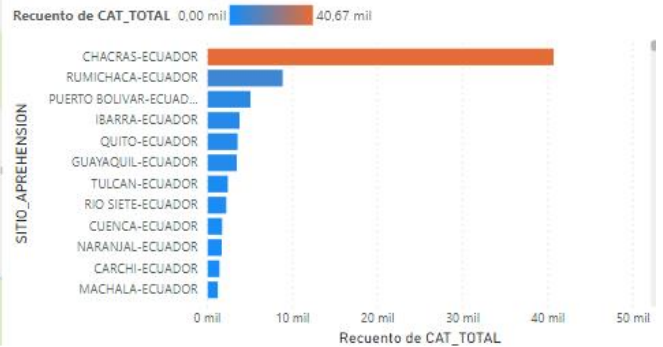
Recuento de CAT\_TOTAL por SITIO\_APREHENSION



TOTAL por SITIO\_APREHENSION

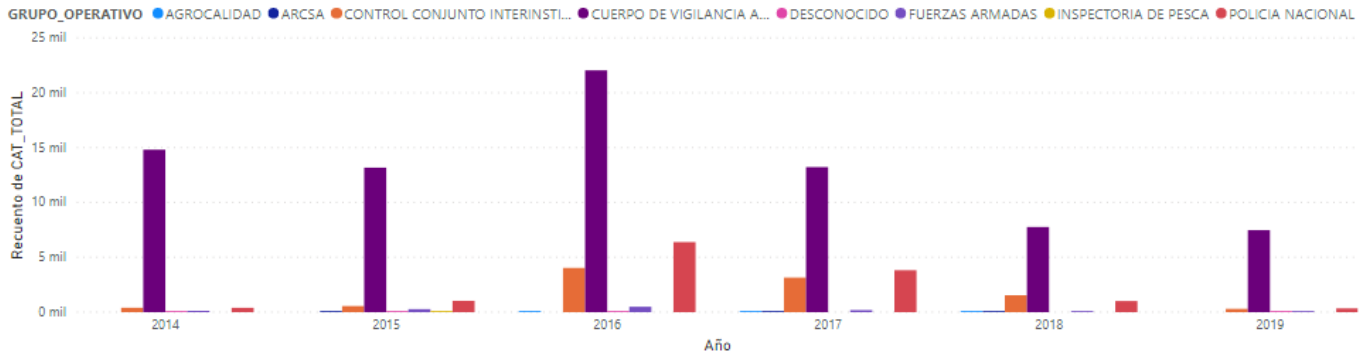


Recuento de CAT\_TOTAL por SITIO\_APREHENSION

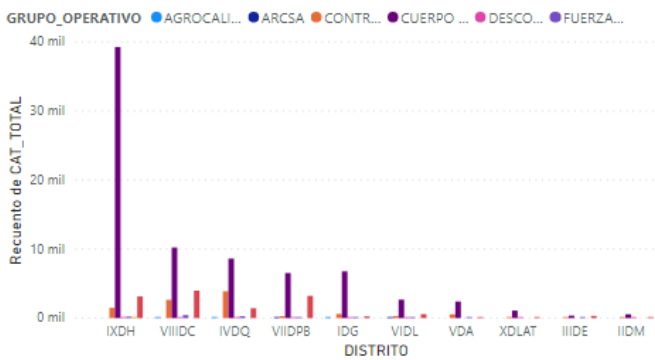


### Pregunta 13:

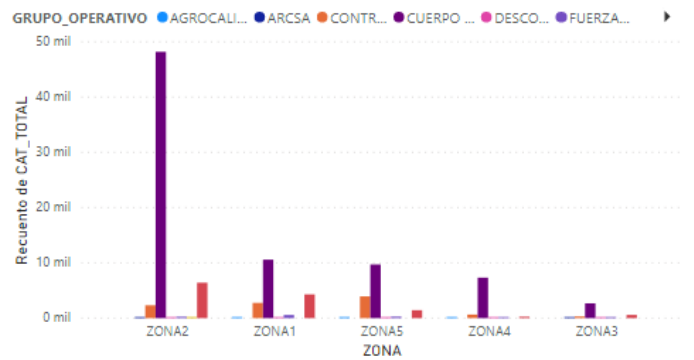
Recuento de CAT\_TOTAL por Año y GRUPO\_OPERATIVO



