UNIVERSIDAD TÉCNICA DEL NORTE

Facultad de Ingeniería en Ciencias Aplicadas

Carrera de Ingeniería en Sistemas Computacionales

ANÁLISIS DE DATOS APLICANDO LAS TÉCNICAS DE DATA MINING (REGLAS DE ASOCIACIÓN Y CLUSTERING) PARA FORTALECER EL COMERCIO ELECTRÓNICO DESCUBRIENDO HÁBITOS DE COMPRA DE PRODUCTOS Y ACCESORIOS DE BICICLETAS EN LA CIUDAD DE TULCÁN.

Trabajo de grado previo a la obtención del título de Ingeniero en Sistemas Computacionales

Autor:

Brayan Guillermo Pérez Paspuel

Director:

Msc. Fausto Salazar

Ibarra – Ecuador

2022

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	
CÉDULA DE IDENTIDAD:	0401762158
APELLIDOS Y NOMBRES:	PÉREZ PASPUEL BRAYAN GUILLERMO
DIRECCIÓN:	IBARRA, AV. ATAHUALPA Y CARLOS EMILIO GRIJALVA
EMAIL:	bgperezp@utn.edu.ec
TELÉFONO MÓVIL	0939395916

DATOS DE LA OBRA	
TÍTULO:	ANÁLISIS DE DATOS APLICANDO LAS TÉCNICAS DE DATA MINING (REGLAS DE ASOCIACIÓN Y CLUSTERING) PARA FORTALECER EL COMERCIO ELECTRÓNICO DESCUBRIENDO HÁBITOS DE COMPRA DE PRODUCTOS Y ACCESORIOS DE BICICLETAS EN LA CIUDAD DE TULCÁN.
AUTOR (ES):	PÉREZ PASPUEL BRAYAN GUILLERMO
FECHA:	06/07/2022
PROGRAMA:	■ **PREGRADO** ☐ **POSGRADO**
TITULO POR EL QUE OPTA:	INGENIERO EN SISTEMAS COMPUTACIONALES
DIRECTOR:	MSC. FAUSTO ALBERTO SALAZAR FIERRO

2. CONSTANCIAS

El autor manifiesta 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 el titular de los derechos patrimoniales, por lo que asume la responsabilidad sobre el contenido de esta y saldrá en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, a los 06 días del mes de julio de 2022

EL AUTOR:

PÉREZ PASPUEL BRAYAN GUILLERMO
040176215-8

CERTIFICADO DEL DIRECTOR DE TRABAJO DE GRADO

UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

CERTIFICACIÓN DEL DIRECTOR

Por medio del presente yo MSc. Fausto Salazar, certifico que el Sr. Brayan Guillermo Pérez Paspuel portador de la cédula de ciudadanía Nro. 040176215-8. Ha trabajado en el desarrollo del proyecto de tesis "ANÁLISIS DE DATOS APLICANDO LAS TÉCNICAS DE DATA MINING (REGLAS DE ASOCIACIÓN Y CLUSTERING) PARA FORTALECER EL COMERCIO ELECTRÓNICO DESCUBRIENDO HÁBITOS DE COMPRA DE PRODUCTOS Y ACCESORIOS DE BICICLETAS EN LA CIUDAD DE TULCÁN", previo a la obtención del título de Ingeniería en Sistemas Computacionales, lo cual ha realizado en su totalidad con responsabilidad y esmero.

Es todo cuanto puedo certificar en honor a la verdad.

En la ciudad de Ibarra, a los días 06 del mes de julio del 2022

Atentamente

FAUSTO ALBERTO SALAZAR FIERRO

Firmado digitalmente por FAUSTO ALBERTO SALAZAR FIERRO
Fecha: 2022.07.06 11:06:20 -05'00'

MSc. Fausto Salazar

TUTOR TRABAJO DE GRADO

DEDICATORIA

Dedico este proyecto de titulación a mi madre Sonia Paspuel y mi padre Guillermo Pérez por brindarme todo su apoyo incodicional, lo cual me motivó a seguir a pesar de muchos obstaculos que se presentaron.

A mis hermanos Jaider y Yamileth, que con sus palabras me daban fuerzas para continuar.

A mi abuela Dolores Ortega por dame palabras de aliento en todo el trascurso de mi carrera, a mis tíos, primos, y todos mis familiares que de alguna manera mostraban su interés por verme conseguir esta meta.

Brayan Pérez

AGRADECIMIENTOS

Primero quiero agradecer a Dios por su fidelidad y misericordia que ha tenido conmigo en todo el trascurso de mi carrera, sin su ayuda nada de esto sería posible.

A mi madre Sonia Paspuel y mi padre Guillermo Pérez que gracias a su esfuerzo y sacrificio hicieron que pueda estudiar esta carrera.

A mis hermanos por su cariño y motivación para no darme por vencido.

Un agradecimiento especial mi tutor MSc. Fausto Salazar y a mi docente PhD. Iván García, por guiarme en la elaboración de este proyecto.

Agradezco a mis abuelos, tíos, primos y amigos que siempre estuvieron al pendiente en todo el transcurso de mi vida universitaria.

A quienes fueron mi segunda familia, mis amigos Tardones, hicieron que esta etapa de mi vida sea la mejor, siempre estaré agradecido con ustedes.

Brayan Pérez

Índice de Contenido

Índice de Figuras

Índice de Tablas

RESUMEN

En esta tesis se han analizado diferentes sets de datos tanto de compradores de bicicletas como datos demográficos de la población objetivo, con el fin de determinar el tamaño del mercado y el arquetipo de compradores de bicicletas para conocer la factibilidad de implementación de una tienda de bicicletas y accesorios.

Para este fin se han validado herramientas estadísticas, proyecciones de la población, emparejamiento de datos, construcción de un prototipo, herramientas de visualización y finalmente algoritmos de modelamiento y pruebas para determinar el segmento de compradores.

Como resultados se obtuvieron algunos datos importantes como: quienes son los mejores compradores; a que segmento de la población se debe enfocar; que tipos de productos son los más vendidos; edades y posible tamaño del mercado.

ABSTRACT

In this thesis, different data sets of both bicycle buyers and demographic data of the target population have been analyzed, in order to determine the size of the market and the archetype of bicycle buyers to know the feasibility of implementing a bicycle shop and accessories.

For this purpose, statistical tools, population projections, data matching, construction of a prototype, visualization tools and finally modeling algorithms and tests to determine the segment of buyers have been validated.

As a result, some important data were obtained, such as: who are the best buyers; which segment of the population should be targeted; what types of products are the best sellers; ages and possible size of the market.

INTRODUCCION

PROBLEMA

Antecedentes

Los comerciantes al momento de la generación de la emergencia sanitaria cambiaron todo su modelo de negocio, comenzaron a divulgar la venta de sus productos a través de las redes sociales, pero en algunos casos que no tienen conocimiento de estrategias para fortalecer su comercio electrónico no tuvieron resultados satisfactorios ya que no pueden descubrir hábitos de compra de sus clientes con el fin de mejorar sus productos y servicios.

Los comerciantes de venta de productos y accesorios de bicicletas no tienen conocimiento de la existencia de técnicas y herramientas tecnológicas que sirven para analizar sus datos de ventas y darles valor, permitiendo tomar mejores decisiones para fortalecer su comercialización.

Situación Actual

Los comerciantes que se dedican a la venta de productos y accesorios de bicicletas en la ciudad de Tulcán no cuentan con conocimientos acerca de lo que es el comercio electrónico y de cómo aplicarlo en la comercialización de sus productos, de igual manera no saben que es el Data Mining y cuáles son las técnicas que les permita identificar las intenciones o hábitos de compra de los clientes.

Planteamiento del problema

Existe un bajo índice de rentabilidad en los negocios orientados a la venta de productos y accesorios de bicicletas en la ciudad de Tulcán.

Aspectos negativos que influyen en el problema:

Económicos. - Los vendedores de productos y accesorios de bicicletas no cuentan con un presupuesto establecido para hacer análisis de datos.

• Costos elevados.

• Bajos ingresos.

Talento Humano. – No encuentran especialistas que puedan hacer análisis de Datos

Tecnológico

• No encuentra mecanismos técnicos para poder recolectar datos o no hay fuentes de datos estandarizadas.

• No cuentan con la suficiente capacitación para manejar las herramientas tecnológicas.

• Ven a las tecnologías como un gasto y no como una inversión que les ayudará a maximizar sus ventas.

Comercial. – No encuentran hábitos de compra para saber lo que necesita el cliente.

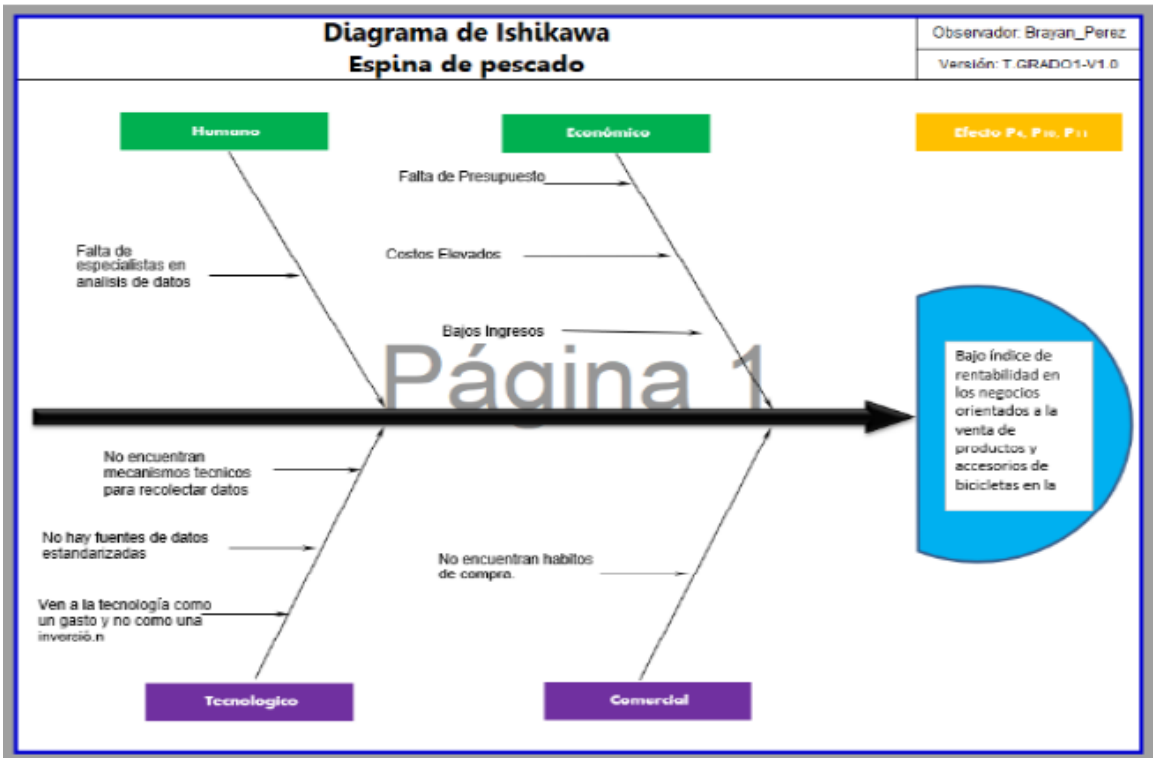


Figura 1 Diagrama de Ishikawa

Objetivos

Objetivo General

Analizar los datos aplicando las técnicas de Data Mining (Reglas de asociación y Clustering) para fortalecer el comercio electrónico permitiendo descubrir hábitos de compra de productos y accesorios de bicicletas en la ciudad de Tulcán.

Objetivos Específicos

a) Realizar el marco teórico sobre el uso y aplicación de Reglas de Asociación y Clustering en Data Mining con Comercio Electrónico.

b) Construir un Data Mining aplicando reglas de asociación y clustering para descubrir hábitos de compra de productos y accesorios de bicicletas.

c) Validar los resultados obtenidos.

Alcance

Con esta investigación se pretende llegar a los comerciantes que se dedican a la venta de productos y accesorios de bicicletas de la ciudad de Tulcán, poniendo en su conocimiento que al aplicar las técnicas de minería de datos que son las Reglas de Asociación y Clustering, acompañado con herramientas informáticas como el manejo del lenguaje de programación Python, se puede realizar un análisis de sus ventas y así darles el mejor servicio a todos sus clientes como también les va a permitir tomar mejores decisiones que ayuden a fortalecer su economía y por lo tanto el comercio electrónico enfocado en este segmento.

Los datos a ser analizados serán obtenidos de un dataset de la empresa Aguila importaciones.

Posteriormente se realiza el procesamiento de texto utilizando Python.

Por último, mostraremos los resultados obtenidos en el prototipo desarrollado.

Las tecnologías que se usarán para el análisis de datos aplicando la técnica Reglas de asociación en Data Mining serán:

- Lenguaje de programación Python.
- Software VScode
- Técnicas de Data Mining Reglas de Asociación y Clustering.

Los módulos que tendrá el prototipo son dos:

- El primero es el descriptivo donde se verá todos los datos sumariados.
- El segundo es el de diagnóstico donde se verá el detalle de los datos.

Figura 2 Metodología Scrum

Metodología

Se utilizará la metodología de investigación documental con la que recolectaremos toda la bibliográfica con respecto al tema de técnicas de Data Mining en especial donde se enfoque sobre las Reglas de Asociación y Clustering su uso y aplicación.

Se obtendrá los datos a los cuales se va a aplicar la técnica reglas de asociación y para luego realizar el procesamiento de texto aplicando Python, por último, se realizará un análisis más profundo de los datos.

Por último, se validará los datos obtenidos de la propuesta mostrando en views o gráficos estadísticos en el prototipo desarrollado.

Justificación

Existe un bajo índice de rentabilidad en los negocios orientados a la venta de productos y accesorios de bicicletas en la ciudad de Tulcán.

El presente proyecto tiene un enfoque hacia dos de los objetivos de desarrollo sostenible:

Objetivo 8.- Trabajo decente y crecimiento económico

El objetivo es estimular el crecimiento económico sostenible mediante el aumento de los niveles de productividad y la innovación tecnológica. Fomentar políticas que estimulen el espíritu empresarial y la creación de empleo es crucial para este fin, así como también las medidas eficaces para erradicar el trabajo forzoso, la esclavitud y el tráfico humano. Con estas metas en consideración, el objetivo es lograr empleo pleno y productivo y un trabajo decente para todos los hombres y mujeres para 2030 (UNDP, 2021).

Objetivo 9.- Industria, innovación e infraestructura

Los avances tecnológicos también con esenciales para encontrar soluciones permanentes a los desafíos económicos y ambientales, al igual que la oferta de nuevos empleos y la promoción de la eficiencia energética. Otras formas importantes para facilitar el desarrollo sostenible son la promoción de industrias sostenibles y la inversión en investigación e innovación científicas (UNDP, 2021).

Justificación Ambiental. – Uno de los mayores contaminantes a nivel mundial es la fabricación de papel, como también la tala indiscriminada de los árboles, por lo que se cree indispensable la reducción del consumo de papel.

CAPÍTULO 1

Marco Teórico

1.1 Comercio Electrónico

1.1.1 Evolución del Comercio Electrónico

En la actualidad, el comercio electrónico es una de las industrias más rentables del mundo, debido a los constantes avances tecnológicos; por lo que todo muestra a que se continuará implementando nuevas tecnologías que favorezcan en el creciente desarrollo del comercio electrónico y se convertirá eventualmente en la forma de negocio más común en la mayoría de los países y regiones del mundo.

Pero antes de ello, se debe tomar en cuenta que el internet durante las últimas décadas se ha convertido en una herramienta de gran importancia para la humanidad, ya que ha revolucionado la sociedad, el comercio y el hombre moderno en términos de organizaciones; ya que las entidades se han visto obligados a adoptar esta herramienta para ser visibles en un mundo globalizado y altamente competitivo.

Dado a que la información que se obtienen de las mismas de forma unipersonal no generan cambios relevantes, ni inteligencia artificial, ni el efecto de las computadoras en los procesos de decisión, determinación política o desarrollo de estrategias; pero que con la ayuda del internet está produciendo profundas transformaciones en la economía, los mercados y las estructuras industriales entero; en bienes y servicios y sus flujos; en la segmentación, en los valores y la comportamiento del consumidor; en los mercados laborales; pero tal vez el impacto ejercido en la sociedad, la política y la visión que tenemos sobre el mundo y sobre de nosotros mismos (González Ó. , 2011).

Por ello, se puede manifestar que el mundo está atravesando una revolución de la información en donde su símbolo es Internet, comparándolo con la revolución industrial en la que su símbolo era la máquina de vapor, estas dos revoluciones tienen un cierto paralelo en la forma en que han modificó la forma de trabajar, vivir e interactuar con la sociedad, la revolución industrial logró cambiar los procesos de mecanización por métodos de automatización de productos industriales quince conceptos básicos de esa época, como los textiles; así como la revolución de la información que surgió con la llegada de las primeras computadoras donde el proceso de informatización cambió a proceso de automatización, generando así un aumento en la eficacia, eficiencia y productividad en las organizaciones.

Consecuente en el año 1989, apareció un servicio en la World Wide Web conocido como la telaraña global (www); el cual fue generado por un grupo de investigadores en Ginebra (Suiza), este método fue idóneo para utilizar la tecnología y vincular con los diversos documentos de las computadoras, integrando textos, gráficos, música, videos, entre otros (Manríquez, 2018).

Adicional se resalta el eje esencia de www es su alto nivel de accesibilidad, acompañado de conocimientos básicos de informática por parte de los usuarios, el desarrollo de 10 tecnologías y telecomunicaciones, han hecho posible el intercambio de la información: mismo que crece a pasos agigantados. Por otro lado, el uso de internet en el sector empresarial dio apertura a una nueva forma de hacer negocios y realizar transacciones comerciales en las que se intercambia un valor por algún bien o servicio, por medio de una plataforma electrónica, que se conoce como comercio electrónico o E-commerce (Rodríguez I., 2014).

El comercio electrónico nominado en su inglés como E-commerce tuvo sus indicios en Estados Unidos, el cual se inició con un intercambio electrónico de datos (IED) entre firmas comerciales u organizaciones, las mismas que lo empleaban con la finalidad de permitir el uso de comprobantes electrónicos como: facturas, notas de ventas, órdenes de compra, cotizaciones etc.

En el año de 1992 surge el primer proyecto de tienda online, la cual tuvo sus indicios con un sistema que imitaba los tablones de anuncios o propagandas que le permitían a los usuarios el adquirir u ofertar libros; esta página con el transcurso del tiempo y de ver la efectividad y la facilidad de comercialización fue evolucionando hasta convertirse en BOOKS.COM; consecuente a ello continuo el desarrollo de productos electrónicos que permitan él envió de datos personales de manera segura por medio del empleo de internet. Años más tardes se incorporó la tecnología celular, dicho sistema facilito el proceso de adquisición de productos por medio del empleo del dispositivo celular.

En 1998 se creó el sistema PayPal, el cual impulsa el comercio electrónico debido a que brinda de las facilidades, accesibilidad y seguridad para efectuar diversos pagos, a partir de este acontecimiento se crean distintas tiendas en líneas que permiten a los usuarios adquirir productos online generando un mercado mucho más dinámico, ágil y eficiente (Álvarez, 2006).

Por ello se puede manifestar que el comercio electrónico ha transformado el mercado, las funciones tradicionales de la intermediación, dado a que ha sido reemplazada y han aparecido nuevos productos y servicios en el mismo. De esta forma se ha modificado la organización en el trabajo, aumentando la flexibilidad que es considerable, ya que el comercio electrónico tiene un efecto catalizador, mismo que acelera, los cambios producidos por la economía,

aumentando de la misma manera la interactividad de la economía y sus vínculos con las pequeñas y medianas empresas, personas e incluso familias (Álamo, 2016).

Según González (2011) manifiesta que el comercio electrónico es la actividad de intercambio de bienes y/o servicios que desarrolla un ofertante bajo un modelo de correlación empresarial basado en interacciones electrónicas que sustituyen la presencia de asesores comerciales y documentación física que respalda una compra (pág. 114).

El E-commerce son aquellas transacciones comerciales que se realizan por medio del uso del internet; debido a que, mediante el empleo del mismo, las organizaciones pueden efectuar negociaciones y dar a conocer la diversidad de productos a los usuarios a través de anuncios o páginas web que son accesibles desde cualquier dispositivo electrónico (Asociación Española de Comercion Electrónico y de Marketing Relacional - AECEM, 2016).

Se puede manifestar que el comercio electrónico se ha encargado de cambiar la ideología del comercio tradicional por uno método innovador, en el cual los bienes, productos y/o servicios son auspiciados en un portal web o aplicaciones móviles de forma dinámica y atractiva para el usuario, además que facilitan el proceso de pago y la entrega de los mismos.

A continuación, se presenta las diferencias entre el comercio tradicional y el comercio electrónico como también el mercado virtual y el físico.

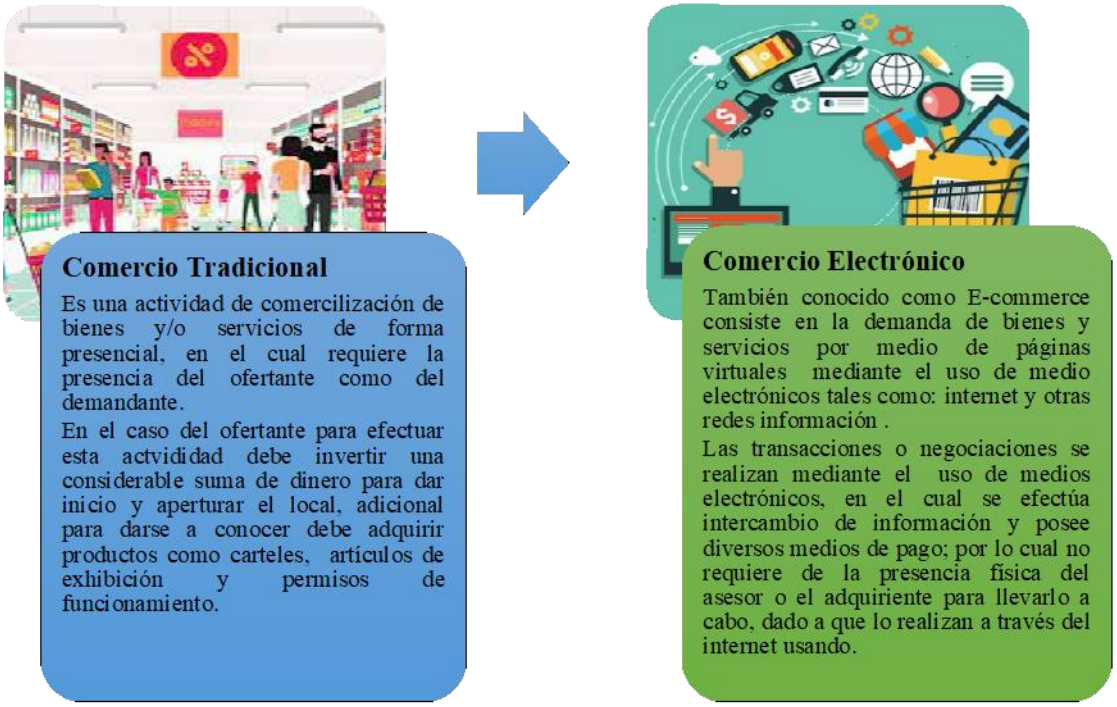


Figura 3 Diferencia entre comercio tradicional y el electrónico

Fuente: Álvarez (2006)

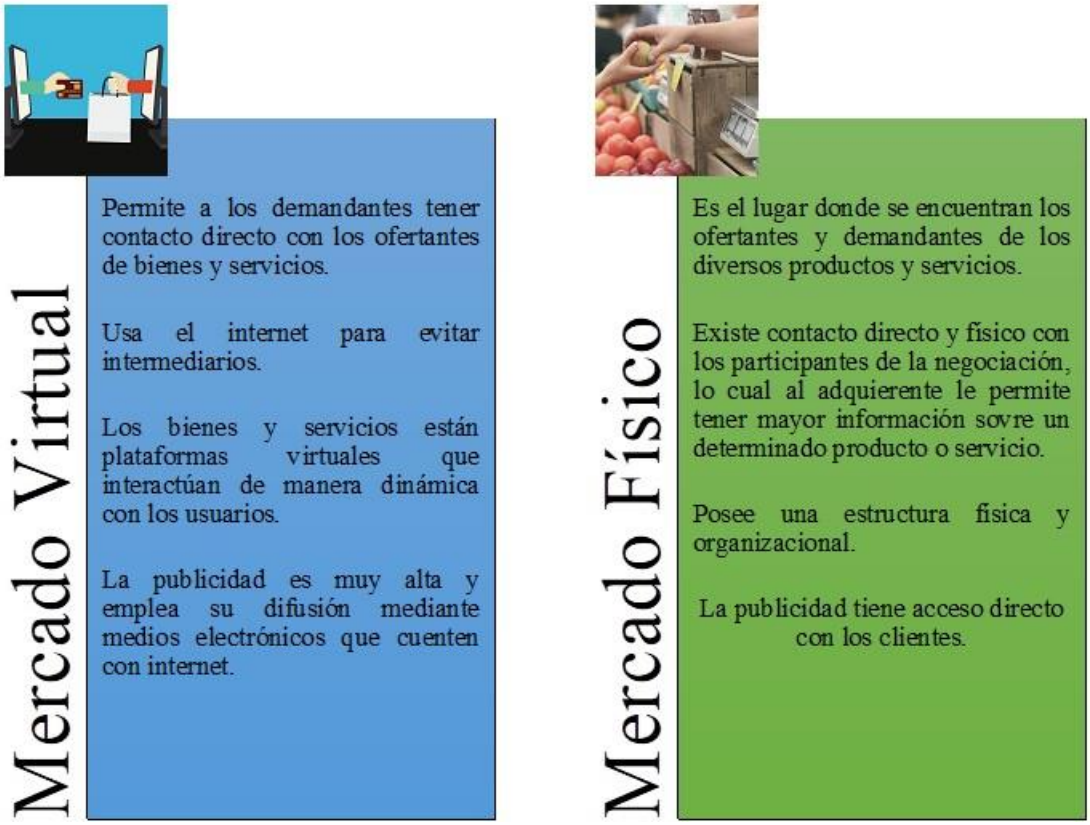


Figura 4 Diferencia entre mercado virtual y físico

Fuente: Vega (2005)

Como se puede observar en las anteriores figuras el comercio electrónico es un medio moderno que mediante el uso del internet y medios electrónicos se puede adquirir u ofertar bienes y servicios; los mismos que están anunciados en páginas web o virtuales, con el fin de optimizar el tiempo de negociación, reducir los costos y maximizar su nivel de acercamiento con los usuarios.

1.1.2 Tipos del Comercio Electrónico

Existen diversos tipos o categorías de E-commerce; estos se relacionan o van ligados a las formas de intercambio comercial o de los actores como: Empresas (B), consumidores (C), administración (G), inversores (I) y los colaboradores o empleados (E), mismos que interactúan en el proceso de compra y venta de bienes y/o servicios.

A continuación, se detallan los tipos de comercio electrónico:

a Business to Consumer B2C

Este tipo comercio es Negocio al Consumidor, en el que el comprador adquiere un bien, producto o servicio de una entidad por medio de sitios virtuales. Para realizar la compra los usuarios ingresan a las tiendas virtuales donde encuentra los distintos artículos de manera rápida, ágil, desde cualquier lugar y sin restricción de horario; el único requisito es que debe estar conectado a internet. Este modelo brinda a los usuarios diversas tiendas que facilitan la interacción de forma directa con el demandante, agilita el proceso de pago y la entrega de los mismos.

Entre las principales ventajas es la automatización y optimiza de los procesos de gestión de compra, reduce costos, mejora el servicio de atención al cliente, reduce el tiempo de espera y entrega, acrecentar el nivel de competitividad de las industrias, entre otros (AECEM, 2014).

b Business to Business B2B

Este tipo de negocio significa Negocios o Negocios, como su nombre lo indica es el proceso de comercialización que se lleva a cabo entre empresas por medio del uso de internet. En esta negociación se enfoca en asesores, proveedores, adquirientes e intermediarios dado a que lo que buscan los actores es buscar proveedores o intermediarios que coadyuven a concretar acuerdos comerciales.

Este es uno de los acuerdos más habituales debido a que sus ventas son al por mayor y no está dirigido al consumidor final.

c Consumer to Consumer C2C

Este modelo se fundamenta en el comercio consumidor a consumidor, en el que los interesados no buscan intermediarios sino ofertar sus productos de forma directa a los consumidores finales a precios accesibles y de ágil negociación; operan desde plataformas virtuales conectadas a internet.

Las ventajas es que ofrece gran variedad de productos, artículos o servicios, coadyuva a la adquisición de nuevos productos, ahorra tiempo de difusión y oferta, puede implementar las pequeñas y grandes empresas (González Ó. , 2011).

d Government to Consumer G2C

El modelo se fundamenta en el comercio entre los gobiernos digitales de los distintos países, el cual tiene como finalidad el facilitar a las naciones la ejecución de trámites y pagos por medio de plataformas virtuales. Es decir, los gobiernos cancelan por un documento o servicio, lo cual les permite ahorra tiempo y dinero además de obtener respaldos electrónicos más seguros.

e Business to Employee B2E

Este tipo de modelo se basa en el comercio Negocio a Empleado, debido a que se centra en la relación comercial entre la institución y sus colaboradores. Esta comercialización se basa en la oferta que propicia la entidad hacia sus miembros o trabajadores por medio de propuestas llamativas que incidan en la mejora del desempeño laboral de los mismos. E incluso el proceso permite que cualquier individuo pueda iniciar un negocio virtual (Torres R., 2010).

f Government to Business G2B

Este comercio de Negocios a Gobierno se trata de la comercialización o negociación entre las empresas y el estado por medio de la tecnología digital, la cual tiene como objetivo el proporcionar a la administración pública facilidades para la adquisición de bienes, productos o servicios de forma ágil, minimice los costos y ahorre tiempo en el proceso de pedido. Las empresas que brindan este tipo de servicios a los gobiernos son las entidades especializadas en proyecto de mercadotecnia, ingeniería, asesoría, entre otros.

g Consumer to Business C2B

El modelo se fundamenta en la comercialización consumidor a empresa, este es un proceso poco habitual dentro del comercio dado a que el consumidor o adquiriente de una organización pueda influir en la difusión del mismo a través de sus páginas virtuales donde se da conocer sus expectativas y nivel de satisfacción del producto adquirido (Torres R. , 2010).

1.1.3 Características del Comercio Electrónico

Una comercialización o negociación para ser considerado como comercio electrónico es necesario que cumpla con las siguientes características:

- Ser un medio de pago electrónico de transcendencia económica.
  El cual debe estar ajustado o el par de la evolución económica y tecnológica, debido a que los clientes, consumidores y usuarios cambian sus gustos y preferencias, lo que conlleva a las instituciones tiendan a satisfacer dichas necesidades bajo el enfoque de innovación que le permita ser competitivo, productivo y eficientes con la finalidad de mantener la fidelidad de los mismos (Torre & Codner, 2017).
- Ser un medio de comercio virtual.
- Donde los actores del comercio no se conocen de manera física debido a que se encuentran en diversos países.
- Medio de comercio de carácter universal.

- Este tipo de comercio es efectuado por diversos individuos desde cualquier parte del mundo siempre y cuando tenga la posibilidad de conectarse a internet; se manifiesta que es universal dado a que no tiene barreras geográficas o de otro tipo (Cisneros, 2018).
- Vincular la innovación tecnológica.
- Se fundamenta en la evolución y avance tecnológico para la comercializa bienes y servicios por medio de plataformas virtuales.
- Ser un medio de difusión de accesible y de bajo costo.
- Este tipo de comercio a ser electrónico no requiere una fuerte inversión en medios publicitarios por lo que reduce los costos de transacción, minimiza los tiempos de demora y respuesta a las necesidades de los usuarios, consumidores y clientes (Durán, 2017).
- Ser un medio rápido, dinámico y ágil.
- Este modelo de comercio refleja una gran diferencia con el comercio tradicional, puesto que el E-commerce se desarrolla con mayor rapidez para la adquisición de un producto, tiene mayores facilidades de pago y entrega de los bienes, y/o servicios dado a que todo se efectúa por medio del internet (Martínez & Ruíz, 2016).

1.1.4 Importancia del Comercio Electrónico

El comercio electrónico es de gran importancia debido a que permite a los empresarios y personas con emprendimientos a lograr una mayor expansión y captación de clientes, usuarios y consumidores por medio de la adaptación del empleo de técnicas de información y comunicación que permite brindar un mayor dinamismo y diversificación en las plataformas virtuales con la finalidad de agilizar los procesos de adquisición, pago, entrega, atención al cliente, entre otros. Además, que constituye como un eje central que coadyuva en el cambio de la matriz productiva para las empresas pequeñas, medianas y grandes organizaciones tanto a nivel local, regional e internacional (MINTEL, 2018).

1.1.5 Ventajas y desventajas del Comercio Electrónico

Según (Seoane, 2005) las principales ventajas y desventajas del E-commerce o comercio electrónico para los ofertantes como para los demandantes son los siguientes:

Ventajas del Comercio Electrónico

Tabla 1 Ventajas del Comercio Electrónico

EMPRESA (OFERTANTE)	COMPRADOR (DEMANDANTE)
Acrecienta el nivel de eficiencia de las	Permite encontrar un portafolio dinámico

EMPRESA (OFERTANTE)	COMPRADOR (DEMANDANTE)
organizaciones y automatiza los procesos. Incrementa el número de ventas, acapara mayor parte del mercado y mejora la atención al cliente. Mejora las relaciones entre la entidad y sus clientes. Minimiza el empleo de intermediarios. Mejora la cadena de distribución de los bienes y/ servicios. Reducir el nivel de inversión en medios publicitarios. Incrementa el nivel de competitividad y productividad organizacional. Plataformas virtuales accesibles para los usuarios las 24 horas al día. Permite un trato más personalizado.	y atractivos de los productos, bienes y/o servicios ofertados por las entidades. Puede acceder a diversas alternativas para adquirir un bien o servicio. Mayor acceso a productos extranjeros. Cuentan con servicios de preventa, post venta y mejor atención al usuario, consumidor o cliente. Pueden comprar bienes, servicios o productos de la localidad o fuera de su localidad de forma ágil y rápida.

Fuente: Seoane (2005)

Desventajas del Comercio Electrónico

Tabla 2 Desventajas del Comercio Electrónico

EMPRESA (OFERTANTE)	COMPRADOR (DEMANDANTE)
Requiere de acuerdo comerciales internacionales para armonizar el proceso de comercio electrónico en la nación.	El lugar donde se adquiere un bien o servicio la mayoría no posee una infraestructura o espacio físico determinado.
Periodos cortos de validez de contratos y negociaciones sin papeles	Al no contar con una infraestructura física provoca en los adquirientes una mayor desconfianza para el proceso de compra como también de pago.
Mayor control de las transacciones internacionales en función a pagos y cobro de aranceles e impuestos.	Genera recelo y desconfianza de los compradores al momento de brindar información personal.
Desconocimiento del medio y posibilidades de fraude.	Tiene a generar desconocimiento de los medios de pago que utiliza en la plataforma virtual.
Carencia de normativa que regule la propiedad intelectual.	Problemas al tratar de comprender un portal o plataforma que está en otro idioma.
Falencia en la seguridad de los medios electrónicos de pago.	Los compradores en su gran mayoría tienen a adquirir productos en los sitios reales antes que el virtual.
Al no contar con un control de la información dificulta la búsqueda de datos específicos.	

Fuente: Seoane (2005)

1.2 Data Mining

Actualmente la minería de los datos o data mining es una nueva tecnología con gran potencia e importancia para la entidad, puesto que permite enfocar, recopilar, tratar, analizar, almacenar y analizar datos que coadyuven a obtener un escenario de la situación real de un determinado acontecimiento de estudio o suceso; el cual facilitará la interpretación del comportamiento y variantes del mismo a los empresarios una manera eficiente y eficaz lo cual contribuirá en la toma de decisión.

Se debe tomar en cuenta que la minería de datos emerge como una nueva innovación tecnológica la cual tiene como finalidad el ayudar a comprender el contenido de la base de datos; que, por lo general estos son el valor total de las materias primas de forma bruta, lo cual dificulta en determinadas ocasiones su análisis. Para ello los usuarios diseñan modelos que permitan mejorar la interpretación de dicha información e incorporan valor agregado al momento de emplear sus conocimientos y experiencia (Molina, 2012).

Figura 5 Relación entre dato, información y conocimiento

Fuente: Molina (2012)

Cuando los profesionales o expertos diseñan o buscan con cuidado modelos para hacer la interpretación del enfrentamiento entre la información y el modelo tiene como finalidad el evitar análisis fallidos y responder de manera rápida a las preguntas gerenciales sin tomar mucho tiempo dado que, para las organizaciones, la minería de datos es esencial y permite descubrir las falencias que posee un determinado proceso o gestión.

Además, que con la gran cantidad de datos emitidos por las entidades es considerada como un bien patrimonial. De manera que si las entidades tuvieran pérdidas parciales o totales de datos generaría diversos perjuicios estructurales y/o económicos para la misma, por lo cual información debe ser protegida pero también explotada. Es por ello, que en los últimos años los empresarios han logrado una transmisión de datos en busca de un mejor manejo y almacenamiento para lo cual se resalta los siguientes factores que contribuirán en el mismo, estos son:

- La disminución del precio de los sistemas de almacenamiento temporales como permanentes.
- Mayor velocidad y capacidad de los procesadores de los equipos de cómputo.

- Incremento del nivel de confiabilidad y facilidad en la transmisión de datos o información relevante.
- Diseño de sistemas administradores en base a datos más poderosos (Molina, 2012).

De acuerdo a Fayyard, la Data Mining es proceso innovador de identificación lícita y novedosa, con gran ponencia y útil que permite entender los patrones y variaciones de una base de datos extensa que favorece en la toma de decisiones (Fayyad, 1996).

Data Mining es la implementación de las buenas prácticas por medio del empleo de técnicas, modelos y algoritmos que se utilizan de forma constante para explorar una base de datos hasta obtener los resultados esperados de modo automático, tomando en cuenta lo que se está buscando. La finalidad es encontrar patrones repetitivos o tendencias que expliquen la variación de los mismos (Pérez & Santín, 2008).

Conformo a los conceptos anteriormente mencionados se recalca que la manera de datos es el conjunto de instrumentos, herramientas y técnicas de análisis de información que a través de la determinación de patrones y combinaciones directas tienden a ser un soporte esencial en el proceso de toma de decisiones.

1.2.1 Arquitectura de datos

La arquitectura de datos hacer referencia específicamente a la gestión, migración y gobernabilidad de los datos, mismos que se detallan a continuación:

- **Gestión de Datos**
  La gestión de datos se efectúa cuando la organización ha optado por realizar una evolución arquitectónica a gran escala de sobre el manejo de los datos, por lo que implementa estrategias y técnicas que permitan integral los datos de manera eficaz, efectiva y oportuna, misma que proporcionara mayor competitividad y productividad a la entidad.
- **Migración de Datos o información**
  La migración de la información se realiza cuando se reemplaza el sistema referencial a uno nuevo, en el cual se traslada los datos de manera cuidadosa y oportuna.
- **Gobernabilidad de datos**
  La gobernabilidad de los datos hace referencia a la seguridad que tiene la entidad al momento de iniciar con el cambio o transformación de las distintas dimensiones de la organización, para lo cual toma en consideración la estructura, el sistema de gestión y el personal (The Open Gruop Standard, 2018).

1.2.2 Etapas de Data Mining

El Data Mining al ser una tecnología innovadora cumple con diversas etapas que agrupan las áreas de una entidad, gobierno, universidad, hospital, entre otros, pero no es un software, puesto que durante su ejecución y desarrollo emplea aplicaciones software en cada una de sus fases, la cual puede ser estadística, de visualización de datos o de inteligencia artificial; estas son:

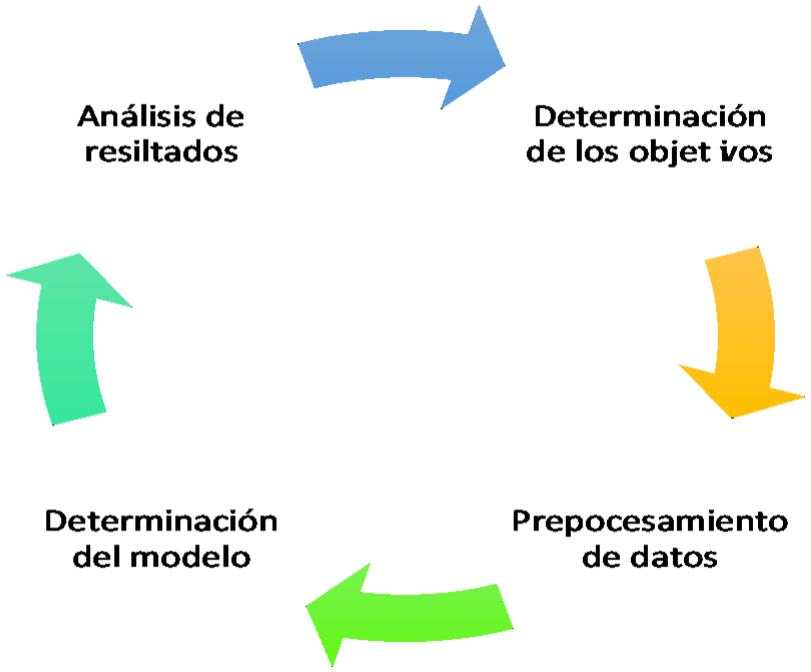


Figura 6 Etapas de Data Mining

Fuente: Molina (2012)

Conforme a la Figura N°6 se puede decir que la Data Mining al cumplir con estas etapas permite descubrir las falencias y problemáticas que tiene una entidad por lo que mediante los resultados obtenidos la emplean como eje esencia para el establecimiento de estrategias y reglas de decisión que permitan a la organización mejorar y ser más competitiva.

Puesto que en la primera etapa se tiene como finalidad la delimitación los objetivos de la persona interesada bajo la alineación de la data mining, en la segunda hace referencia a la selección, depuración, reducción y transformación de la base de datos; este proceso conlleva aproximadamente el 70% de la ejecución del modelo.

En la tercera etapa: Determinación del modelo se inicia con el proceso de análisis estadístico de los datos mediante cuadros y figuras que evidencia la primera aproximación de la situación real de la entidad; consecuente se analiza los objetivos determinados y las actividades a realizarse para cumplir el mismo por medio del uso de algoritmos diseñados en distintas área de la inteligencia artificial y finalmente en la última etapa se verifica que los

resultados obtenidos sean coherentes y correlacionales a los estadísticos, puesto que al determinar su efectividad facilitará el proceso de toma de decisiones (Molina, 2012).

1.2.3 Extracción del conocimiento en Data Mining

La Data como se ha expuesto anteriormente es un modelo que permite analizar grandes bases de datos de manera estadística, visualización de datos o inteligencia artificial, lo cual de manera proactiva y oportuna brinda información relevante sobre un suceso bajo la recopilación de manera masiva de información, uso de equipo informático idóneo y algoritmos.

Puesto que, para la extracción de conocimiento es necesario e importante estar relacionado con el proceso de descubrimiento, reconocimiento e identificación de patrones y tendencias válidas e información útil que mediante la conversión arroje los resultados esperados. A continuación, se presenta el proceso de extracción del conocimiento:

Figura 7 Proceso de Extracción de conocimiento en Data Mining

Fuente: Hernández, Ramírez y Ferri (2004)

1.3 Técnicas de Data Mining

La categorización de la Data Mining (DM) se puede realizar de distintas formas, ya que en la práctica dependerá de técnica que se adopte; entre las cuales están la predictiva que se fundamenta en variables dependientes e independientes; la descriptiva que expresa que todas las variables tienen el mismo estatus y las Auxiliares.

Las técnicas predictivas son aquellas que expresan que para el desarrollo de un modelo se requiere como base datos previos que permitan contrastar el después de un proceso de minera de datos antes de ser validado. Esto se debe a que como primera fase del modelo debe identificar de manera objetiva partiendo información que muestre el comportamiento o

tendencia inicial de un determinado estudio, reconocer si se realiza algún tipo de ajustes o estimaciones que altere la información preliminar (Pérez & Santín, 2008).