Recuento de CAT\_TOTAL por DISTRITO y GRUPO\_OPERATIVO



Recuento de CAT\_TOTAL por ZONA y GRUPO\_OPERATIVO





Anexo 4  
Formato sugerido para ingreso de datos (Macro Excel)

REGISTRO DE APREHENSIONES				
		Fecha	13/06/2020	
Numero	CVAT-OPE-AA-2020-0001	Hora	12:01:00	
Producto				
Cantidad	Unidades	Precio	Total	
3,00	Unidad	\$450,00	\$1.350,00	
Grupo	SubGrupo	Descripcion	Status	Marca
Electronica y sus Accesorios	Televisores	Televisor	Nuevo	LG
CVA				
Distrito	Destacamento	Origen Aprehensión	Grupo Operativo	Bodega
VIII DC	Carchi	Patrullaje de rutina (Orden de patrulla)	Turno de Guardia	Dirección Distrital Tulcan
Sitio				
Zona	Provincia	Cantón	Latitud	Longitud
Zona1	Esmeraldas	Cantón Esmeraldas	0.9592	-79.65397
Vehículo				
Tipo	Modelo	Nro. Placa (Letralnicial)	Nro. Ejes	Combustible
Sedán	2015	C (Carchi)	Automóviles, todoterrenos o camionetas	Gasolina
Color	Capacidad Carga (TM)			
Plata	1TM			
Infractor				
Género	Etnia	Edad (años)	Nacionalidad	Estado Civil
Masculino	Mestizo	33	Ecuatoriano	Casado
Discapacidad	Antecedente Penal			
No	No			
GUARDAR REGISTRO		LIMPIAR REGISTRO		

## REFERENCIAS

- Adriana, T., Morales, A., Lucía Rodríguez, T. M., Jorge, I., & Rodríguez Rodríguez, E. (2005). *Prototipo de software para la asociación de datos "UDAssociate."* 567–588. [http://cidc.udistrital.edu.co/investigaciones/documentos/revistacientifica/rev7/Unidad 28 pags 567-588.pdf](http://cidc.udistrital.edu.co/investigaciones/documentos/revistacientifica/rev7/Unidad%20pags%20567-588.pdf)
- Aduana del Ecuador, . (2020). *Quiénes somos UVA*. <https://www.aduana.gob.ec/quienes-somos-uva/>
- Aguirre Yacup, N., & Walteros Alcázar, M. A. (2018). Identificación De Patrones Delictivos En Colombia Durante El Periodo 2010-2016 Mediante El Uso De Técnicas De Minería De Datos. *Gestión Calidad y Desarrollo En Las Facultades de Ingeniería*, 315. <https://bit.ly/305BmII>
- Amat Rodrigo, J. (2018). *Reglas de asociación y algoritmo apriori con R*. [https://rpubs.com/Joaquin\\_AR/397172](https://rpubs.com/Joaquin_AR/397172)
- Arce, D., Lima, F., & Orellana et al., M. (2018). Descubriendo patrones de comportamiento entre contaminantes del aire : Un enfoque de minería de datos. *Enfoque UTE*, 9(4), 168–179. <http://scielo.senescyt.gob.ec/pdf/enfoqueute/v9n4/1390-6542-enfoqueute-9-04-00168.pdf>
- Beltrán Martínez, B. (2001). *Minería de Datos* [Benemérita Universidad Autónoma de Puebla]. <https://www.cs.buap.mx/~bbeltran/NotasMD.pdf>
- Borja-Cevallos, R. (2020). *Contrabando*. Derecho Ecuador. <https://www.derechoecuador.com/contrabando->
- Borondics, F. (2020). *Orange Data Mining*. <https://orange.biolab.si/>
- Calabrese, J., Esponda, S., Pasini, A., & Boracchia, M. (2019). *Guía para evaluar calidad de datos basada en ISO / IEC 25012*. 694–706.
- Camana, R. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. *Revista Tecnológica ESPOL-RTE*, 29(1), 170–183.
- Camaná, R., & Torres, R. (2017). Descubrimiento del estilo de aprendizaje dominante de estudiantes de la carrera de tecnología en análisis de sistemas. *In Crescendo*, 8(2), 193. <https://doi.org/10.21895/incres.2017.v8n2.04>
- Castillo-Rojas, W., Medina-Quispe, F., & Vega-Damke, J. (2017). Esquema de visualización para modelos de clústeres en minería de datos. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, 21, 67–84. <https://doi.org/10.17013/risti.21.67-84>
- Chiang, L. H., Pell, R. J., & Seasholtz, M. B. (2003). Exploring process data with the use of robust outlier detection algorithms. *Journal of Process Control*, 13(5), 437–449. [https://doi.org/10.1016/S0959-1524\(02\)00068-9](https://doi.org/10.1016/S0959-1524(02)00068-9)