Una vez validada la información se procede con la estimación de los parámetros elegidos, el diagnostico que valide la veracidad del modelo y consecuente se inicia con la predicción sobre los resultados obtenidos, el cual coadyuvará a predecir los valores a futuro de las variables dependientes por medio del uso de herramientas de análisis como la regresión, árbol de decisiones, redes neuronales y análisis discriminante, entre otros instrumentos que se puede apreciar a continuación:

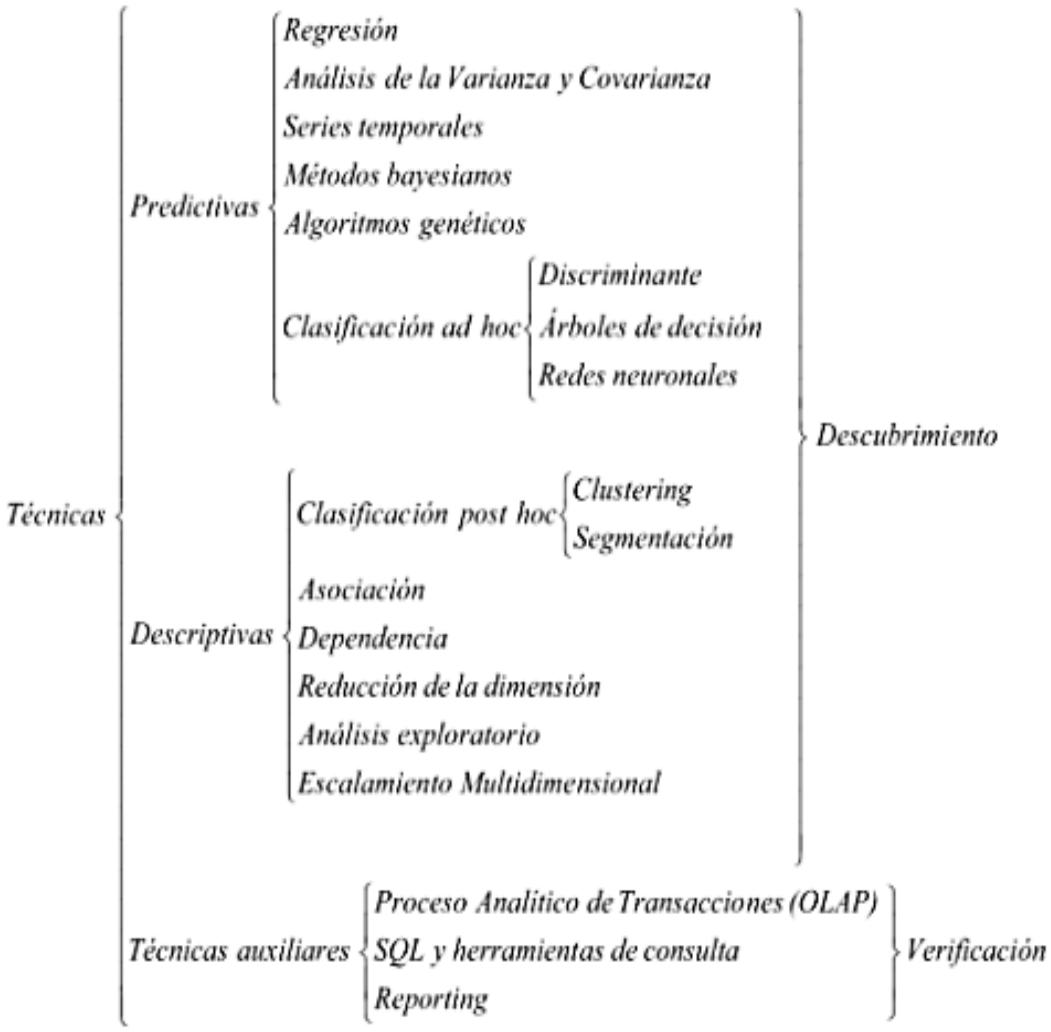


Figura 8 Técnicas de Data Mining

Fuente: Pérez & Santín (2008)

Como se puede apreciar en la Figura No 8 los mecanismos de análisis como el árbol de problemas y las redes neuronales se encargan de elegir un atributo como raíz para el desarrollo de las variables más relevantes.

En el caso de las técnicas descriptivas son aquellas que no se asignan ningún tipo de variables, dado que la carencia de variables dependientes e independientes como tampoco existe un modelo previo para el estudio. Por lo que, este conlleva a la creación o diseño de un modelo en base al reconocimiento de patrones o tendencias de un objeto de estudio. Como se puede observar en la figura a este tipo de técnicas se incluyen otras técnicas como clustering, de segmentación, asociación, dependencia, exploratorias y de reducción de dimensiones hasta el descubrimiento de los datos esperados. Y finalmente las técnicas auxiliares son aquellas que se apoyan de manera superficial y limitada a las técnicas de predicción y descriptivas debido a que buscan la verificación de variables (Pérez & Santín, 2008).

Tabla 3 Técnicas predictiva

Técnicas	
Predictiva	
Regresión	Es una herramienta ampliamente utilizada en ingeniería especialmente en informática, big data y extracción de datos.
Series temporales	Se las puede utilizar para describir una variedad de eventos a largo plazo, son trabajos remotos basados en DTW.
Análisis varianza y covarianza	Es una forma de comparar uno o más métodos, en el cual se utiliza la t de student. Por otro lado, el análisis de la covarianza se emplea dos variables razonables.
Métodos bayesianos	Es un método de inferencia estadística, donde se puede inferir la probabilidad de una hipótesis si es cierta.
Algoritmos genéticos	Es un proceso de búsqueda de un problema concreto, donde se utiliza mecanismos que simulan los de la evolución de las especies.

Elaboración propia

Tabla 4 Técnicas descriptiva

Técnicas	
Descriptiva	
Asociación	Técnica importante en la extracción de datos donde implica encontrar asociaciones interesantes en forma de factores que elevan entre los valores de las propiedades de os objetos en una base de datos.
Reducción de la	Es un requisito importante para un sistema de aprendizaje,

dimensión	ejecutando la reducción de número de variables en la recopilación de datos.
Análisis explicativo	Es el campo de la estadística y la informática con la aplicación de intentos para identificar patrones en grandes bases de datos.
Escala multidimensional	El modelo de visualización puede ser 2D, 3D o multidimensional. Se han desarrollado varias herramientas visuales para integrarse con los datos, y algunas funciones sobre este tema se han incluido en [VIS95].

Elaboración propia

1.3.1 Reglas de asociación

La característica de las reglas de asociación es que describe la correlación entre elementos de un conjunto de datos relevantes.

Son reglas que clasifican casos, en los que se pueden ejecutar árboles de decisión y patrones a partir de datos de entrada. La información de entrada es un conjunto de variables o atributos.

Las técnicas de regla de asociación utilizan diferentes algoritmos ayudando a optimizar los datos, teniendo en cuenta la supervivencia de los mejores para que puedan adaptarse, así se puede mejorar las tareas de las organizaciones y optimizar el rendimiento de la misma (González C. , 2020).

Características de la regla de asociación

- Observe el apoyo y la confianza mínimos, y también se debe obedecer a un subconjunto de ellos.
- Si algún elemento no está limitado por el nivel más bajo, no debe considerarse un superconjunto.
- Genere el resultado de una sola regla para construir dos o más en secuencia.
- El trabajo realizado depende de la cobertura mínima requerida.
- Si un grupo de elementos no pasa la prueba de soporte, ninguno de sus superconjuntos pasará.

Probablemente la extensión más exitosa es FP-Growth, porque puede calcular de manera eficiente el conjunto de elementos frecuentes en el ejemplo utilizando la estructura de datos de árbol FP. Utiliza un método que puede descomponer tareas en subtareas más pequeñas (Berzal, 2018).

Por ejemplo, al realizar compras online, se tienen en cuenta las sugerencias, que explican que un cliente que compró un determinado artículo también realizó una compra de otro

artículo. Este es un ejemplo de reglas de asociación. A la hora de saber qué elementos coexisten más, el algoritmo FP-Growth juega un papel importante (Berzal, 2018).

Las reglas de asociación que las de categoría funcionan con atributos discretos. Por esa razón, hay que acaecer enfoques comunes y discretos antaño de cavar jerarquías predefinidas. Asimismo, hay que entrar acrecentar la soltura y acelerar la largo de las reglas no siempre que te encuentras una menstruación de academia quiere mencionar que sea útil (López A. S., 2017).

Las reglas de asociación tienen muchas aplicaciones diferentes, tales como:

- Ayudar a tomar una decisión
- Diagnosticar y anticipar alarmas en las comunicaciones
- Análisis de información de ventas
- Distribución de mercancías en almacenes
- Segmentación de clientes en base a hábitos de compra
- Es como reglas de calificación
- Sin embargo, también se encuentra usando el proceso de superposición, en el lado derecho de la base puede aparecer cualquier par de atributos de valor.
- Para encontrar este tipo de base, debe considerar todas las combinaciones posibles de pares de atributos y valores de enteros.
- Para recortarlo más tarde usando:
    - ✓ Cobertura: el número de casos predichos correctamente
    - ✓ Exactitud: la proporción del número de casos en los que se aplica la regla

Métricas

El soporte es una métrica importante, porque las reglas con un soporte muy bajo solo pueden aparecer por casualidad. Desde una perspectiva empresarial, la regla de bajo soporte no tiene sentido, porque no tiene sentido promocionar productos que los clientes rara vez compran al mismo tiempo. Por lo tanto, el soporte generalmente se usa para eliminar reglas sin sentido. Además, el grado de soporte tiene una propiedad deseada, que se puede utilizar para el descubrimiento eficaz de las reglas de asociación (Martínez, 2020).

La confianza es confiable para el argumento basado en reglas para una regla dada $X \Longrightarrow Y$, cuanto mayor sea la confianza, más probable es que incluya Y en la transacción de X. La probabilidad condicional de Y también se puede estimar para una X dada (Amat, 2018).

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} = \frac{s(X \cup Y)}{s(X)}$$

1.3.2 Clustering

El Clustering o la clusterización es una instrucción importante adentro del Machine learning. Este pensamiento desarrolla una acción cardinal que le permite a los algoritmos de instrucción automatizada lavar el cerebro y conocer de manera adecuada los datos con los que desarrollan sus actividades.

El clustering es una ocupación que tiene como ártico principal salir el agrupamiento de conjuntos de objetos no etiquetados, para durar construir subconjuntos de datos amigos como Clústeres. Cada clúster adentro de un grafo está simpatizante por un sumario de objetos o datos que a términos de grafología resultan similares entre sí, pero que poseen nociones diferenciales con respecto a otros objetos pertenecientes a la totalidad de datos y que pueden conformar un clúster independiente (Monrroy, 2016).

Tipos de validación

La validación externa y la validación interna son los dos tipos más importantes de validación de grupo. La principal diferencia es si se utiliza información externa para la verificación, es decir, información que no es producto de la tecnología de ensamblaje utilizada.

A diferencia de las técnicas de validación externa, las técnicas de validación interna miden el ensamblaje basándose únicamente en información de datos. Evalúan la calidad de la estructura del ensamblaje sin más información que el propio algoritmo y sus resultados.

Dado que la validación externa mide la calidad de la agregación al conocer la información externa de antemano, se utiliza principalmente para determinar el algoritmo de agregación óptimo en un conjunto de datos determinado.

Las métricas de validación interna se pueden utilizar para elegir el mejor algoritmo de agrupamiento, así como el número óptimo de agrupaciones sin ninguna información adicional.

En la práctica, la información externa, como las etiquetas de clase, a menudo no está disponible en muchos casos de aplicación.

Métricas de Validación Externa

Cuando se tiene información externa, como la categoría de cada dato, el siguiente análisis es común y ampliamente utilizado: tiene la categoría de cada dato en el conjunto de datos, es decir, tiene de antemano cuántos grupos y a qué grupo pertenece cada dato.

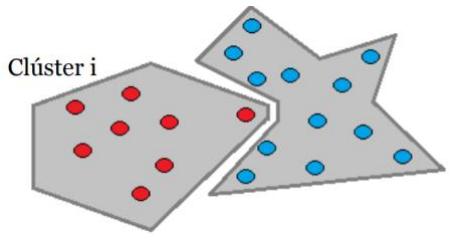


Figura 9 Métricas de validación externa

Una vez que se ha realizado el agrupamiento utilizando un algoritmo previsto para tal fin (k-means, DBSCAN) el algoritmo puede proponer un nuevo conjunto de datos, diferente de los definidos por las clases previamente conocidas:

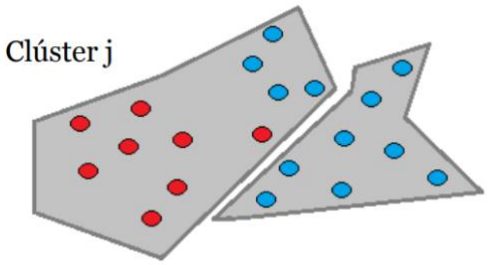


Figura 10 Metricas de validación

Como se mencionó anteriormente, la idea de una validación externa es que una vez que se completa el algoritmo de agregación, el conjunto al que se asigna cada elemento se compara con la etiqueta de clase que se le dio anteriormente (información externa).

Métodos de Clustering

- **Algoritmo de k-medias**

Es un método clásico en el que se aplica un proceso de agrupamiento, en este caso el algoritmo busca los mejores centroides para realizar la correlación de manera que los mejores centroides estén lo más cerca (Monrroy, 2016).

El algoritmo de k-modes comienza dividiendo el conjunto de datos en grupos (k grupos) y luego refinando iterativamente cada grupo mediante operaciones repetidas de fusión y división.

El algoritmo k-Modes es la primera extensión del algoritmo k-Means para la agregación de datos categóricos. Sigue la misma idea del algoritmo k-Means y la estructura del algoritmo no cambia, siendo la principal diferencia la medida de similitud utilizada para comparar objetos.

Las principales características de este algoritmo son:

- Usa la escala de diferencias para comparar cosas.
- Reemplace el uso de promedios con modas.
- Utilice un método basado en la frecuencia para actualizar los modos.

El algoritmo k-Modes está diseñado para la agregación exclusiva de grandes conjuntos de datos categóricos.

El algoritmo k-mean resuelve el problema de optimización, que es una función para mejorar (reducir) la suma de las distancias al cuadrado de cada objeto a su centro de grupo.

Los objetos están representados por un vector real de tamaño d (x1, x2, ..., xn) y el algoritmo de k medias construye k grupos donde la suma de las distancias de los objetos es la más pequeña, en cada grupo S = {S1, S2, …, Sk}, en su centro. El problema se puede formular de la siguiente manera:

$$\min_{\mathbf{S}} E(\boldsymbol{\mu_i}) = \min_{\mathbf{S}} \sum_{i=1}^{\kappa} \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2$$

donde S es el conjunto de datos cuyos elementos son los objetos xj representados por vectores, donde cada uno de sus elementos representa una característica o atributo. Tendremos k grupos o clusters con su correspondiente centroide μi.

La principal ventaja del método k-mean es que es simple y rápido. Pero el valor de k debe especificarse y el resultado final depende de la configuración de los baricentros. En principio, no converge con un mínimo global sino hacia un mínimo local.

- **Clustering jerárquico**

Es un método de más visualización práctica porque contiene una forma de damerograma, se lo puede realizar tanto en forma divisiva o aglomeraría. Permite analizar alternativas diferentes de los grupos. Dependiendo de los objetivos del proyecto se podrá resolver eligiendo el grupo específico que durante el proceso permite ajustar a resolver el problema (Roldán, 2015).

La elección de una métrica apropiada influenciará la forma de los grupos, ya que algunos pueden estar cerca unos de otros de acuerdo a una distancia y más lejos de acuerdo a otra. Por ejemplo, en un espacio 2-dimensional, la distancia entre el punto (1,0) y el origen (0,0) es siempre 1 de acuerdo a las normas usuales, pero la distancia entre el punto (1,1) y el origen (0,0) puede ser 2,$\sqrt{2}$ o 1 bajo la distancia Manhattan, la distancia euclidiana o la distancia máxima respectivamente.

Algunos de los indicadores comúnmente utilizados para la agrupación jerárquica son:

Names	Formula
Distancia euclidiana	$\|a-b\|_2 = \sqrt{\sum_i (a_i - b_i)^2}$
Distancia euclidiana al cuadrado	$\|a-b\|_2^2 = \sum_i (a_i - b_i)^2$
Distancia Manhattan	$\|a-b\|_1 = \sum_i \|a_i - b_i\|$
distancia máxima	$\|a-b\|_\infty = \max_i \|a_i - b_i\|$
Distancia de Mahalanobis	$\sqrt{(a-b)^\top S^{-1}(a-b)}$ donde S es la matriz de covarianza
Similitud coseno	$\frac{a \cdot b}{\|a\|\|b\|}$

Figura 11 Indicadores para la agrupación jerárquica

La agrupación jerárquica tiene la ventaja obvia de que se puede utilizar cualquier medida de distancia. De hecho, las notas en sí no son necesarias: solo use la matriz de distancia.

1.4 Scrum

Es un proceso en el que se aplican regularmente un conjunto de buenas prácticas, colaboran en forma de equipo y obtienen los mejores resultados del proyecto. Estas prácticas se apoyan entre sí y su elección surge del estudio de métodos eficientes de trabajo en equipo. En Scrum, se carga de la entrega parcial y regular del producto final donde la innovación, la productividad y la competitividad son elementales para priorizar beneficios que aportan al receptor del proyecto (Subra & Vannieuwenhuyze, 2018).

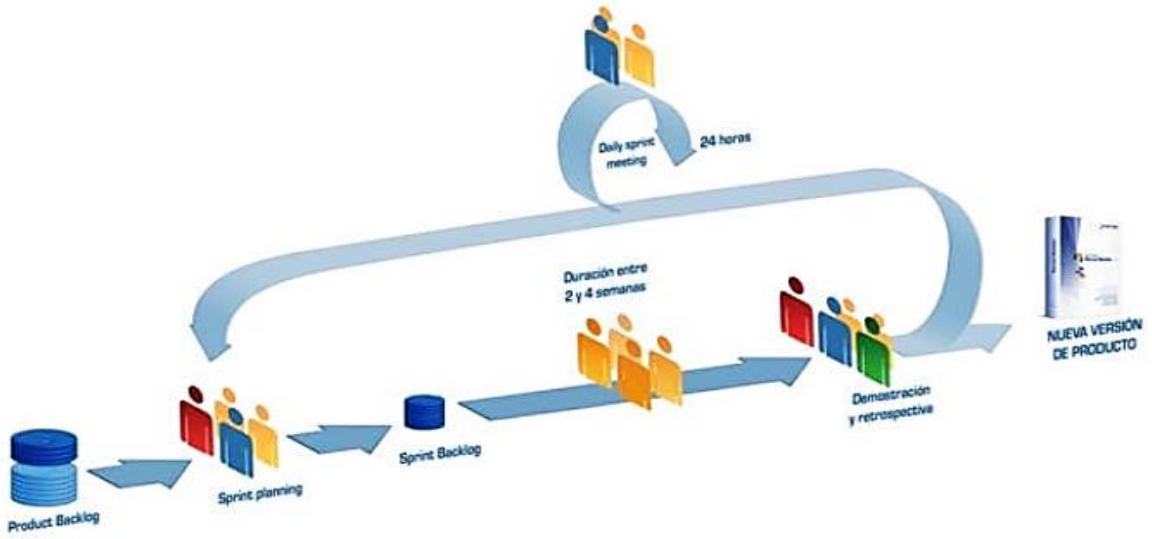


Figura 12 Metodología Scrum

Fuente: Dimes (2015)

Por tanto, Scrum es miembro por las entidades con la finalidad generar especialmente un entorno adecuado para el desarrollo planes en entornos complejos, en los que necesita obtener resultados lo antes posible; para ello se debe tomar en cuenta que los requisitos cambian o están poco definidos por motivos de innovación, la competitividad, la flexibilidad y la productividad. Adicional se lo puede utilizar de resolver situaciones en el que no se cumple con las exigencias o necesidades de los clientes, es decir se encarga de solventar la situación que no puede satisfacer las necesidades de los clientes con respecto al tiempo de entrega que en ocasiones suele ser demasiado largo, el costo se dispara o la calidad es inaceptable (Subra & Vannieuwenhuyze, 2018).

El Scrum es un salvavidas para las organizaciones que se enfrentan a dificultades que les limite continuar con una metodología de Cascada o que ni siquiera empleen algún tipo de software que facilita el manejo de información. Es decir, el Scrum es un instrumento que contribuye en la creación de un software que facilite la gestión de procesamiento de información de manera fácil, rápida u sencilla (Dimes, 2015).

En otras palabras, se puede decir expresar que el Scrum es un instrumento que gestiona el desarrollo de in software cuya finalidad es maximizar el retorno de inversión de la entidad, además de acrecentar la productividad y competitividad de la organización.

1.4.1 Cuando se utiliza

Este método se emplea cuando el cliente se entusiasma y busca que su proyecto crezca de forma continua, dicha interacción permite que cualquier persona u organización al momento a linear el software a los objetivos corporativos accedan y estén predispuestos al cambio. Debido a que este suceso invita a todos los miembros de la organización a laborar de forma activa en los procesos de innovación, a través de la motivación, compromiso y trabajo en equipo.

1.4.2 Proceso del Scrum

En Scrum, los proyectos se ejecutan en un período corto de tiempo y una duración fija (la iteración suele ser de 2 semanas, aunque en algunos equipos es de 3 a 4 semanas, que es el límite máximo de retroalimentación para objetos y reflejos). Cada iteración debe proporcionar un resultado completo, es decir, el incremento del producto final puede entregarse al cliente con el mínimo esfuerzo según sea necesario.

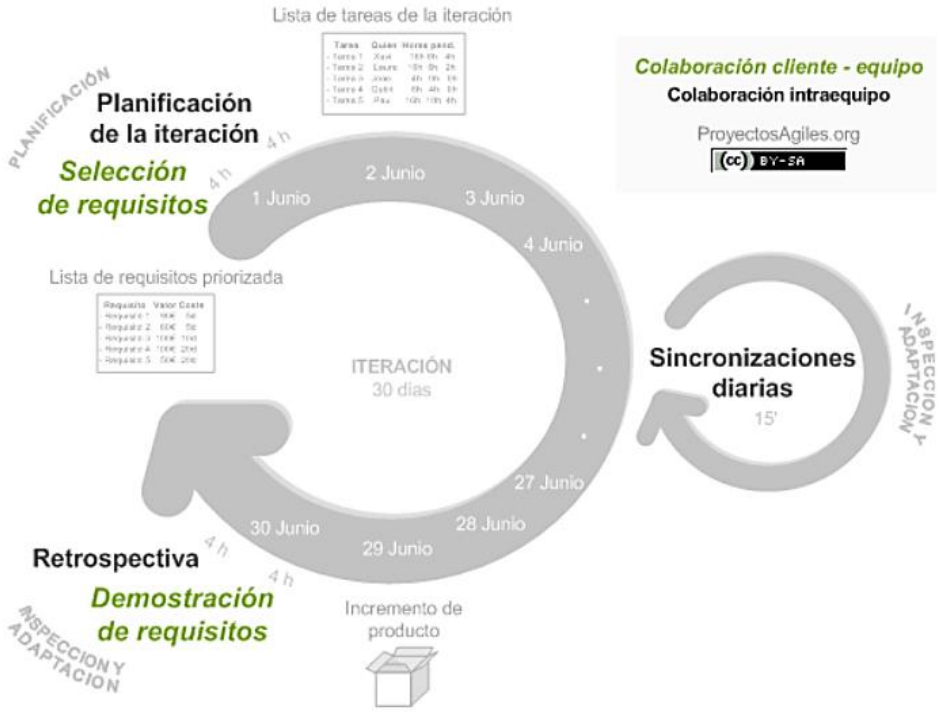


Figura 13 Procedo de la metodología Scrum

Fuente: Rad & Turley (2019)

El proceso comienza con una lista de objetivos / requisitos prioritarios para el producto, que sirve como plan de proyecto. En esta lista, el cliente (propietario del producto) prioriza los objetivos y proporciona un precio equilibrado basado en el costo (el equipo estima según la "definición de finalización") y lo divide en iteración y entrega (Rad & Turley, 2019).

1.4.3 Roles del Scrum

Los roles de Scrum son muy importantes para asegurar la implementación de algún proyecto, Scrum tiene tres roles principales.

- **Product Owner** (Dueño del producto)

Es el responsable de decidir sobre el trabajo que necesita hacerse o maximizar el producto, provecto, que se esté ejecutando y se compone de las siguientes tareas.

- ✓ Gestiona prioridades.
- ✓ Representante del negocio.
- ✓ Intraemprendedor.

- **El Scrum Master**

Es una persona servicial que ayuda al equipo a tomar las mejores decisiones para la utilización de la metodología de Scrum, su principal función es la responsabilidad del ROI en el proyecto.

- **Team (Equipo de desarrollo)**

Es un conjunto de profesionales, que poseen conocimientos necesarios que permiten ejecutar el proyecto, se compromete al inicio de cada sprint. Sprint es el tiempo o el periodo de duración que va de 1 a 4 semanas, que sean de preferencia con intervalos cortos.

1.4.4 Fases del Scrum

Las fases de la metodología de Scrum se resumen en 5 pasos o etapas de implementación.

1 **Inicio**

En esta etapa, examinará su trabajo y descubrirá las necesidades reales de su carrera. Las preguntas que surgen son: ¿Qué quieres? ¿Qué debo hacer?

2 **Planificación y estimación**

En esta etapa se incluye los siguientes pasos:

- ✓ Crear y comprometer historias de usuario.
- ✓ Identificar y estimar tareas.
- ✓ Crear el sprint backlog o interacción de tareas.

Esta puede ser una de las etapas más importantes de un proyecto, ya que las tareas se asignan a cada grupo y el tiempo se calcula de acuerdo con la prioridad del proyecto.

3 **Implementación**

En la etapa de implementación se realiza un concilio en donde se discuten el sprint y se explora como desarrollar el encargo de los grupos en el proyecto, tiene los siguientes procesos:

- ✓ Crear entregables.
- ✓ Realizar daily stand-up.
- ✓ Refinanciamiento del backlog.

4 **Revisión y retrospectiva**

Después de que se haya completado la preparación e implementación, el proyecto debe revisarse, esto quiere decir criticar o evaluar el trabajo del equipo local.

Los pasos importantes para implementar en esta etapa incluyen:

- ✓ Demostrar y validar el sprint.

- ✓ Retrospectiva del sprint.

5 Lanzamiento

Volverá a los resultados del trabajo y la entrega del producto. Aquí debe reali zar dos tareas específicas.

- ✓ Enviar entregables.
- ✓ Enviar retrospectiva del proyecto.

- **Conclusiones**

  Las técnicas de diseño Scrum son muy útiles a la hora de desarrollar software, pero también se utilizan en todo tipo de negocios y operaciones donde la colaboración es importante.

1.4.5 Beneficios del Scrum

Los principales beneficios que proporciona la implementación de la metodología Scrum son las siguientes:

- **Cumplimiento de expectativas.**

  Los clientes, consumidores y usuarios al convivir con un entorno altamente cambiante y dinámico provocan que sus expectativas cambien, es por ello que las organizaciones deben mantenerse a la par del desarrollo tecnológico con el afán de lograr mantenerse en el mercado y cumplir los objetivos corporativos.

- **Flexibilidad a los cambios.**

  Hace referencia a la capacidad de reacción y cambio que tiene las organizaciones ante permutaciones que requieren innovación y evolución para satisfacer las necesidades de los consumidores o clientes.

- **Disminución del Time to Market.**

  Se genera cuando los clientes empiezan a realizar funcionalidades más relevantes antes que culmine el proyecto.

- **Mejora la calidad del software.**

  Conforme a la metódica de trabajo, la necesidad y capacidad para cumplir con la versión de funcionalidad el software mejorar y brindará a los usuarios mayor satisfacción en el uso.

- **Acrecienta el nivel de productividad y competitividad de la empresa.**

  Eliminado la burocracia a nivel estructural en la organización y la adaptación de estrategias que incentiven al personal se lograr obtener un mayor nivel de productividad y competitividad del personal lo cual a su se verá reflejado en los resultados de la entidad.

- **Maximiza e impulsa el retorno de la inversión (ROI).**
  La implementación de la metodología únicamente con las prestaciones logrará que la institución recupere su inversión inicial.
- **Reduce los tiempos de ocio y optimiza los recursos.**
  Establecido la metodología coadyuvará a establecer los tiempos de cada uno de los procesos de gestión empresarial, lo cual optimizará los recursos como el tiempo de perdida que generaría una determina tarea.
- **Minimiza el riesgo.**
  El hecho de proceder con la implementación de una metodología nueva en una organización representa un alto riesgo en el caso de que el personal no esté dispuesta a trabajar y ser partícipe del cambio (Rad & Turley, 2019).

CAPÍTULO 2

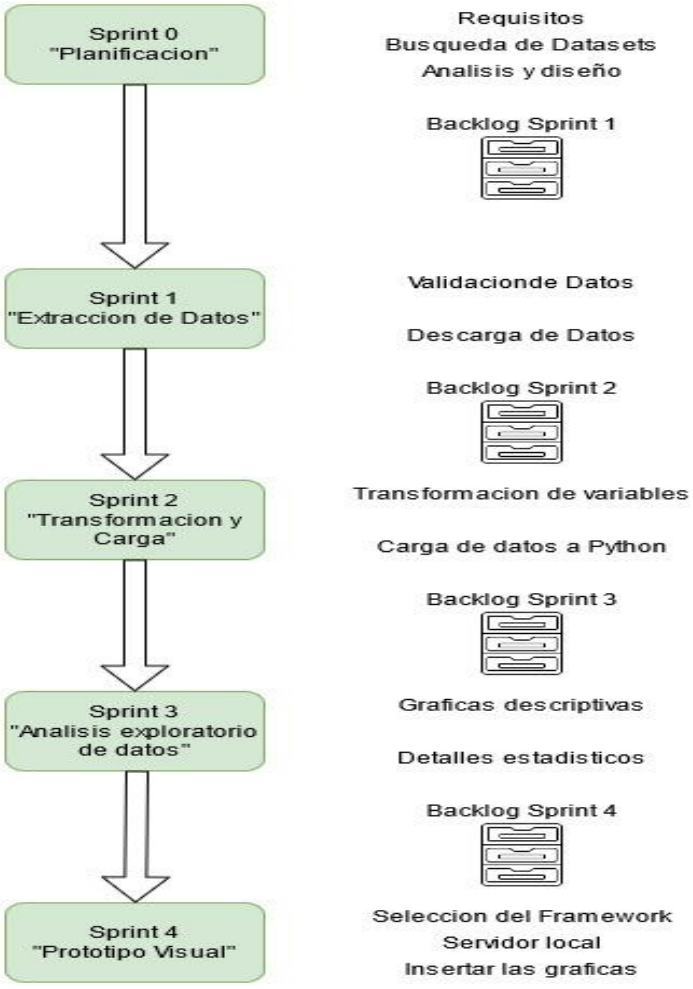
Desarrollo

2.1 Recolección de datos

2.1.1 Metodología AGILE SCRUM para el prototipo

Para la realización del prototipo se ha utilizado la metodología Agile Scrum que permite visibilizar el desarrollo del prototipo al mismo tiempo que flexibiliza los métodos y cambios mientras se desarrolla. Hemos dividido el desarrollo del prototipo inicial en 5 Sprint de trabajo (contando la planificación), como se muestra en la gráfica siguiente:

Figura 14 Sprint para el desarrollo del prototipo.



La metodología de Agile Scrum se basa en los siguientes principios:

- Transparencia.
- Responsabilidad.
- Mejora continua.
- Individuos e interacciones sobre procesos y herramientas.
- Producto de trabajo al final de cada iteración.

Se utilizan varias técnicas para la implementación de la metodología Agile Scrum, que se incluyen en este trabajo. Para implementar estas técnicas, podemos utilizar diversas herramientas que permitirán que todo el trabajo sea transparente y visible para todos los participantes de un proyecto.

Para crear las historias de usuario y medir así la cantidad de trabajo realizada en cada Sprint se utilizó la plataforma Trello online.

Trello es una herramienta de colaboración visual que le permite organizar y priorizar proyectos de una manera divertida, flexible y gratificante. Un tablero de Trello es una serie de listas, con un montón de tarjetas adjuntas y repletas de potentes funciones y automatización.

2.1.2 Limpieza de datos

Para realizar la limpieza de datos primero encontramos los valores vacios en el set de datos.

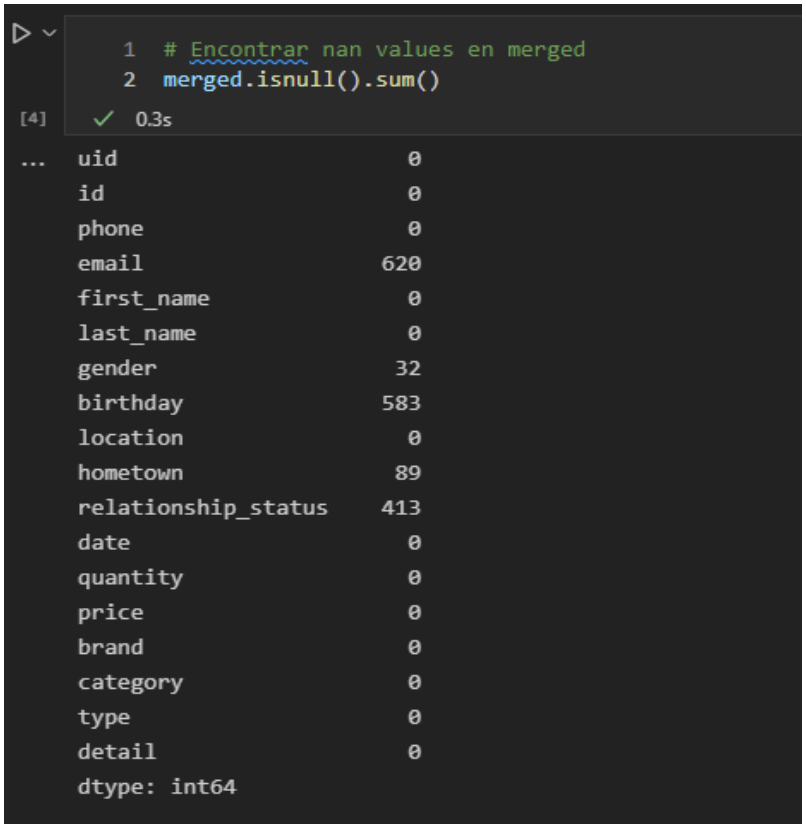


Figura 15 Limpieza de datos

Luego vimos si existen caracteres especiales en los nombre

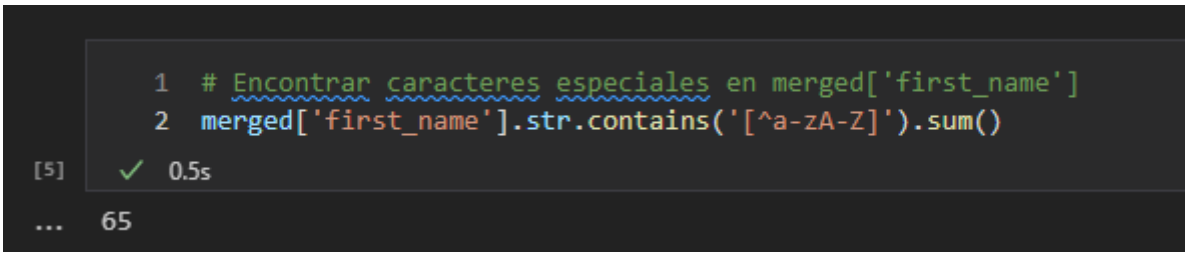


Figura 16 Caracteres

Como hubo varios caracteres no solo en nombre sino en apellidos y cuidades se utilizó una transformación de caracteres especiales a comunes utilizando el método “maketrans” para todas las columnas q contienen caracteres especiales.

Luego ya que el programa por defecto tiene el idioma indones, se reemplazaron estas palabras por su equivalente al inglés.

```
# Eliminar caracteres especiales usando str.maketrans
a,b = 'áéíóúüñÁÉÍÓÚÜÑ','aeiouunAEIOUUN'
trans = str.maketrans(a,b)

# Aplicar translate
merged['first_name'].str.translate(trans)
merged['first_name'] = merged['first_name'].str.translate(trans)
merged['last_name'] = merged['last_name'].str.translate(trans)
merged['birthday'] = merged['birthday'].str.translate(trans)
merged['location'] = merged['location'].str.translate(trans)
merged['hometown'] = merged['hometown'].str.translate(trans)
merged['relationship_status'] = merged['relationship_status'].str.translate(trans)
merged['brand'] = merged['brand'].str.translate(trans)
merged['category'] = merged['category'].str.translate(trans)
merged['type'] = merged['type'].str.translate(trans)
merged['detail'] = merged['detail'].str.translate(trans)
```

✓ 0.5s

```
# Replace หมั้นแล้ว to Engaged in merged["relationship_status"]
merged['relationship_status'] = merged['relationship_status'].replace({'หมั้นแล้ว':'Engaged', 'มีแฟนแล้ว':'In a relationship',
complicated"})
```

Para finalizar se cambió el formato del precio que contenía el signo “$” a valores numéricos tipo float utilizando:

```
compradores['price'] = compradores['price'].str.replace(",","").str.extract(r'([0-9]+)', expand = False)
compradores['price'] = compradores['price'].astype(float)
compradores['price'].describe()
```

[10]

```
count     623.000000
mean      380.876404
std       461.357332
min         8.000000
25%        40.000000
50%       125.000000
75%       499.000000
max      1350.000000
Name: price, dtype: float64
```

2.1.3 Datos demográficos de Tulcán

De acuerdo al Instituto Nacional de Estadísticas y Censos INEC (2017), la última proyección de la población del cantón Tulcán calculada para el año 2020 es de 102,395 habitantes. Teniendo esto en consideración y los resultados del Fascículo Tulcán correspondientes al censo del 25 de noviembre de 2001, se pueden inferir los datos de la

población de Tulcán siguiendo la distribución de este último censo (INEC, Fascículo Provincial Carchi, 2010).

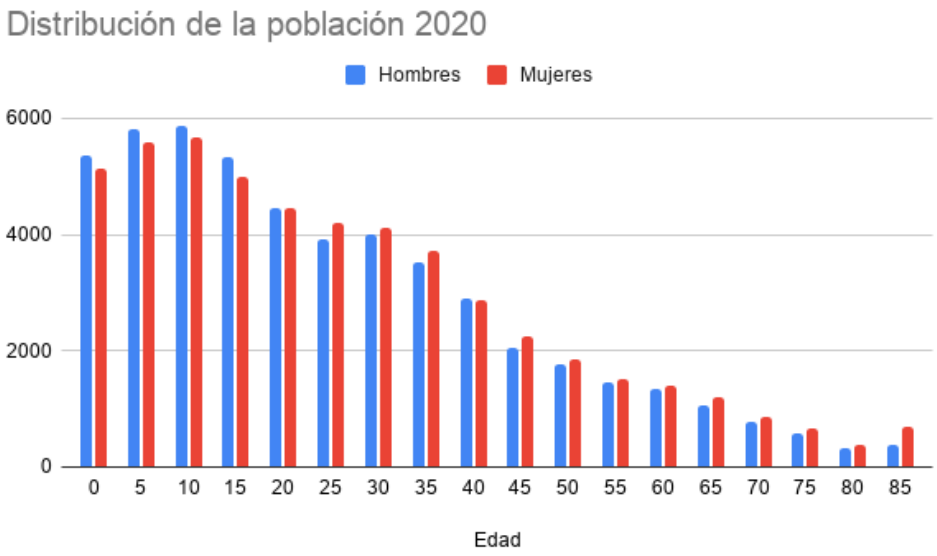


Figura 17 Distribución por edad y género de la población de Tulcán

Del mismo modo se han calculado el número de habitantes clasificados por nivel de instrucción.

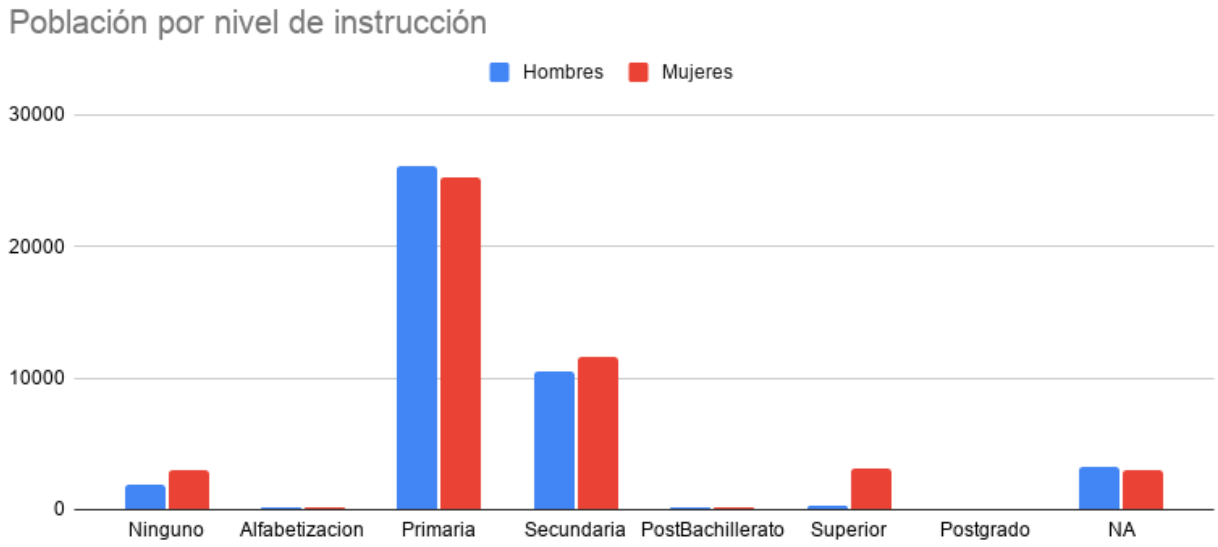


Figura 18 Población por género y nivel de instrucción

Así mismo se han obtenido las gráficas de acuerdo a la ocupación de la población de Tulcán.

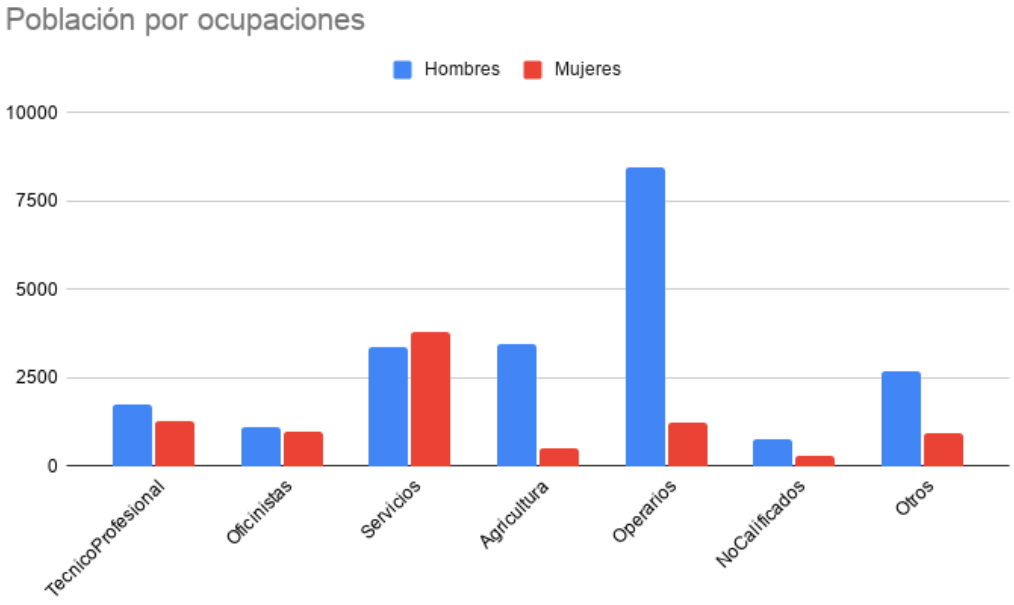


Figura 19 Población por ocupación

Y por último, para el estado civil.