- Chiriboga Zambrano, G. (2015, July). Perfil Criminológico. *Perfil Criminológico Nro.15*, 2. [www.flacsoandes.edu.ec](http://www.flacsoandes.edu.ec)
- Cisneros Buitrón, S. A. (2019). *Detección de patrones de deserción estudiantil utilizando técnicas descriptivas de agrupamiento, asociación y atípicos en minería de datos para la gestión académica en la Universidad Técnica del Norte*. Universidad Técnica del Norte.
- de la Fuente Fernández, S. (2001). *Análisis de conglomerados*.
- Dirección Nacional Jurídica. (2018). *Código Orgánico del la Producción, Comercio e Inversiones*. 1–113.
- Flores Urgiles, C. M., & Ortiz Amoroso, M. S. (2018). Revisión de algoritmos para la detección de valores atípicos. *Killkana Técnica*, 2(1), 19. [https://doi.org/10.26871/killkana\\_tecnica.v2i1.287](https://doi.org/10.26871/killkana_tecnica.v2i1.287)
- Gironés, J., Casas, J., & Minguillón et al., J. (2017). *Minería de datos: modelos y algoritmos* (Editorial).
- Han, J., Kamber, M., & Pei, J. (2013). Minería de Datos Conceptos y Técnicas. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9). Morgan Kaufmann Publishers. <https://doi.org/10.1017/CBO9781107415324.004>
- Hernández, J., Ramírez, M. J., & Ferri, C. (2004). *Introducción a la Minería de Datos*. Pearson Educación.
- Hernández Valadez, E. (2006). *Algoritmo de clustering basado en entropía para descubrir grupos en atributos de tipo mixto* (Vol. 52, Issue 2). [https://doi.org/10.1016/0009-3084\(90\)90158-N](https://doi.org/10.1016/0009-3084(90)90158-N)
- Huang, W., & Chen, Y. (2017). The multiset EM algorithm. *Statistics and Probability Letters*, 126, 41–48. <https://doi.org/10.1016/j.spl.2017.02.021>
- Instituto Ecuatoriano de Normalización, I. (2014). *NTE INEN-ISO/IEC 25012*.
- ISO 25000. (2018). *ISO 25012*. ISO. <https://iso25000.com/index.php/normas-iso-25000/iso-25012>
- Jiménez Berríos, L. H. (2015). *APLICACIÓN DE UN SISTEMA DE ALERTA TEMPRANA BASADA EN LA MINERÍA DE DATOS PARA IDENTIFICAR PATRONES DELICTIVOS EN LA CIUDAD DE CHICLAYO*. UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO.
- Knime. (2019). *Software KNIME*. <https://www.knime.com/software-overview>
- Lara Gutiérrez, H. G., Lara Ruiz, M. G., & Hernández Hernández, V. (2016). Análisis de un caso práctico aplicando el algoritmo K means mediante weka (Waikato environment for knowledge analysis). *Ciencia Huasteca Boletín Científico de La Escuela Superior de Huejutla*, 4(7), 1–6. <https://doi.org/10.29057/esh.v4i7.1135>
- Lara Torralbo, J. A. (2014). *Minería de datos* (C. D. E. FINANCIEROS (ed.)). [www.udima.es](http://www.udima.es)