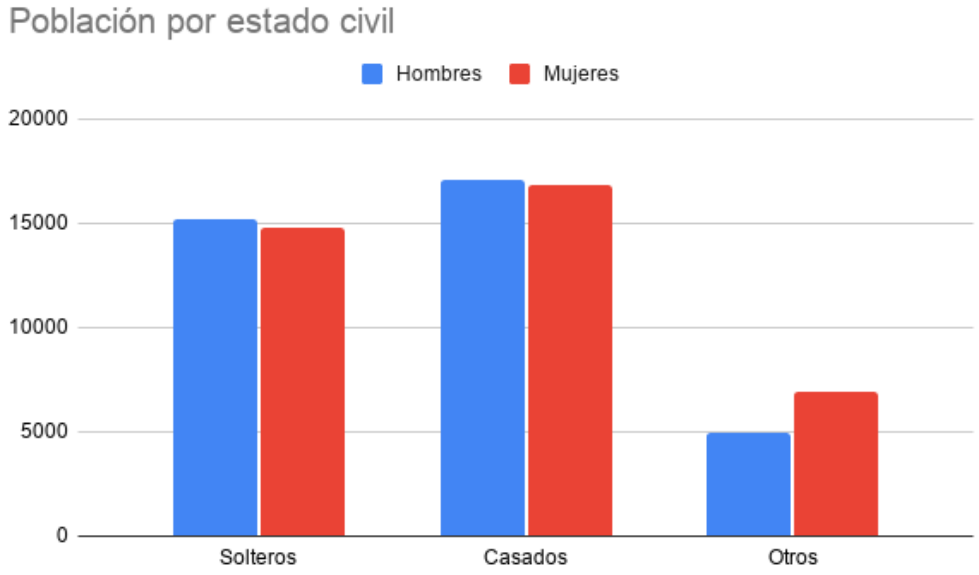


Figura 20 Población por estado civil

2.1.4 Ciclismo en Tulcán

El INEC (2017) con razón del día mundial de la bicicleta publicó un artículo en 2016 con estadísticas sobre el uso de bicicletas en Ecuador. De donde se resalta que el 49.38% de la población ecuatoriana maneja bicicleta por lo menos una vez a la semana y que el 34.09%

utiliza la bicicleta a diario. Además, el 38.3% de estos ciclistas tienen edades comprendidas entre 5 y 14 años (INEC, INEC, 2021).

Por otro lado, el Gobierno Autónomo Descentralizado Municipal de Tulcán, GAD de Tulcán, con motivo de la creación de ciclo vías emergentes para la cuidad de Tulcán realizó por su parte un análisis de viabilidad (Gobierno Autónomo Descentralizado Municipal de Tulcán, 2020). Del cual se destaca que el 90.27% de la muestra aseguraron tener 1 o más bicicletas en su hogar. Además, que las distancias recorridas por los ciclistas tienen la siguiente distribución: (Túlcan, 2021).

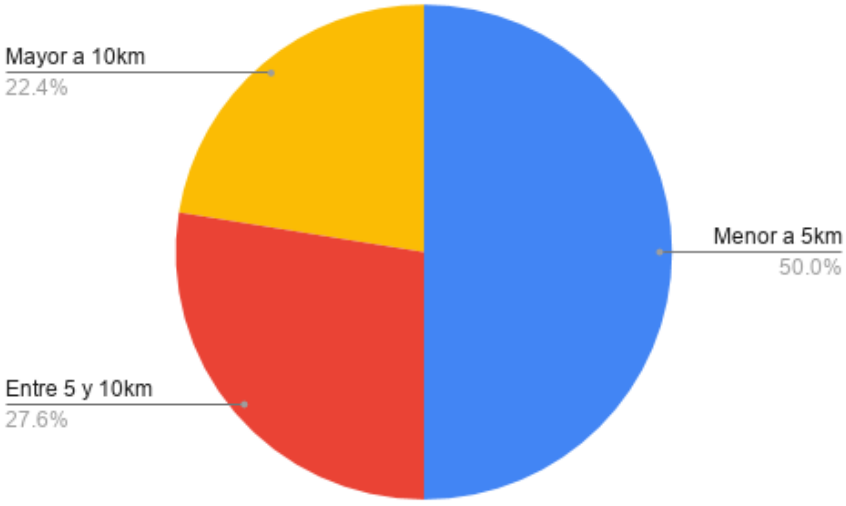


Figura 21 Distribución de ciclistas por recorrido

2.1.5 Ventas de bicicletas

Para poder centrarse en un segmento de mercado es necesario comprender quienes son los clientes y quienes son los usuarios de las bicicletas. Para esta investigación se cuenta con datos de una empresa importadora nacional de bicicletas y accesorios la cual nos ha facilitado sus datos de venta de su sede en Tulcán (Álvaro, 2015).

Para poder obtener gráficas, las variables categóricas han sido convertidas a numéricas a través de la creación de diccionarios de Python. Estos datos fueron recolectados para compradores con datos demográficos, la distribución de estos datos es la siguiente: (Dedhia, 2021).

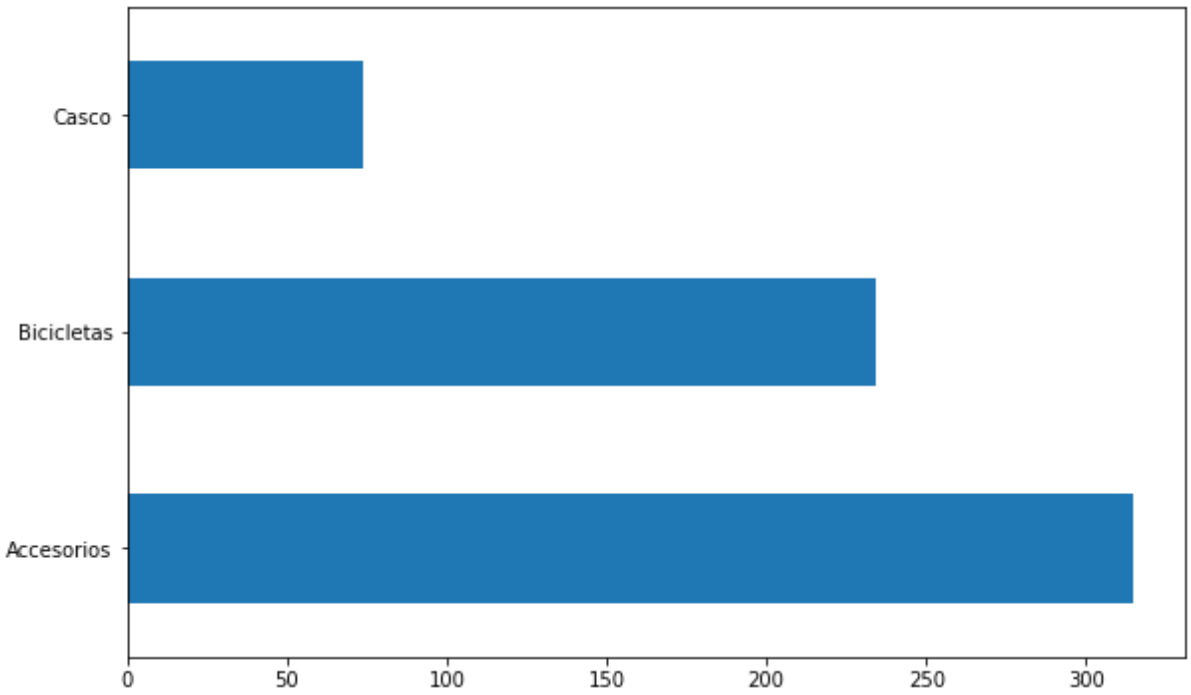


Figura 22 Compras en el almacén por categoría

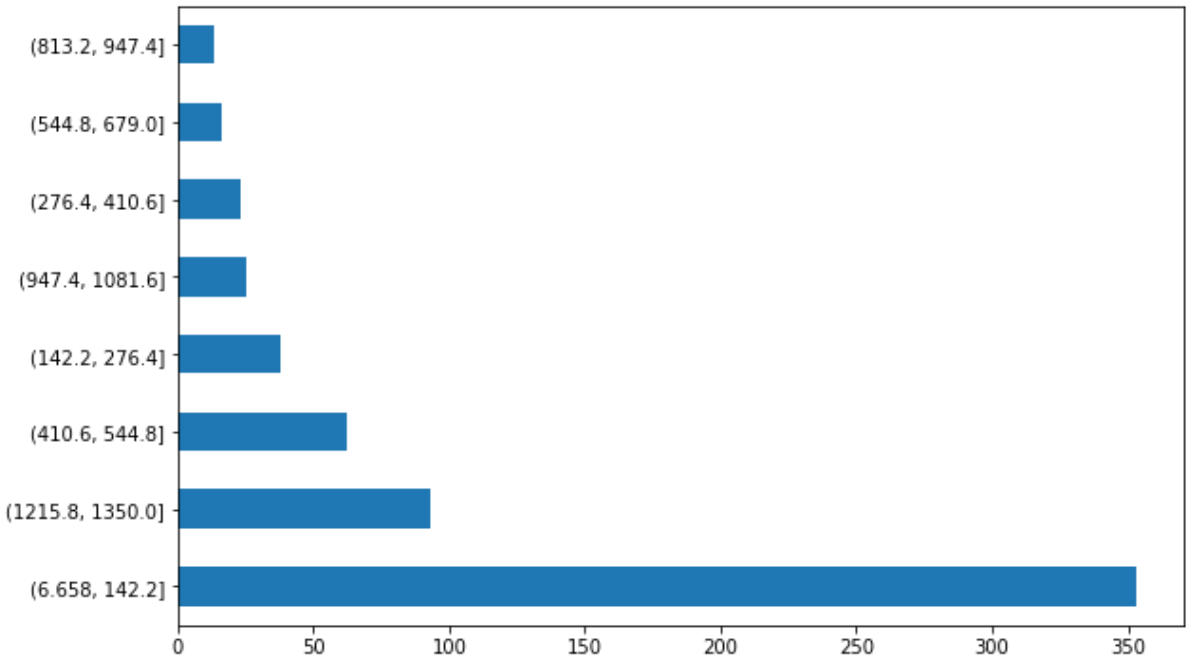


Figura 23 Compras en el almacén por rango de precios

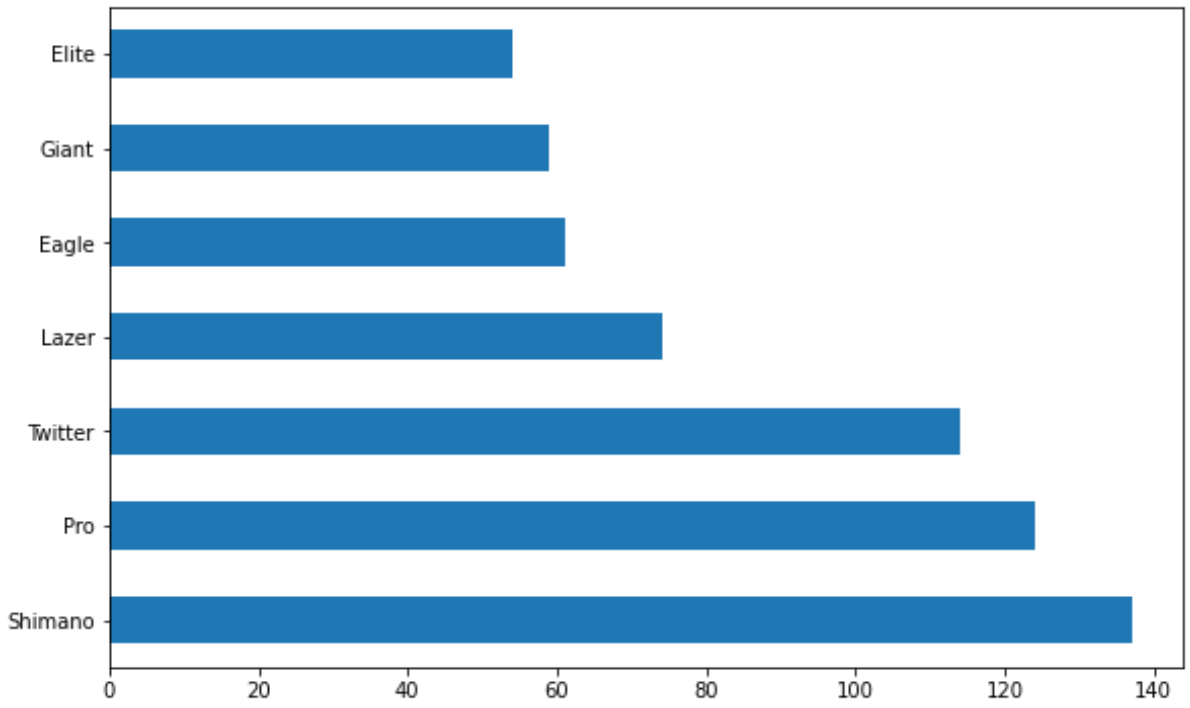


Figura 24 Compras en el almacén por marcas

2.1.6 Estados Financieros de la venta de bicicletas 2016-2020

Dado que la información de las ventas es del año 2016 y los datos de la población están inferidos con la proyección a 2020 es necesario hacer el análisis de los estados financieros de las empresas de compraventa de bicicletas y accesorios durante el periodo comprendido entre 2016 y 2020. Para tal fin, se han analizado los datos de la Superintendencia de Compañías correspondientes al CCIUU:

- G4649.92. Venta al por mayor de bicicletas, partes y accesorios incluyen los artículos deportivos.

De este análisis pudimos obtener las siguientes gráficas:

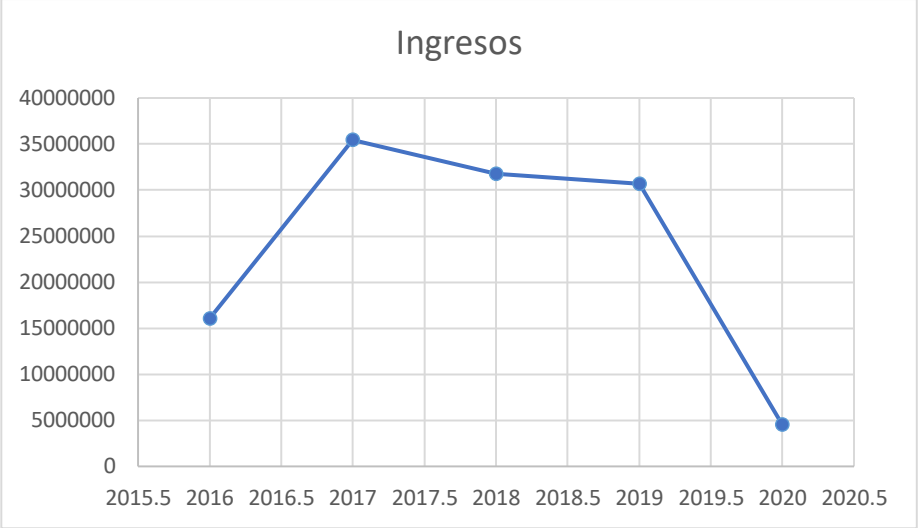


Figura 25 Ingresos por venta de bicicletas 2016-2020

En el periodo 2016-2017 la industria refleja un incremento de más del 200% en Ingresos.

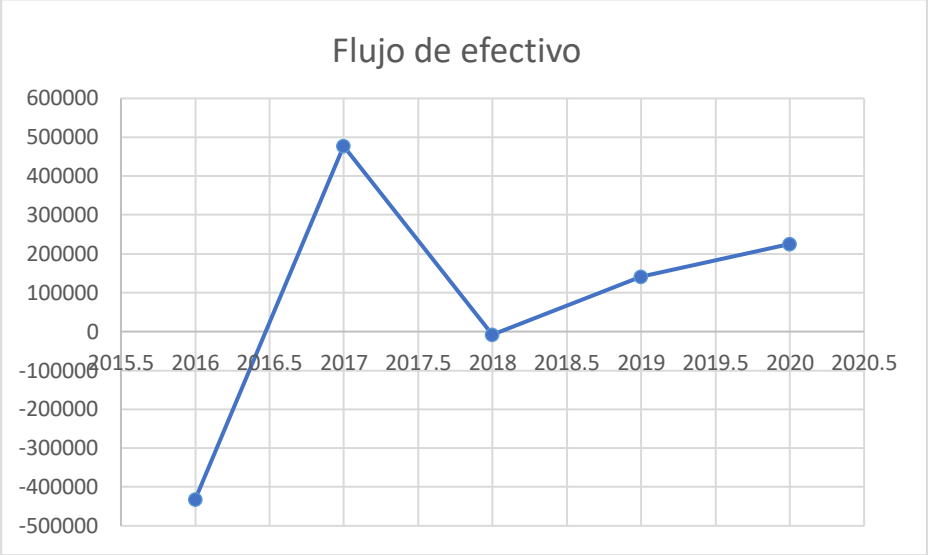


Figura 26 Flujo de efectivo 2016-2020

De igual forma para el Flujo de Efectivo de ventas de bicicletas con incremento del 111%.

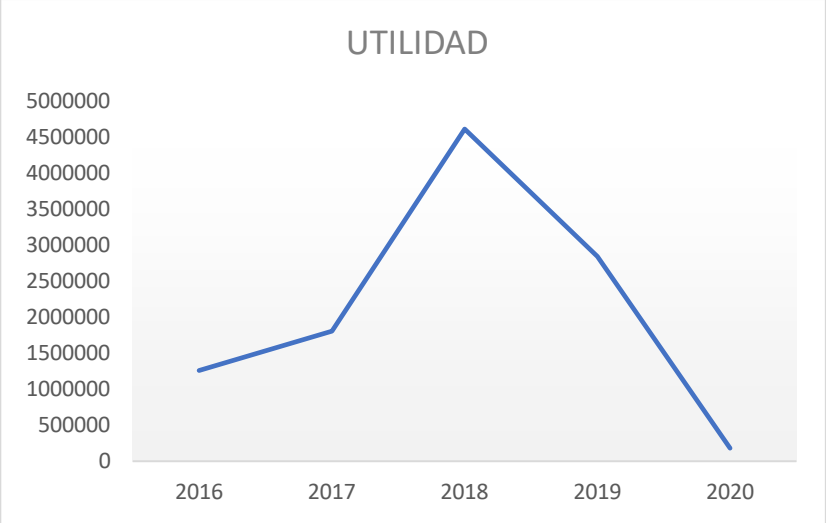


Figura 27 Utilidad 2016-2020

Mientras que la utilidad no tuvo un incremento tan significativo en dicho periodo. No obstante, en 2018 se puede ver su pico más alto con un crecimiento del 255%.

En el periodo 2017-2019 hay una caída en los ingresos hasta llegar a su punto más bajo con menos de 5'000'000.00 USD para el año 2020.

Una variable importante es la utilidad que tienen las empresas y se observa que en el periodo 2018-2020 cae abruptamente. Esto se pude explicar por la situación de la emergencia sanitaria que se desató desde principios de 2020 y que ha tenido repercusiones socioeconómicas fuertes en el territorio ecuatoriano ya que, por un periodo prolongado, los locales que no pertenecen a primera necesitad se vieron obligados a cerrar. A esto se le suma la baja capacidad de respuesta por parte de los proveedores de bicicletas ya que la demanda subió considerablemente por motivos de movilidad consecuencia de la pandemia.

2.1.7 Visualización del prototipo

Para la visualización del prototipo se ha usado el entorno de trabajo Dash. Dash es un entorno de trabajo para Python que nos ayuda a crear aplicaciones web analíticas. Dash está escrito sobre Flask, Plotly y React JS. Es ideal para construir visualizaciones con interfaces de usuario personalizadas. Dash se renderiza en cualquier navegador y se puede montar en servidores tanto locales como para esta investigación como servidores de nube. Estas características lo hacen ideal para el desarrollo de aplicaciones visuales con datos desde Python y se puede acceder desde cualquier dispositivo que tenga un navegador (González, 2018).

Se han adaptado los datos "crudos" a tipo objeto para poder así hacer uso del graficador de Dash, para esto la variable Edad se ha cortado en "bins" lo que crea un arreglo con las

edades comprendidas en un rango determinado en este caso hemos hecho un corte cada 5 años lo que nos da como resultado un arreglo de 13 categorías entre 40 y 75 años, esto a parte de ayudarnos con Dash nos permite relacionar directamente la información de compradores con los datos socio-demográficos de Tulcán directamente. A este arreglo de categorías finalmente se lo ha convertido a objecto usando la herramienta de Python “as_type()”.

Para las primeras graficas del visualizador de prototipo se ha encontrado el porcentaje de cada categoría de cada columna de la base de datos y se ha graficado esto en un arreglo de graficas tipo “pie” especificando estos últimos resultados en la gráfica además del detalle al pasar el mouse sobre los mismos (Chaturvedi, 1996).

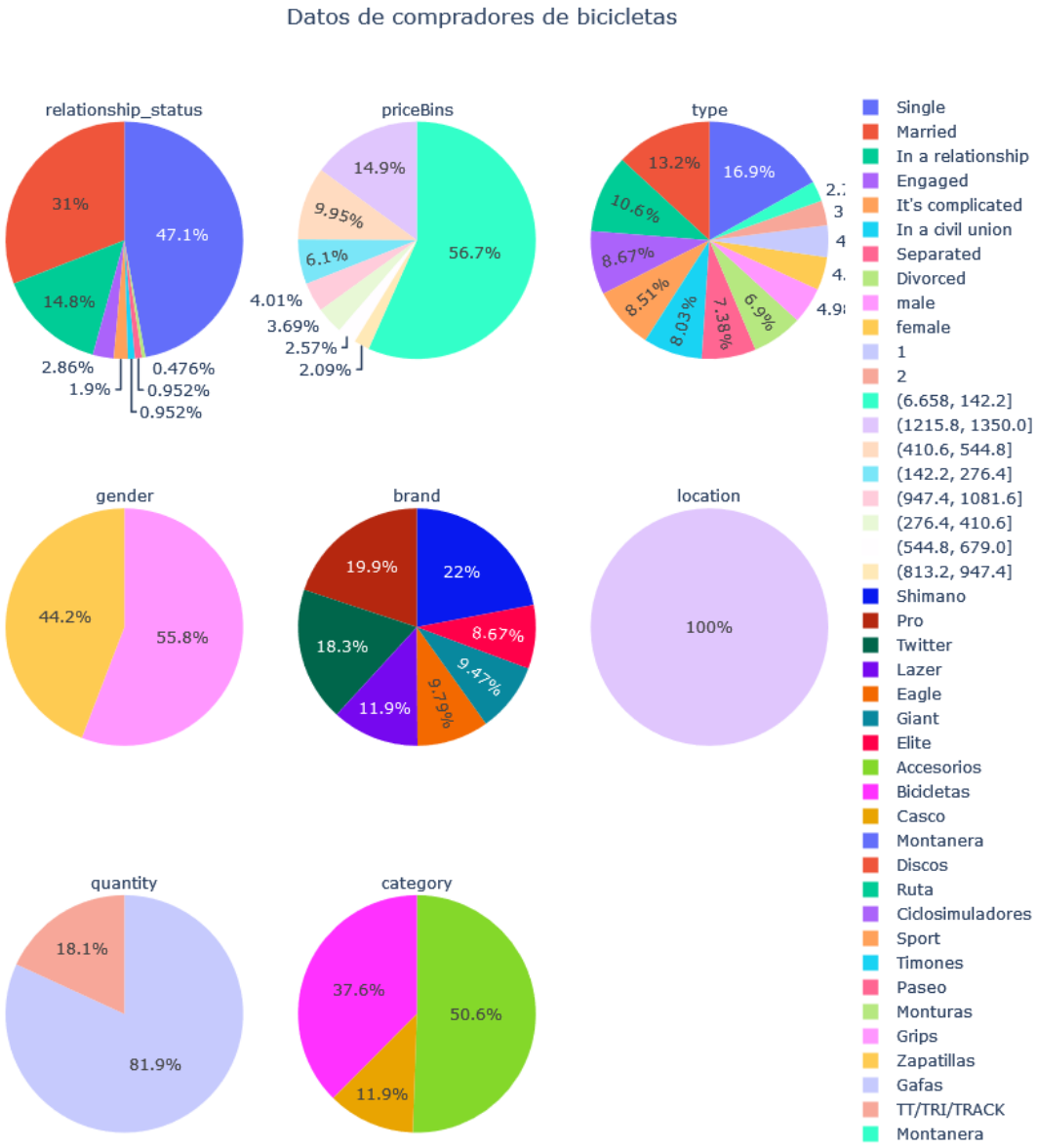
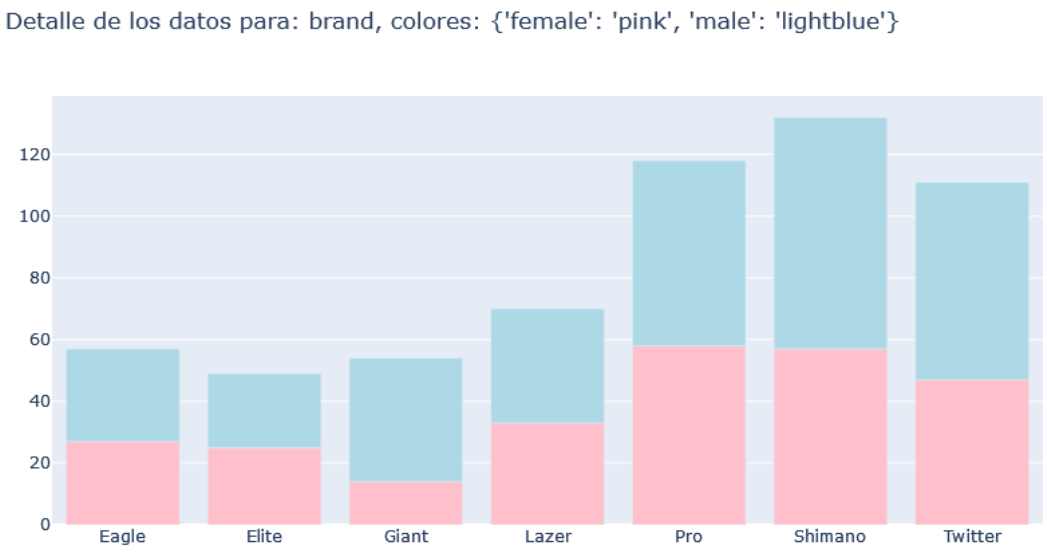


Figura 28 Gráfica de porcentajes para cada una de las columnas.

Para la gráfica de detalle, en cambio, primero se ha creado un Dropdown con todas las columnas de la base de datos, luego, para la edad se han seleccionado los colores: Rosa para Mujeres y Celeste para los Hombres, con esto junto lo que se crea es una gráfica dinámica que dependiendo de la columna que seleccionemos en el dropdown genera automáticamente el detalle de los datos, separados por genero para su revisión e inspección. Al igual que las gráficas anteriores esta muestra los detalles al pasar el mouse sobre cualquier parte del gráfico.

Figura 29 Ejemplo de grafica de Detalle/Diagnostico.



2.2 Implementación de las Reglas de asociación

2.2.1 Algoritmo Apriori

Apriori fue uno de los primeros algoritmos desarrollados para encontrar reglas de asociación y sigue siendo uno de los más utilizados, y consta de dos pasos:

- Identifica todos los elementos recurrentes que ocurren con una frecuencia superior a un cierto umbral (grupos de elementos recurrentes).
- Convierta estos grupos recurrentes en reglas de asociación.

Para ilustrar el trabajo del algoritmo, se utiliza un ejemplo simple. Considere la siguiente base de datos para un centro comercial donde cada fila representa una transacción.

El algoritmo Apriori determina los valores mínimos de los indicadores de soporte y espera para la base. Pongamos en nuestro ejemplo un apoyo mínimo del 30% y esperemos que al menos el 80%.

Este algoritmo se implementa en dos etapas principales. El primero es encontrar el conjunto de elementos recurrentes de un conjunto de operandos, y el segundo es crear reglas de asociación a partir de estos elementos. La primera etapa es la más costosa desde el punto de vista computacional porque realiza una gran cantidad de lecturas o transferencias de datos, un número que depende del conjunto potencial de productos analizados y la mayoría de los esfuerzos de optimización del algoritmo se dirigen a optimizar el tiempo de ejecución durante esta etapa (Amat, 2018).

En este punto, el algoritmo Apriori utiliza un sistema de búsqueda que, en el proceso, calcula cada grupo de elementos regulares de tamaño k. Esto se usa para encontrar en la siguiente iteración el conjunto de elementos repetidos de tamaño k+1.

2.2.2 Metodología Apriori

El funcionamiento del algoritmo Apriori empieza con la obtención de los llamados "conjuntos de ítems frecuentes", los cuales son aquellos conjuntos cuyos ítems superan un umbral que define un valor mínimo para la medida de soporte. Debido al amplio uso del algoritmo Apriori, desde que se formalizó la inducción de reglas de asociación, la obtención de los conjuntos de ítems frecuentes es una tarea común en dichos algoritmos.

El algoritmo a priori ayuda a relacionar los datos, a partir de reglas de asociación, las cuales son basadas en las variables definidas en el pre-procesamiento. Este algoritmo es muy usado en la minería de datos por su confianza y certeza.

Este algoritmo solo puede buscar reglas entre atributos simbólicos, por lo que requiere que todos los atributos numéricos sean eliminados. El algoritmo de minería de reglas de asociación preferido se utiliza para encontrar conjuntos de elementos frecuentes en una base de datos transaccional. La calidad de las reglas de asociación depende del favor y la confianza (Martínez, 2020).

2.2.3 Implementación del Algoritmo Apriori

En primer lugar, se crea una lista en donde se guarden las transacciones con las variables que se desea analizar, es decir, las posibles asociaciones con las variables requeridas y seleccionadas.

```
transactions = []

for i in range(0,2002):
    transactions.append([str(aso.values[i,j]) for j in range(0,3)])
```

Como segundo lugar, se realiza la función que permitirá realizar el metodo a priori de reglas de asociacion, calcular el respectivo suport, condence y lift.

```python
def apriori(transactions, **kwargs):
    """
    Executes Apriori algorithm and returns a RelationRecord generator.

    Arguments:
        transactions -- A transaction iterable object
                        (eg. [['A', 'B'], ['B', 'C']]).

    Keyword arguments:
        min_support -- The minimum support of relations (float).
        min_confidence -- The minimum confidence of relations (float).
        min_lift -- The minimum lift of relations (float).
        max_length -- The maximum length of the relation (integer).
    """
    # Parse the arguments.
    min_support = kwargs.get('min_support', 0.1)
    min_confidence = kwargs.get('min_confidence', 0.0)
    min_lift = kwargs.get('min_lift', 0.0)
    max_length = kwargs.get('max_length', None)

    # Check arguments.
    if min_support <= 0:
        raise ValueError('minimum support must be > 0')

    # For testing.
    _gen_support_records = kwargs.get(
        '_gen_support_records', gen_support_records)
    _gen_ordered_statistics = kwargs.get(
        '_gen_ordered_statistics', gen_ordered_statistics)
    _filter_ordered_statistics = kwargs.get(
        '_filter_ordered_statistics', filter_ordered_statistics)

    # Calculate supports.
    transaction_manager = TransactionManager.create(transactions)
    support_records = _gen_support_records(
        transaction_manager, min_support, max_length=max_length)
```

Figura 30 Implementación Apriori

```python
    # Calculate ordered stats.
    for support_record in support_records:
        ordered_statistics = list(
            _filter_ordered_statistics(
                _gen_ordered_statistics(transaction_manager, support_record),
                min_confidence=min_confidence,
                min_lift=min_lift,
            )
        )
        if not ordered_statistics:
            continue
        yield RelationRecord(
            support_record.items, support_record.support, ordered_statistics)
```

Posteriormente se aplica el algoritmo a priori en el set de datos limpio.

```
rules = apriori(transactions,
                min_support=0.003,
                min_confidence=0.2,
                min_lift=3,
                min_length=2)
```

Finalmente se realiza una nueva función para presentación de los datos.

```
def inspect(results):
    rh          = [tuple(result[2][0][0]) for result in results]
    lh          = [tuple(result[2][0][1]) for result in results]
    supports    = [result[1] for result in results]
    confidences = [result[2][0][2] for result in results]
    lifts       = [result[2][0][3] for result in results]
    return list(zip(rh, lh, supports, confidences, lifts))
```

2.2.4 Análisis de resultados de Apriori

Análisis preliminar

Los resultados son pocos, pero valiosos para el estudio, se debe tomar en cuenta que al realizar el algoritmo este puede duplicar los resultados cuando acción y consecuencia cambian de lugar.

Soporte

El soporte del algoritmo a priori se realiza mediante la razón del número de transacciones que contienen un cierto elemento de interés entre el total de transacciones posibles dentro de los productos (Amat, 2018).

Confianza

El cálculo de la confianza se realiza mediante la razón entre el soporte de las transacciones conjuntas de un cierto articulo entre el soporte de un cierto producto base de la siguiente manera (Amat, 2018).

$$Confidence(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)}$$

Lift

El lift es una medida que cuantifica la relación entre un conjunto de productos analizados en las transacciones, se calcula mediante el cociente del soporte observado dividido entre el soporte teórico que se obtiene en las transacciones y se lo calcula de la siguiente manera.

$$lift = \frac{Confidence(p \rightarrow q)}{Support(d)}$$

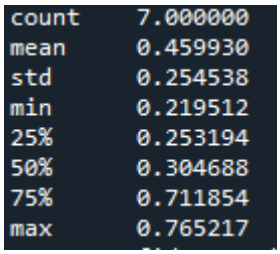
La interpretación para este indicados se lo realiza usando un umbral de 1, mientras que más se acerque al valor 1, menos asociación se obtendrá, mientras que sobrepasa o entre menor sea el valor de 1, la asociación entre categorías será de manera positiva o negativa respectivamente (Martínez, 2020).

2.2.5 Interpretación de algoritmo de reglas de asociación

Para este caso y como análisis a las preferencias de las personas en conjunto de género, categoría y tipo de producto, se obtiene las categorías para genero de masculino (male) y femenino (female), en cuanto a las categorías de los productos estos se dividen en accesorios, casco y bicicletas, para el tipo de producto se tiene las siguientes categorías.

- Paseo
- Discos
- Timones
- Ciclosimuladores
- Montañera
- TT/TRI/TRACK
- Monturas
- Montañera
- Ruta
- Grips
- Zapatillas
- Sport
- Gafas

Como se observa, en la gráfica, el algoritmo de asociación a priori nos sugiere que tomando en consideración el género, la categoría y el tipo de producto adquirido en este centro de venta de bicicletas, en promedio, el promedio de las probabilidades de adquisición de un producto es de alrededor del 46%.



Con respecto a los resultados del support, de manera general, no sobrepasan el 10%, por lo quesu venta no es muy frecuente y en cuanto al confidence, inicialmente se tiene una probabilidad alta en productos de deporte.

En cuanto al lift de los resultados se observa que es superior a 1, por lo que puede decirse que el género, la categoría de productos y el tipo de productos están asociados positivamente.

Tomando en cuenta los casos particulares se puede ver que de una persona que ha comprado un accesorio para hacer deporte tienen una probabilidad del 72.84% de comprar un casco, de las personas que adquirieron productos de TT/TRI/TRACK, la probabilidad de que una persona compre un casco es del 27.16%, de las personas que compraron un artículo para paseo, la probabilidad de que estas compre una bicicleta y sean mujeres es del 76.52%, por otro lado una persona que haya comprado artículos para deporte, la probabilidad de que esta sea hombre y compre un casco es del 69.53%, en el caso de comprar un artículo de TT/TRI/TRACK, la probabilidad de que esta sea mujer y haya comprado un casco es de 23.47% asimismo del mismo grupo de personas que compraron este tipo de productos la probabilidad de que sea hombre y compre un casco es del 30.46%.

Tabla 5 Resultados del método a priori

Product 2	Product 1	Support	Confidence	Lift
('Casco',)	('Sport',)	0,093651	0,728395	7,777778
('Casco',)	('TT/TRI/TRACK',)	0,034921	0,271605	7,777778
('female', 'Bicicletas')	('Paseo',)	0,033333	0,219512	3,096105
('female', 'Casco')	('Sport',)	0,046561	0,765217	8,170965
('male', 'Casco')	('Sport',)	0,04709	0,695313	7,424523
('female', 'Casco')	('TT/TRI/TRACK',)	0,014286	0,234783	6,72332
('male', 'Casco')	('TT/TRI/TRACK',)	0,020635	0,304688	8,725142

De esta manera, dicho de otras palabras, la probabilidad es más altas de compra de productos se dan en artículos de deporte, principalmente en la compra de cascos, en cuanto al género, las mujeres tienden a comprar artículos de seguridad más frecuentemente que los hombres, por lo que es recomendable que los productos de seguridad se ofrezcan mayormente a mujeres que a hombres.

2.2.6 Algoritmo de FP – Growth

```
# Metoto de FP - Growth
frecquent = pfg.find_frequent_patterns(transactions = transactions,
                                       support_threshold = 0.7)

rules = pfg.generate_association_rules(patterns=frecquent,
                                       confidence_threshold=0.7)
for i,j in zip(rules.keys(),rules.values()):
    print(i,j)
```

Figura 31 Algoritmo de FP – Growth

A continuación, se presenta el resultado del algoritmo de Frequent Pattern Growth (FP – Growth) en donde se observa que las recomendaciones que se obtienen con el algoritmo son superiores que usando el algoritmo a priori de las reglas de asociación, su funcionamiento es similar que con el algoritmo a priori, sin embargo, en este caso, el algoritmo se toma las asociaciones en base a condiciones de una manera similar a un árbol de decisión, los resultados del algoritmo se presentan a continuación.

En el resultado se puede observar los posibles artículos que se pueden sugerir de acuerdo a ciertas características de los clientes registrados con una probabilidad análoga al lift en el caso del individuo a priori.

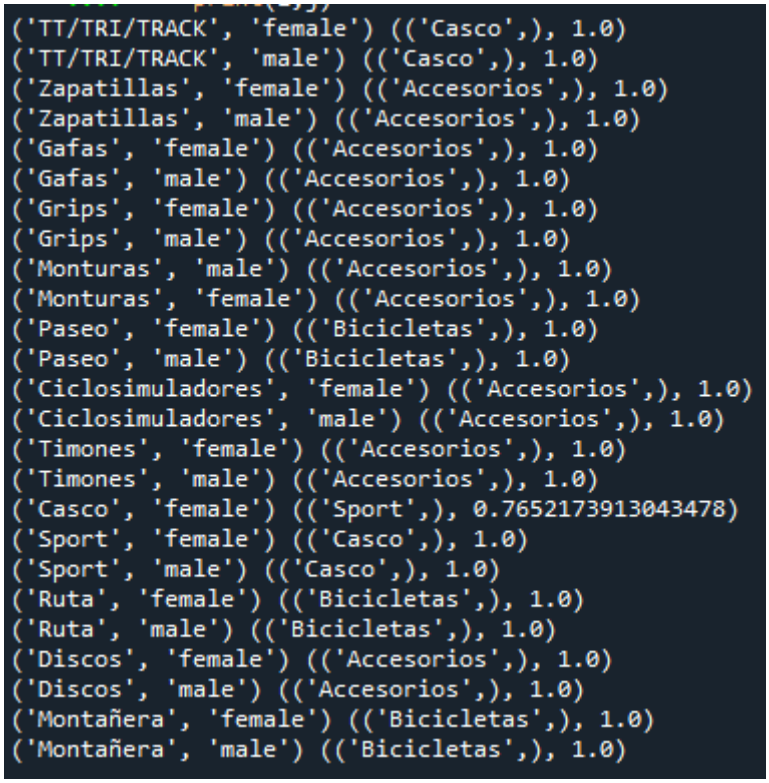


Figura 32 Resultados posibles artículos

2.2.7 Comparación entre el algoritmo a priori y el algoritmo de FP – Growth

Tabla 6 Comparación entre el algoritmo a priori y el algoritmo de FP – Growth

A priori	FP Growth
El método a priori genera grupos de elementos mediante emparejamientos, los cuales pueden ser entre dos, tres, etc elementos.	El algoritmo FP Growth genera patrones en base a patrones frecuentes.
A priori genera los conjuntos a partir de candidatos, en donde se parte de un elemento y se amplía el conjunto según corresponda.	FP – Growth usa un análogo a un árbol de decisión que usa para asociar elementos a cada uno de los items.
El metodo a priori necesita de varios escaneos de la data, lo que consume recursos y tiempo.	FP Grwoth necesita un solo escaneo para determinar las asociaciones mejorando la eficiencia del algoritmo.
Utiliza una búsqueda de patrones no tan profunda.	La búsqueda de patrones se realiza a profundidad

2.2.8 Correlación de datos (Verificación de datos)

Los datos de todas las fuentes no tienen celdas vacías ni inconsistencias sin embargo se debe hacer un análisis de los datos correspondientes a ventas con respecto a los datos demográficos de Tulcán.

Primero graficamos una matriz de correlación entre las variables para ver posibles sesgos de los datos o variables. El resultado fue el siguiente:

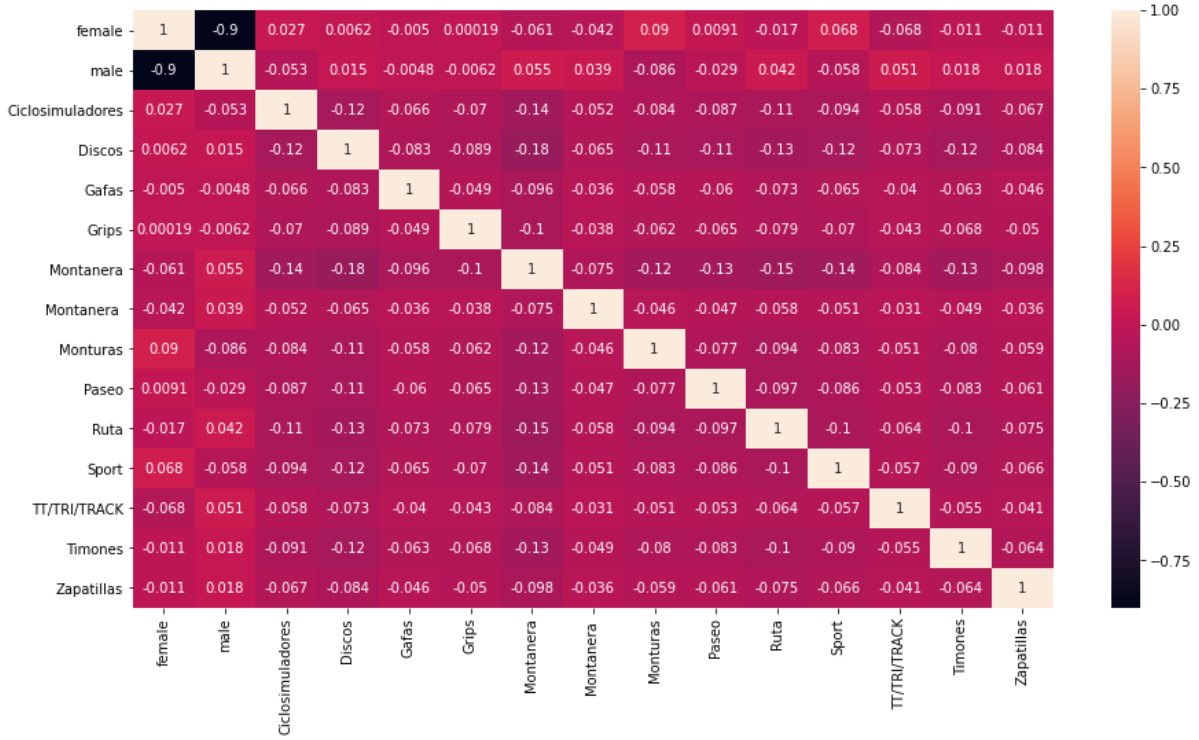


Figura 33 Matriz de Correlación de las variables del set de datos.

Podemos observar en la matriz de correlación que las únicas variables muy ligadas unas con otras corresponden a la marcas y modelos exclusivos unos de otros. Ninguna otra correlación de la gráfica es tan significativa como para considerarse similitud.

2.3 Procesamiento y clústering de datos

2.3.1 Variables seleccionadas

Para efecto y objetivos del estudio las variables seleccionadas para conocer hábitos de compra son las de género, categoría de los productos y el tipo de producto que se adquieren por los clientes.