- Larrañaga, P., Inza, I., & Moujahid, A. (2012). *Tema 14. Clustering*. 1–11.  
[https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=2ahUKEwig3-ne88\\_nAhXOrFkKHbisDRUQFjADegQIBxAB&url=http%3A%2F%2Fwww.sc.ehu.es%2Fccwba yes%2Fdocencia%2Fmmcc%2Fdocs%2Ft14clustering.pdf&usg=AOvVaw2IBAoHOrXTyVbnd9Ld6M-n](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=2ahUKEwig3-ne88_nAhXOrFkKHbisDRUQFjADegQIBxAB&url=http%3A%2F%2Fwww.sc.ehu.es%2Fccwba yes%2Fdocencia%2Fmmcc%2Fdocs%2Ft14clustering.pdf&usg=AOvVaw2IBAoHOrXTyVbnd9Ld6M-n)
- Machine Learning Group. (2011). *WEKA El banco de trabajo para el aprendizaje automático*. University of Waikato. <https://www.cs.waikato.ac.nz/ml/weka/>
- Microsoft. (2020). *Visualización de datos | Microsoft Power BI*.  
<https://powerbi.microsoft.com/es-es/>
- Minguillón, J., & Casas, J. (2017). Minería de datos: modelos y algoritmos. In *Minería de datos: modelos y algoritmos* (Editorial).
- Montenegro Cochao, G. R. (2005). *Normalización de Datos*.
- Código Orgánico Integral Penal, 1 (2014).
- Nigro, H. O., Xodo, D., Corti, G., & Terren, D. (2003). *KDD (Knowledge Discovery in Databases): Un proceso centrado en el usuario*.  
[http://sedici.unlp.edu.ar/bitstream/handle/10915/21220/Documento\\_completo.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/21220/Documento_completo.pdf?sequence=1)
- Orellana, M., & Cedillo, P. (2020). Detección de valores atípicos con técnicas de minería de datos y métodos estadísticos. *Enfoque UTE*, 11(1), 56–67.  
<https://doi.org/10.29019/enfoque.v11n1.584>
- Perez Lopez, C., & Gonzáles Satin, D. (2007). *Minería de datos* (Paraninfo (ed.)).
- Pineda, B. (2009). *Método de Selección de Atributos por Clase*. Instituto Nacional de Astrofísica, Óptica y Electrónica.
- Porras, P., & Conde, J. (2019). Habilidades interpersonales y retos del Jefe de Proyecto en tiempos de crisis. *1e Congreso En Dirección de Operaciones*.
- Posada Hernández, G. J. (2016a). Elementos básicos de estadística descriptiva para el análisis de datos. In *Elementos básicos de estadística descriptiva para el análisis de datos*. (Fondo Edit).
- Posada Hernández, G. J. (2016b). Elementos básicos de estadística descriptiva para el análisis de datos. In *Elementos básicos de estadística descriptiva para el análisis de datos*.  
<https://elibro.net/es/ereader/udla/127436?page=8>.
- Quiroga-Juárez, C. A., & Villalobos-Escobedo, A. (2015). Análisis del comportamiento bursátil de las principales bolsas financieras en el mundo usando el análisis multivariado (análisis de componentes principales PCA) para el periodo de 2011 a 2014. *Revista CEA*, 1(2), 25.  
<https://doi.org/10.22430/24223182.122>