2.3.2 Algoritmo K-Means o K-Modes

A lo largo del capítulo anterior se observó una serie de datos y el 100% de estos son del tipo categórico, al intentar hacer un gráfico de dispersión de los mismos se puede notar que al ser categórico las distancias tienen mucho sesgo, y al ser el algoritmo K-means un comparador de distancias entre grupos no resulta una buena opción, es por esto que López (2017) creó una extensión de k-Means llamada k-Modes, esta resuelve la clusterización a través de diferencias entre grupos. En esta investigación optaremos por el algoritmo k-Modes (Martínez, 2020).

El algoritmo de disimilaciones de k-modes se ha desarrollado inicialmente para la dimensión vertical con el fin de proporcionar una agrupación más robusta que la agrupación de un solo enlace. El algoritmo de k-modos es básicamente una modificación del algoritmo de k-means auto organizado de Hartigan sin necesidad de reasignación de centros a grupos. Esta adaptación conserva muchas de las ventajas de la técnica de auto organización de Hartigan al tiempo que alivia sus debilidades, como la sensibilidad a las condiciones iniciales, y puede lograr una calidad de agrupación arbitrariamente alta con un número finito de iteraciones.

Costo de K-Modes

El método de K - modes al ser un método iterativo que depende de un cierto número, en este caso, k modas, para poder realizar sus agrupaciones en clúster, es importante establecer un numero de modas optimo para poder realizar un algoritmo adecuado y eficiente, es así que se utiliza el método del codo para analizar el numero óptimo de k, el cual es escogido al momento de determinar un punto de inflexión en el costo del algoritmo, para este caso el valor óptimo de k es de 2, por lo que se encontraran dos clúster con su respectivo centroide.

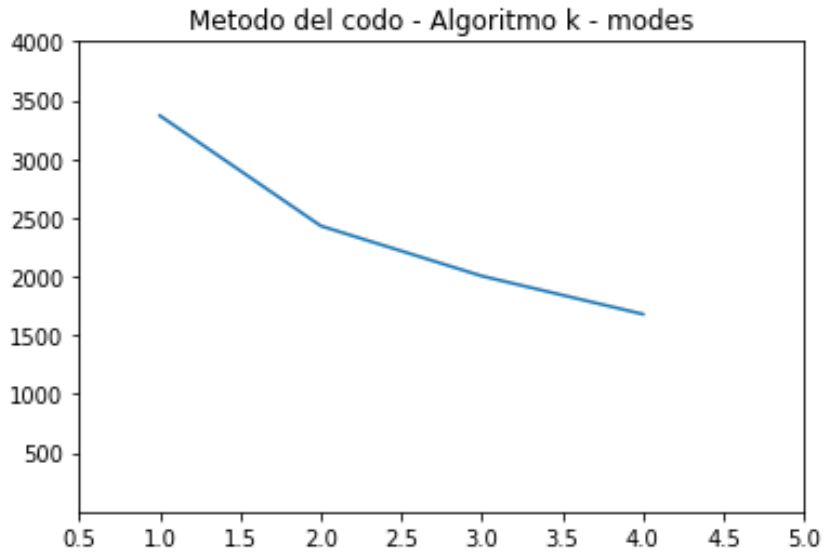


Figura 34 Método del codo algoritmo k-modes

2.3.3 Resultado: Arquetipo de comprador

En el resultado de las preferencias a traves del algoritmo de K – modes, se obtuvo dos centroides correspondientes a cada uno del clúster, el primero formado por género masculino, categoría accesorios y tipo de producto discos, por otro lado, el otro clúster se forma por el

género femenino, la compra de bicicletas de tipo montanera, esto da un claro panorama acerca de cuáles son las preferencias en estos dos grupos de compradores.

Tabla 7 Arquetipo de comprador

Gender	Category	Type
Male	Accesorios	Discos
Female	Bicicletas	Montañera

2.3.4 Algoritmo KNN (vecinos más cercanos)

Al igual que los anteriores algoritmos, un algoritmo útil para poder predecir intención de compra es el algoritmo de vecinos mas cercanos, ya que este algoritmo en base a características comunes intenta predecir la intención de una persona por adquirir un cierto producto, existen ciertas ventajas, a continuación, se presenta el algoritmo para realizar el análisis de la intención de compra de varios artículos individuales de acuerdo a categorías, marcas y tipos de producto.

A continuación, se presenta el algoritmo utilizado para realizar el algoritmo de vecinos más cercanos para cada una de las categorías de tipo, marca y producto.

```
# Dicotomizacion de variables
dgender = pd.get_dummies(data['gender'])
dbrand = pd.get_dummies(data['brand'])
dcategory = pd.get_dummies(data['category'])
dtype = pd.get_dummies(data['type'])
drel = pd.get_dummies(data['relationship_status'])

# Base de datos con variables dicotomizadas
KNN = pd.concat([dgender,dbrand,dcategory,dtype,drel],axis=1)
KNN.head()

KNN.columns


# Definicion de columnas a utilzar en el analisis
x = KNN[['male','female','Divorced', 'Engaged', 'In a civil union',
         'In a relationship', 'It\'s complicated', 'Married', 'Rumit',
         'Separated', 'Single']]

ys = KNN[['Eagle', 'Elite', 'Giant', 'Lazer', 'Pro', 'Shimano',
         'Twitter', 'Accesorios', 'Bicicletas', 'Casco', 'Ciclosimuladores',
         'Discos', 'Gafas', 'Grips', 'Montañera', 'Monturas',
         'Paseo', 'Ruta', 'Sport', 'TT/TRI/TRACK', 'Timones', 'Zapatillas',
         'Divorced', 'Engaged']]



# Importacion de librerias
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25,
                                                    random_state=0)


k = list(range(2,6))
```

```
def Knn(x, ys, k):
    p = []
    presition = []
    errors = []
    K_i = []

    for prod in ys.columns:
        x = x
        y = ys[prod]
        x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25,
                                                            random_state=0)


        sc = StandardScaler()


        x_train = sc.fit_transform(x_train)
        x_test = sc.transform(x_test)

        for K in k:
            classifier = KNeighborsClassifier(n_neighbors = K,
                                              metric = 'minkowski',
                                              p = 2)

            classifier.fit(x_train,y_train)

            y_pred = classifier.predict(x_test)

            cm = confusion_matrix(y_test, y_pred)

            presicion = round((cm[0,0]+cm[1,1])/sum(sum(cm)), 2)
            error = round(1 - presicion, 2)

            p.append(prod)
            K_i.append(K)
            presition.append(presicion)
            errors.append(error)

    df = list(zip(p,K_i,presition,errors))

    DF = pd.DataFrame(df,columns=['Product','K','Presition','Error'])
    return DF
```

Figura 35 Algoritmos de vecinos

```
performance = Knn(x, ys, k)

index = list(range(0,performance.shape[0],4))

performance = performance.sort_values(['Product','Presition'],ascending=[True,False])

bests = performance.loc[index]

bests.sort_values('Presition',ascending=False,inplace=True)
```

En la tabla 7 se puede observar el resultado del proceso de K vecinos más cercanos (KNN), por cuestiones de costo computacional y mejor precisión en las predicciones que realiza el modelo, se ha tomado un número de vecinos para la cauterización de 2, de esta manera poder agrupar a las personas con una intención en su compra de acuerdo con otros individuos que han tenido las mismas características, en este caso tomando en cuenta la relación sentimental y el género.

De esta manera se puede realizar una predicción de la cantidad de datos predichos correctamente, se observa que hay una mayor cantidad de datos predichos correctamente en todas las marcas, tipos y categorías de los productos, mayormente superan el 80% de los

datos predichos correctamente, a excepción de Shimano, bicicletas y accesorios que son inferiores.

Tabla 8 Resultado del procesos K vecinos

i	Product	K	Presition	Error	i	Product	K	Presition	Error
1	Engaged	2	1	0	13	Sport	2	0,9	0,1
2	TT/TRI/TRACK	2	0,96	0,04	14	Eagle	2	0,89	0,11
3	Zapatillas	2	0,95	0,05	15	Casco	2	0,87	0,13
4	Monturas	2	0,95	0,05	16	Lazer	2	0,87	0,13
5	Gafas	2	0,95	0,05	17	Discos	2	0,86	0,14
6	Grips	2	0,94	0,06	18	Montañera	2	0,83	0,17
7	Giant	2	0,93	0,07	19	Twitter	2	0,81	0,19
8	Paseo	2	0,92	0,08	20	Pro	2	0,81	0,19
9	Timones	2	0,91	0,09	21	Shimano	2	0,75	0,25
10	Ruta	2	0,91	0,09	22	Bicicletas	2	0,52	0,48
11	Ciclosimuladores	2	0,91	0,09	23	Accesorios	2	0,5	0,5
12	Elite	2	0,91	0,09					

Tabla 9 Comparación entre los modelos

Modelo Característica	K - modes	K – Nearest Neighbor
Tipo de algoritmo	No supervisado	No supervisado
Forma de clusterizar	Modelo iterativo por categorías	Modelo de asociación por parentesco
Tipo de distribución necesaria	No paramétrica	No paramétrica
Aplicaciones	Descubrir hábitos de compra actuales de los individuos, sin embargo, es un poco limitante en cuanto a un resultado explícito de intensión de compra de un conjunto de bienes.	Principalmente realizar recomendaciones personalizadas de productos a individuos con características similares, pronosticar gustos en productos de nuevos clientes en base a clientes regulares con gustos similares.
Algoritmo	Se comienza seleccionando un k número de clusters que corresponden a	El método de vecinos más cercanos se basa en las características de las observaciones,

	observaciones aleatorias del set de datos, de esta manera se realiza una iteración en cada columna correspondiente a cada valor de cada uno de los clusters, en caso de que la categoría coincida, en este cluster se suma 1, caso contratio un cero, esto se realiza con todos los clusters seleccionados al azar, se crean nuevos clusters con las menores distancias en cada uno de los clusters creados, una vez se obtienen estos clustes se realiza el mismo procedimiento previo hasta que el numero de observaciones por cluster no tenga una variabilidad notable.	se selecciona un número de vecinos a tomar en cuenta en la cauterización, de esta manera considerando el numero de vecinos mas cercanos, el cluster que tenga mas observaciones cerca o parecidas a la observación, será el cluster al que pertenezca este elemento nuevo.
Ventajas	Es un algoritmo que permite conocer tendencias en la combinación de productos, es útil para formar combos o promociones de productos. Es de fácil implementación. La distribución de los datos no debe ser paramétrica. Los datos no necesitas muchos	Permite estimar gustos de nuevos clientes en base al conocimiento previo de clientes antiguos con características similares. La distribución de los datos no debe ser paramétrica.

	tratamientos para implementarlo.	
Desventajas	Es un algoritmo computacionalmente costoso cuando el número de productos es muy amplio por su naturaleza iterativa. No es un resultado explicito (numérico).	El tratamiento de los datos es mayor que en los otros dos modelos. Es más complicado su implementación ya que necesita de datos previos para poder realizar el entrenamiento del algoritmo.

2.4 Comparación teórica

En cuanto al algoritmo de reglas de asociación, se ha demostrado que en caso de un sistema de recomendación y minería de datos como el que se desea implementar tiene un buen resultado, si bien este tipo de técnicas son algo complicadas de evaluar ya que son algoritmos no supervisados, sin embargo como se observa en el estudio realizado por Gonzales (2018) se llega a la conclusión de que el sistema es muy recomendable para cuando se desea conocer acerca de preferencias y posibles productos que se adquirirían por los individuos tomando en cuenta hábitos de compra de individuos con hábitos similares, de la misma manera que en el presente estudio, las reglas de asociación son usadas para generar combinaciones de artículos y en base a reglas probabilísticas bayesianas, determinar la probabilidad de que un artículo sea adquirido en conjunto con otros, a pesar de tratarse de artículos distintos, el método y la manera en la que se lo utiliza es el mismo.

De igual manera en el caso del algoritmo de K – modes aplicado a la segmentación de mercados realizado por Chaturvedi (1996), se llega a la conclusión de que el rendimiento en este algoritmo en ciertos casos no es muy alto, el cálculo análogo que se realizo en este caso para la clasificación de 1540 personas resulta ser ineficiente, dando un total de predicciones correctas de entre 50% y 60%, que para un modelo de predicción se considera bajo, sin embargo el estudio fue realizado incluyendo variables cuantitativas, siendo el método K – modes un algoritmo de aprendizaje automático para variables cualitativas o categóricas.

En el trabajo realizado por Chaturvedi (1996), se aplica el algoritmo de K – modes para realizar una clusterizacion geográfica mediante el cual agrupa se agrupa a individuos con características en común y analizar donde se concentra geográficamente los clientes del negocio.

En el estudio de (Riveros et.al, 2017) se utiliza el algoritmo de K Nearest Neighbors para realizar un sistema de recomendaciones de productos, de igual manera que en el presente estudio, se trata de hallar similitudes en los habitos de compra de clientes antiguos para realizar sugerencias a nuevos clientes que puedan verse interesados en los artículos del negocio, de esta manera dar un mejor servicio y personalizar las opciones que se le pueden presentar al cliente en nuevas visitas.

2.5 Pruebas

2.5.1 Tamaño del mercado

Habiendo encontrado ya el arquetipo de comprador, vamos a comparar los datos demográficos de Tulcán para encontrar el tamaño del mercado y así determinar la viabilidad del proyecto.

El tamaño del mercado relativo a la población total es de 11.4%.

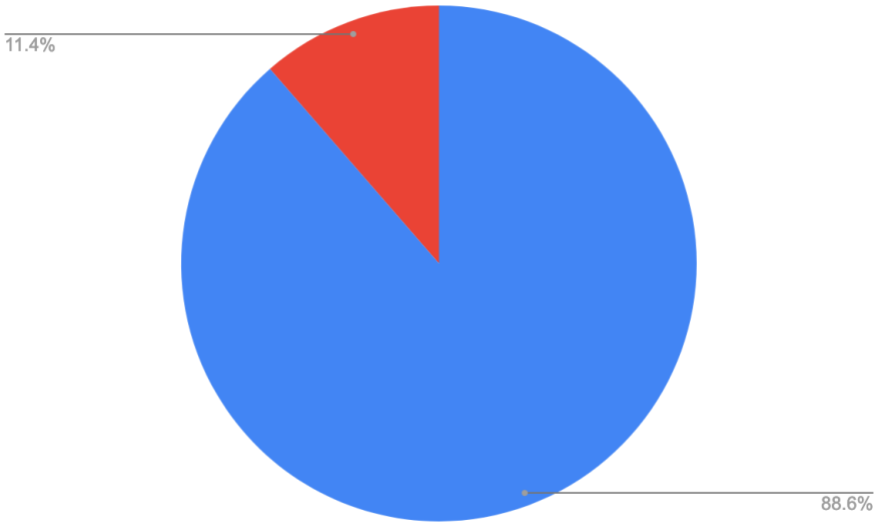


Figura 36 Tamaño del mercado con respecto a la población total.

En América Latina la bicicleta se ha convertido en una forma económica y eficaz de desplazarse en la distancia más corta, encontrar trabajo y reducir la contaminación. Durante años, las bicicletas han sido la principal alternativa de movilidad para trabajar y estudiar en la mayor parte de América Latina. Como consecuencia, las tiendas de bicicletas han comenzado a abrirse vendiendo bicicletas nuevas y usadas, llevando accesorios como cascos, candados o bombas. En estos mercados también se están construyendo las primeras fábricas de bicicletas, así como servicios de mensajería en bicicleta que ofrecen la rapidez como principal ventaja.

2.5.2 Análisis de la competencia

Existe un competidor importante dentro de la ciudad de Tulcán, sin embargo, no existe una tienda de bicicletas especializada para un mercado específico, con asesoría personalizada y complementos de moda.

Puede resultar difícil determinar qué competidor en una ciudad tiene más posibilidades de éxito. Con esto en mente, nos hemos encargado de completar un análisis de cada competidor en términos de demografía para cada mercado en el que están tratando de ingresar.

Los competidores son pocos y no tienen ninguna diferenciación, lo que representa una oportunidad de negocio para nosotros. Además de esto, crearíamos un plan de marketing para dar difusión a la empresa.

2.5.3 Plan de marketing

El plan publicitario fomentará las técnicas comerciales para lograr una situación en el mercado que asegure la resistencia y, posteriormente, el desarrollo y avance de la organización. Para su preparación se considerará el perfil de potencial interés, los costos de los artículos y administraciones de la organización, y cómo hacer que los clientes elijan a nuestra empresa antes que cualquiera de sus rivales inmediatos (Martínez, 2020).

Teniendo en cuenta los diversos animadores comprometidos con el negocio, se propondrán metodologías comerciales para atraer a los clientes esperados. De esta forma se diseccionarán de forma independiente las cuatro "P's" que deben concentrarse en la mejora de cualquier plan impulsor, que son las que acompañan:

- Productos y servicios
- Precio
- Punto de venta o distribución
- Publicidad

Ya se han visto las líneas vitales distintivas de la organización. La compañía intentará seguir estas líneas para lograr una progresión de los objetivos comerciales establecidos en los distintos años bajo investigación (este registro se diseccionará hasta el quinto año). Los destinos se concentrarán en los ejercicios con el mejor efecto en los resultados y serán meticulosos en cuanto a objetivos alcanzables.

En esta línea, se establecerán las metas para el primer año y los siguientes. Se debe considerar que las metas están destinadas a cumplirse, a la luz de la forma en que es una organización de reciente creación y que comenzará en un momento de emergencia monetaria, por lo que serán razonables y tradicionalistas

CAPÍTULO 3

Resultados

3.1 Validación de Resultados

Una vez aplicadas las técnicas de recolección de la información, se procede a realizar el tratamiento correspondiente de los datos para finalizar con el modelo y análisis de los mismos, por cuanto la información que arrojará este modelo será la indique las conclusiones a las cuales llega esta investigación.

En este capítulo presentamos los resultados del modelado de los datos obtenidos en nuestra experimentación. Estos resultados muestran la mejor alternativa de arquetipo de clientes para una tienda de bicicletas y accesorios en Tulcán y las características particulares de cada clúster de clientes. Se destaca especialmente las variables que han influido significativamente en la mejora de la selección de clústeres y en su evolución, ofreciendo las posibles razones que han podido dar lugar a dichos resultados.

3.1.1 Valoraciones Iniciales

En esta sección haremos el análisis para cada categoría del clúster de compradores y analizaremos así el arquetipo al que debe apuntar nuestra empresa de compraventa de bicicletas y accesorios en Tulcán.

3.1.2 Género

La variable de género ha sido tokenizada usando la siguiente correspondencia:

- 0: Hombres
- 1: Mujeres
- 2: Otros

En esta categoría de compradores, los hombres superaron aunque no mucho en cantidad a las mujeres.

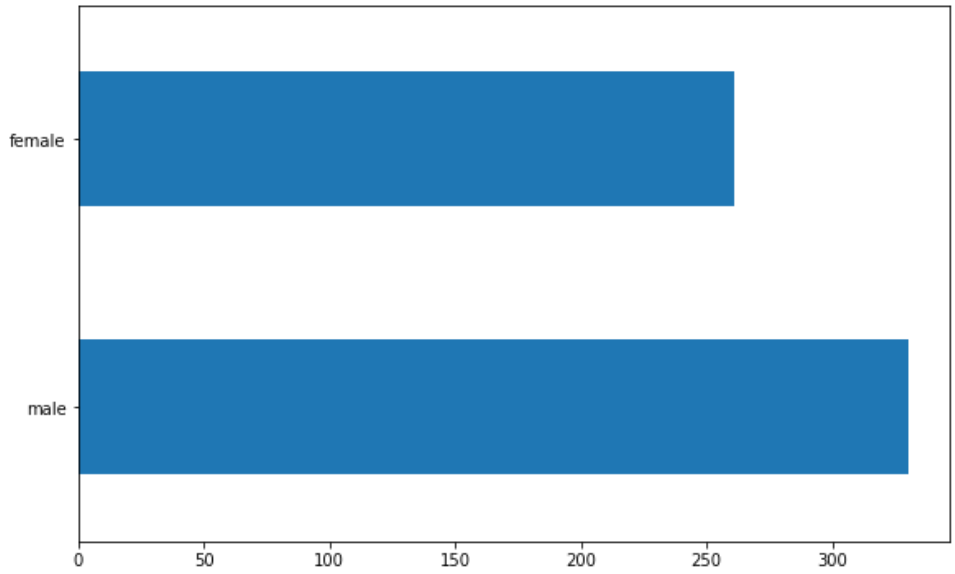


Figura 37 Género

3.1.3 Estado Civil

La variable de estadio civil ha sido tokenizada usando la siguiente correspondencia:

- 0: Solteros
- 1: No especificado
- etc...

Dentro de esta categoría no hay una gran diferencia entre las personas con estado civil, Soltero y Casado, sin embargo, las personas casadas superan con poco a los solteros al comprar bicicletas.

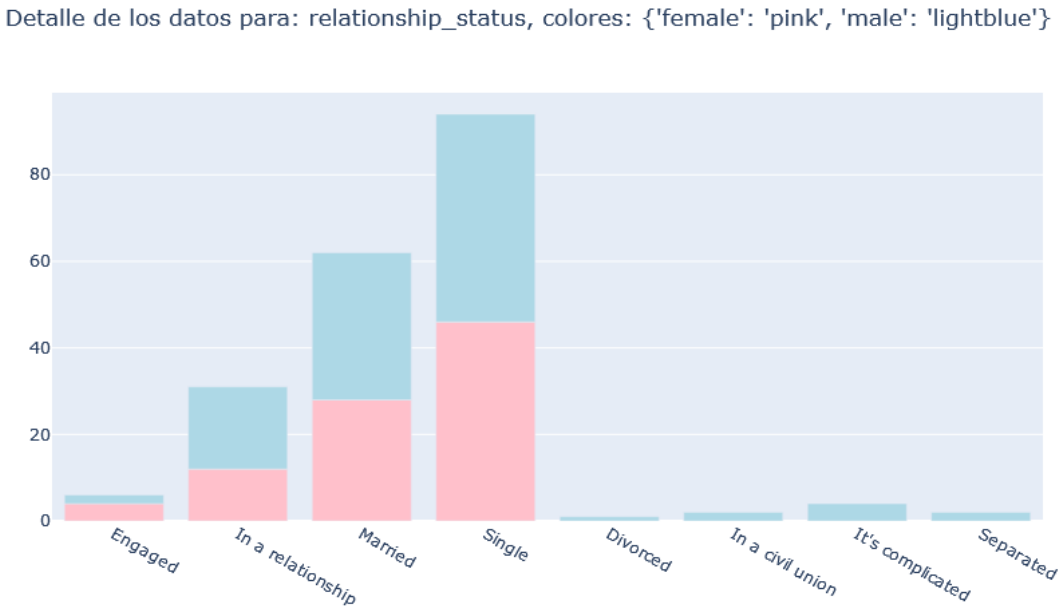


Figura 38 Estado civil de compradores

3.1.4 Cantidad de productos

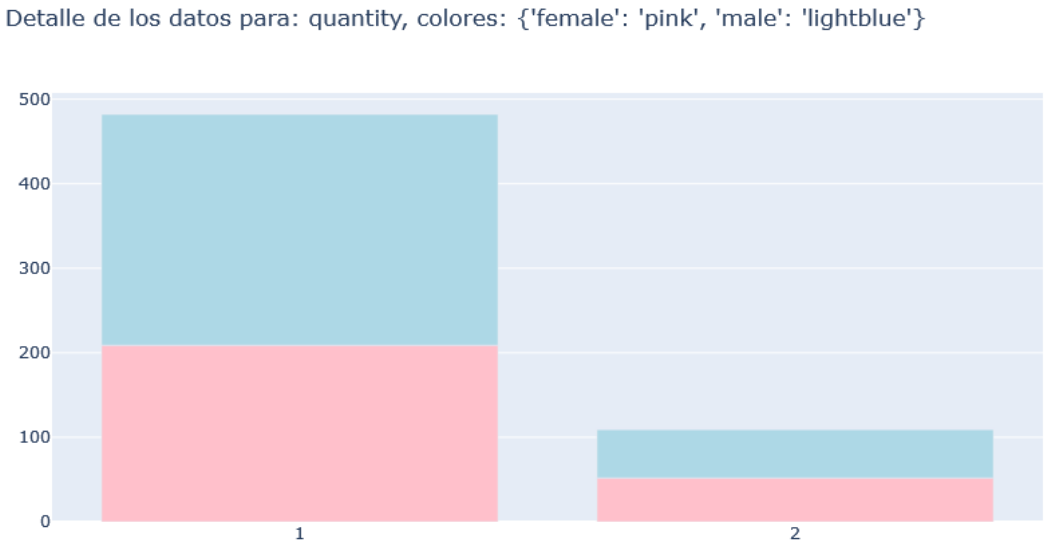


Figura 39 Cantidad de compras

3.1.5 Rangos de precios

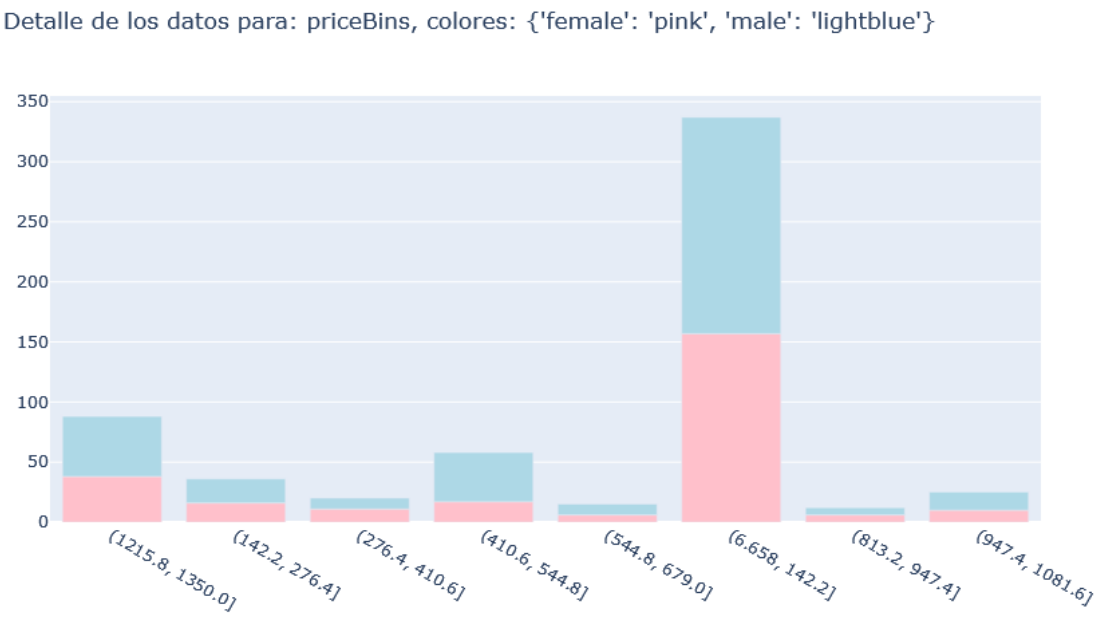


Figura 40 Cantidad de compras por rango de precios

3.1.6 Marcas

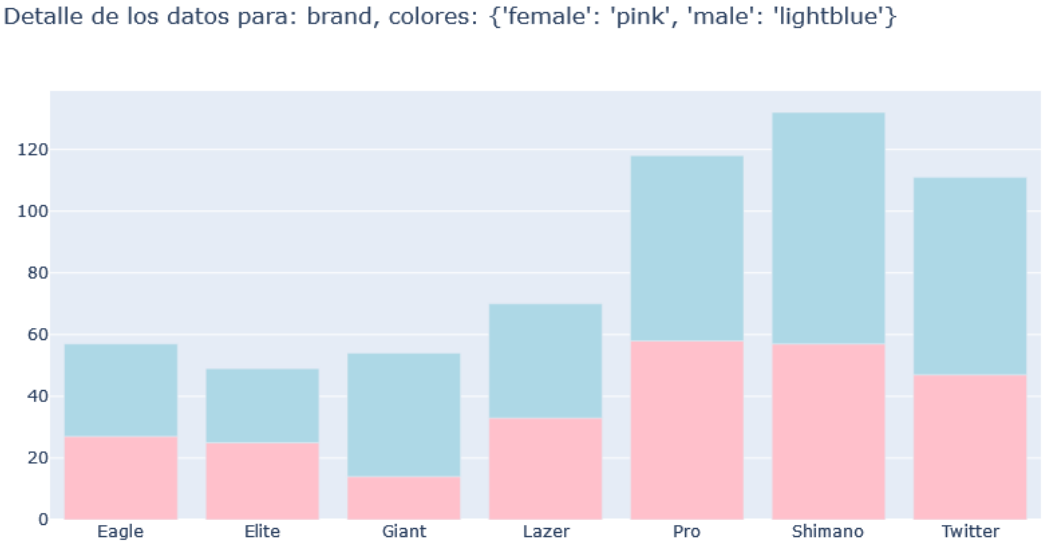


Figura 41 Marcas

3.1.7 Categorías

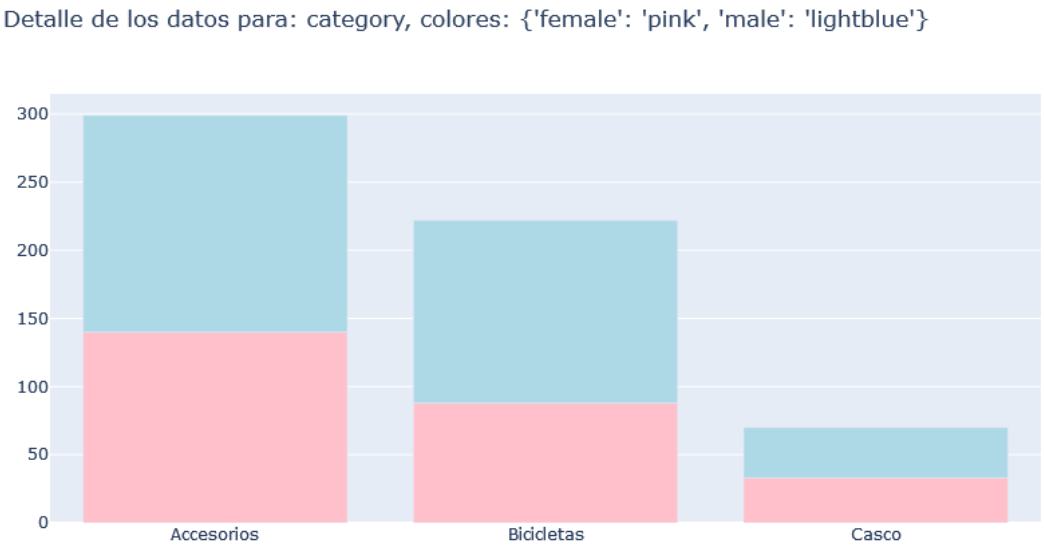


Figura 42 Compras por marca y género

3.1.8 Tipos

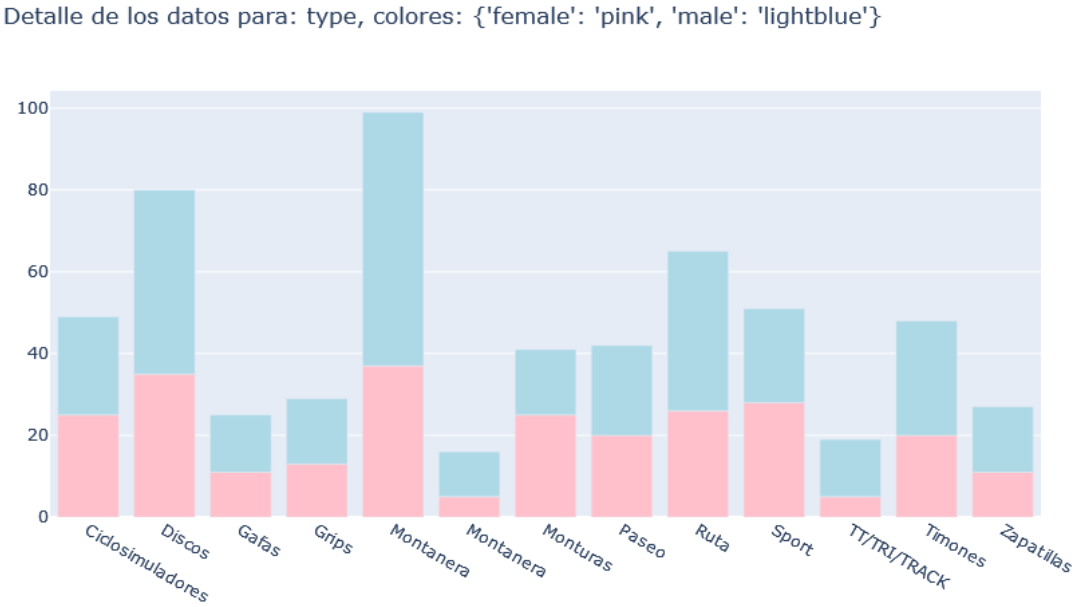


Figura 43 Compras por tipo de producto

3.2 Interpretación de resultados

Sumarizando estos datos encontrados y sus gráficas correspondientes del clúster de resultados podemos obtener ciertas conclusiones:

- La mayor parte de compradores son varones.
- Piden un solo artículo.
- Principalmente compran bicicletas nuevas.
- Principalmente compran bicicletas montañeras y accesorios.
- Principalmente compran la marca Twitter.

Estos resultados se interpretan como que la mayor parte de consumidores son padres, de estrato medio-alto, comprando bicicletas a sus hijos, eso tiene sentido al comparar estos resultados con la frecuencia vs edad mostradas en el capítulo 2. En el que se observa claramente que la mayor parte de la población que usa bicicletas son niños y adolescentes (Rey, Fajardo, & Cubides, 2017).

Esto nos dará una ventaja competitiva con respecto a los demás distribuidores ya que no existe un distribuidor enfocado a ese mercado de niños y adolescentes que ofrezcan productos de alta calidad y seguridad que son las características que buscan los padres al momento de hacer una compra.

3.3 Análisis de impactos

El mercado de bicicletas para niños se divide en dos categorías: tipo y aplicación. Los ciclistas, las partes interesadas y otros participantes del mercado global de Bicicletas para niños obtendrán una ventaja competitiva al utilizar la investigación como un recurso valioso. Para el período 2015-2026, el análisis segmentario se centra en las ventas, los ingresos y las previsiones por tipo y aplicación.

Los impactos externos que puedan afectar el negocio se enumeran en esta sección y se dará una breve introducción a cada uno de ellos.

3.3.1 COVID-19

Market Research Store ha publicado un nuevo estudio sobre la industria mundial de bicicletas para niños con el fin de proporcionar un mejor conocimiento de todo el análisis y valoración del mercado en un solo lugar. La dinámica del mercado se examina a través de las tendencias históricas de crecimiento, las condiciones actuales y las posibilidades de crecimiento futuro en este análisis (Amat, 2018).

No hay un mercado en el planeta que no se haya visto afectado por la pandemia actual. Muchas empresas se han visto obstaculizadas por la pandemia de COVID-19, y el mercado mundial de bicicletas para niños no es una excepción. Para combatir la epidemia, el gobierno y las naciones han implementado una serie de medidas estrictas, incluido el bloqueo y ajustes

a una serie de reglas industriales, para permitir que diversas empresas se mantengan a flote en el mercado.

Mientras el país lucha por abordar la situación de COVID, las ventas minoristas de bicicletas continuarán sólidas en 2021, pero la cadena de suministro no ampliará la disponibilidad de inventario "a pedido" a medida que los precios suban durante el año (Martínez, 2020).

3.3.2 Suministro

El brote de coronavirus del año pasado provocó un auge del ciclismo, que se prolongó hasta 2021, lo que resultó en una escasez en todo el país. Los fabricantes se esfuerzan por satisfacer la demanda de un número cada vez mayor de personas que quieren salir y andar en bicicleta (Amat, 2018).

En comparación con el mismo período en 2019, las ventas de bicicletas aumentaron en un 55 por ciento entre diciembre de 2020 y febrero de 2021, según Matt Powell, asesor senior de la industria y vicepresidente de la firma de investigación de mercado NPD Group. Powell le dijo a CBS News que las ventas de bicicletas son parte de una tendencia mayor de los estadounidenses que compran más equipos para actividades al aire libre este verano (Martínez, 2020).

Los problemas no terminan cuando la bicicleta está terminada y empacada; todavía tiene que enviarse. Este es otro obstáculo más para el sector. El espacio para los contenedores de envío es escaso y caro, y hay retrasos en los puertos.

A pesar de esto, todas las personas con las que he hablado hasta ahora parecen optimistas. Las ventas de equipos están aumentando y los problemas se están solucionando, aunque lentamente. El equilibrio está en camino. Lamentablemente, todavía no ha llegado y, por el momento, puede resultar complicado localizar una bicicleta o una pieza.

Conclusiones

El algoritmo por reglas de asociacion es un buen metodo en el caso de que el interés del negocio sea realizar un sistema de recomendación de productos para los clientes, ademas de que permite ver de mejor manera la probabilidad de que un producto sea adquirido conjuntamente con otros productos.

El algoritmo de K – modes principalmente sirve al negocio para conocer acerca de cuales son las tendencias de compra de los clientes, de esta manera realizar un sistema de 'combos' o promociones conjuntas, ofertando productos conjuntamente.

En el caso de que el negocio tenga un tiempo en funcionamiento y tenga una base de datos más amplia, el algoritmo de K – nearest neighbors se lo puede implementar para conocer hábitos de compra entre individuos similares, de esta manera realizar recomendaciones de productos a clientes nuevos tomando en cuenta la información previa registrada de clientes antiguos con las mismas características sociales, económicas, etc.

En base a los resultados del algoritmo de aprendizaje de reglas de asociación, las probabilidades mas altas en la adquisición de productos, es la que se encuentra en la categoría de deporte, alrededor del 70% de las personas que han adquirido productos para deporte han adquirido principalmente artículos de protección, específicamente cascos, de este grupo de personas, las que más adquieren este tipo de estos productos son las mujeres con alrededor del 76%, mientras que los hombres lo hacen en menor medida con alrededor del 69% de la intención de compra.

En relación con los resultados obtenidos con el algoritmo de K – modes se observo que los hombres tienen un mayor consumo en accesorios de tipo disco, mientras que, en el caso de las mujeres, principalmente adquieren bicicletas montaneras.

En cuanto al algoritmo de K – Nearest Neighbors se observa que se la capacidad predictiva que tiene en cuanto a los modelos es alta ya que supera el 80%, dependerá las necesidades del negocio establecer un umbral más alto para tomar la estimación puntual de la probabilidad de una predicción considerada como adecuada, a excepción de shimano, bicicletas y accesorios, para los demás productos el algoritmo funciona me marea apropiada, de esto se puede decir que se puede realizar una predicción aceptable en el caso de la llegada de nuevos clientes.

Recomendaciones

En la actualidad el conocimiento sobre el mercado y los hábitos de los clientes brinda una gran ventaja en la competitividad de los negocios, principalmente cuando la competencia se encuentra en constante cambio y evolución, es así como la implementación de un algoritmo que permita realizar automáticamente un sondeo de los gustos y preferencias de consumidores es fundamental para ganar un espacio dentro del mercado y las técnicas de aprendizaje automático son la mejor alternativa como solución a muchas de estas necesitades en los negocios.

Es así como con lo antes mencionado el manejo, optimización e implementación continua de algoritmos de aprendizaje automático son una gran apuesta, en el presente caso de la tienda de bicicletas para el caso de realizar un sistema de recomendación de productos basado en compras anteriores de los clientes mediante el algoritmo de reglas de asociación, asi mismo dar seguimiento a productos nuevos, tomar en cuenta características de nuevos productos a ofertar para poder relacionarlos con productos ya en existencias, de esta manera poder ofrecer una mayor variedad de productos a los clientes.

En caso de que se desee conocer tendencias para poder ofrecer paquetes de productos o realizar ventas en conjunto con otros productos, el algoritmo de K – Modes es una opción muy viable ya que permitirá encontrar tendencias en la compra de productos.

En cuanto al cliente, si se cuenta con un volumen más amplio de información, se puede realizar un sistema para conocer la aceptación de nuevos clientes a los productos que se oferten actualmente, el algoritmo de K – Nearest Neighbors se recomienda para predecir posibles intenciones de compra de nuevo mercado que acuda al negocio de bicicletas.

De manera general se recomienda utilizar los algoritmos de aprendizaje automático de acuerdo a las necesitades del negocio, reglas de asociación para un sistema de recomendaciones, K – Modes para analizar tendencias de compra de conjuntos de productos y K – Nearest Neighbors para predecir hábitos de compra entre individuos con características similares.

Bibliografía

Hernández, Fernández, & Baptista. (2017). *Metodología de la Investigación Científica.* México: McGrawHill.

Delos Arcos, J. R., & Hernandez, A. A. (2019). Efficient Apriori algorithm using enhanced transaction reduction approach. *2019 IEEE 13th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*. https://doi.org/10.1109/tssa48701.2019.8985482

Gaikwad, P. R., Kamble, S. D., Thakur, N. V., & Patharkar, A. S. (2017). Evaluation of Apriori algorithm on retail market transactional database to get frequent Itemsets. *Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering*. https://doi.org/10.15439/2017r83

Intellipaat. (2022, February 18). *Data science apriori algorithm in Python - Market Basket Analysis*. Intellipaat Blog. Retrieved March 10, 2022, from https://intellipaat.com/blog/data-science-apriori-algorithm/

Programador CLIC. Algoritmo a priori - programador clic. (n.d.). Retrieved March 10, 2022, from https://programmerclick.com/article/95571908197/

Sutisnawati, Y., & Reski, M. (2019). Looking for transaction data pattern using APRIORI algorithm with Association Rule Method. *IOP Conference Series: Materials Science and Engineering*, *662*(2), 022078. https://doi.org/10.1088/1757-899x/662/2/022078

Hernández, J., Ramírez, M., & Ferri, C. (2004). *Introducción a la Minería de datos.* España: Pearson Educación.

INEC. (2010). *Fascículo Provincial Carchi.*

INEC. (16 de Abril de 2017). *Ecuador en Cifras*. Obtenido de https://www.ecuadorencifras.gob.ec/24-millones-de-personas-usaron-la-bicicleta-en-2016/

INEC. (2021). *INEC.*

López, A. S. (2017). Reglas de asociación en una Base de datos . *Redalyc.com*, 16-20.

López, S. (2017). Algoritmos de Agrupamiento. *Instituto Nacional de Astrof´ısica, Optica y Electronica*, 58.

MINTEL. (12 de febrero de 2018). *Comercio electrónico, una oportunidad para el desarrollo de los negocios a través de la web*. Obtenido de Ministerio de Telecomunicaciones y de la Sociedad de la Información de la República del Ecuador: https://www.telecomunicaciones.gob.ec/comercio-electronico-una-oportunidad-para-el-desarrollo-de-negocios-a-traves-de-la-

web/#:~:text=%C2%BFPor%20qu%C3%A9%20es%20importante%20su,dinamiza%20y%20diversifica%20la%20econom%C3%ADa.

Molina, L. (10 de noviembre de 2012). *Data Mining: Torturando a los datos hasta que confiesen.* Obtenido de Universidad Oberta de Cataluña: https://www.businessintelligence.info/resources/assets/dss/molina-torturando-datos.pdf

Monrroy, M. (2016). El concepto clúster. *CLACSO-CONACYT*, 56-89.

Rodríguez, I. (2014). *Marketing Digital y Comercio Electrónico.* Catalunya- España: Ediciones Pirámide.

Rodríguez, J. (28 de Febrero de 2019). *Blau Comunicación y Eventos*. Obtenido de https://www.blaucomunicacion.es/el-blog-de-blau/produccion_audiovisual/video-interactivo/#Tipos_de_videos_interactivos

Roldán, N. (2015). ¿PARA QUÉ QUEREMOS CLÚSTERS? *Redalyc.com*, 45-56.

Seoane, E. (2005). La nueva era del Comercio Electrónico: TIC al servicio de la gestión empresarial. *Dialnet*, 165.

Subra, J., & Vannieuwenhuyze, A. (2018). *Scrum: Un método ágil para sus proyectos.* Barcelona: Ediciones ENI.

Torre, G., & Codner, D. (2017). *Fundamentos del Comercio Electrónico. 2da Edición.* Buenos Aires: Universidad Virtual de Quilmes.

Tulcán, M. (2021). *Proyecto ciclovía emergente para la ciudad de Tulcán.*