- Ramírez-Murillo, H., Torres-Pinzón, C. A., & Forero-García, E. F. (2019). Photovoltaic Potential Estimation by Means of Data Mining in Four Colombian Cities. *Tecnológicas*, 22(46), 77–97. <https://doi.org/https://doi.org/10.22430/22565337.1345>
- RapidMiner Inc. (2020). *RapidMiner \_ Best Data Science & Machine Learning Platform*. <https://rapidminer.com/>
- Rodriguez Miranda, F. O. (2018). *IMPLEMENTACIÓN DE KDD PARA MEJORAR EL PROCESO DE IDENTIFICACION DE ESTILOS DE APRENDIZAJE EN LA UNIVERSIDAD AUTÓNOMA DEL PERÚ*. Universidad Autónoma del Perú.
- Ruiz, G. (2015). Rutas fronterizas del contrabando en Ecuador. *Perfil Criminológico Nro.15*, 3–6. [repositorio.flacsoandes.edu.ec/bitstream/10469/7512/2/BFLACSO-PC15.pdf](http://repositorio.flacsoandes.edu.ec/bitstream/10469/7512/2/BFLACSO-PC15.pdf)
- Salazar Méndez, D. (2015). *Los rostros del contrabando rutas fronterizas*.
- San-Segundo, F., & Marvá, M. (2016). *Postdata 1.0*.
- Santamaria Ruiz, W. (2006). *Técnicas de Minería de Datos Aplicadas en la Detección de Fraude:Estado del Arte. January 2006*.
- SAS Institute Inc. (2020). *Analítica avanzada. Donde el poder se encuentra con el propósito*. [https://www.sas.com/en\\_us/solutions/analytics.html#view-all-products](https://www.sas.com/en_us/solutions/analytics.html#view-all-products)
- SENAE, A. del E. (2016). *Manual Específico Para Licores Aprehendidos Sujetos Al Etiquetado Fiscal*.
- Resolución Nro. SENAE-SENAE-2017-0345-RE, (2017).
- SENAE, A. del E. (2018). *Informe de Gestión 2017*. <https://doi.org/10.1017/CBO9781107415324.004>
- SENAE, A. del E. (2020). *SENAE incrementa cifras de aprehensiones de mercancía ilegal*. <https://www.aduana.gob.ec/senae-incrementa-cifras-de-aprehensiones-de-mercancia-ilegal/>
- Suárez Rodríguez, J. M. (2015). *Caracterización de los hurtos a personas que afectan la localidad Los Mártires de la ciudad de Bogotá, mediante la implementación de algoritmos de agrupamiento de minería de datos espaciales y apoyado en una infraestructura de datos espacial*. [Universidad Distrital Francisco José De Caldas]. <http://repositorio.unan.edu.ni/2986/1/5624.pdf>
- Timarán-Pereira, R., Hernández-Garzón, G., & Quemá-Taimbud, N. (2017). Identificación de lesiones no fatales en la cartografía del municipio de pasto con la técnica de grupamiento. *Revista De Investigación, Desarrollo E Innovación*, 8(1), 147. <https://doi.org/10.19053/20278306.v8.n1.2017.5793>
- Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A., & Alvarado-Pérez, J. C. (2016). *Proceso de conocimiento en bases de datos The Process of Knowledge Discovery on Databases. 2016*, 63–86.

- Universo, E. (2019). *Salario Básico de Ecuador se fija en \$400 para 2020*.  
<https://doi.org/10.16309/j.cnki.issn.1007-1776.2003.03.004>
- Velandia Rodríguez, R. A., Guataquira, A., & Prada, J. (2018). Prototipo de un modelo clasificador para la toma de decisiones a partir de la percepción actual de la seguridad en Bogotá. *Tecnología, Investigación y Academia*, 6(2), 18–24.
- Ventara-Hitachi. (2019). Pentaho - Data Integration and Analytics Platform. In 2019.  
<https://www.hitachivantara.com/en-us/products/data-management-analytics/pentaho-platform.html>
- Vila Espinosa, D. P. (2019). *Detección de patrones de deserción estudiantil utilizando técnicas predictivas de clasificación y regresión de minería de datos, para la gestión académica en la Universidad Técnica del Norte*.