**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

# Cesión de derechos de autora

# Certificación del trabajo de grado

# Dedicatoria

# Agradecimientos

**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.

**Summary**

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.

Índice de Contenido

[Cesión de derechos de autora II](#_Toc99032950)

[Certificación del trabajo de grado III](#_Toc99032951)

[Dedicatoria IV](#_Toc99032952)

[Agradecimientos V](#_Toc99032953)

[INTRODUCCION 1](#_Toc99032954)

[PROBLEMA 1](#_Toc99032955)

[ANTECEDENTES 1](#_Toc99032956)

[SITUACIÓN ACTUAL 1](#_Toc99032957)

[PLANTEAMIENTO DEL PROBLEMA 1](#_Toc99032958)

[OBJETIVOS 2](#_Toc99032959)

[OBJETIVO GENERAL 2](#_Toc99032960)

[OBJETIVOS ESPECÍFICOS 2](#_Toc99032961)

[ALCANCE 2](#_Toc99032962)

[METODOLOGÍA 4](#_Toc99032963)

[JUSTIFICACIÓN 4](#_Toc99032964)

[CAPITULO I 6](#_Toc99032965)

[Marco Teórico 6](#_Toc99032966)

[1.1. Comercio Electrónico 6](#_Toc99032967)

[1.1.1. Evolución del Comercio Electrónico 6](#_Toc99032968)

[1.1.2. Tipos del Comercio Electrónico 9](#_Toc99032969)

[1.1.3. Características del Comercio Electrónico 11](#_Toc99032970)

[1.1.4. Importancia del Comercio Electrónico 12](#_Toc99032971)

[1.1.5. Ventajas y desventajas del Comercio Electrónico 12](#_Toc99032972)

[1.2. Data Mining 14](#_Toc99032973)

[1.2.1. Arquitectura de datos 15](#_Toc99032974)

[1.2.2. Etapas de Data Mining 16](#_Toc99032975)

[1.2.3. Extracción del conocimiento en Data Mining 17](#_Toc99032976)

[1.3. Técnicas de Data Mining 17](#_Toc99032977)

[1.3.1. Reglas de asociación 20](#_Toc99032978)

[1.3.2. Clustering 20](#_Toc99032979)

[1.4. Scrum 21](#_Toc99032980)

[1.4.1. Cuando se utiliza 22](#_Toc99032981)

[1.4.2. Proceso del Scrum 23](#_Toc99032982)

[1.4.3. Roles del Scrum 23](#_Toc99032983)

[1.4.4. Fases del Scrum 24](#_Toc99032984)

[1.4.5. Beneficios del Scrum 25](#_Toc99032985)

[CAPITULO II 27](#_Toc99032986)

[Desarrollo 27](#_Toc99032987)

[2.1. Recolección de datos 27](#_Toc99032988)

[2.1.1. Metodología AGILE SCRUM para el prototipo 27](#_Toc99032989)

[2.1.2. Limpieza de datos 28](#_Toc99032990)

[2.1.3. Datos demográficos de Tulcán 30](#_Toc99032991)

[2.1.4. Ciclismo en Tulcán 33](#_Toc99032992)

[2.1.5. Ventas de bicicletas 33](#_Toc99032993)

[2.1.6. Estados Financieros de la venta de bicicletas 2016-2020 35](#_Toc99032994)

[2.1.7. Visualización del prototipo 37](#_Toc99032995)

[2.2. Reglas de asociación 39](#_Toc99032996)

[2.2.3. Variables a predecir 43](#_Toc99032997)

[2.3. Procesamiento y clústering de datos usando Python 43](#_Toc99032998)

[2.3.1. K-Means o K-Modes 43](#_Toc99032999)

[2.3.2. Costo de K-Modes 44](#_Toc99033000)

[2.3.3. Número de clústeres 45](#_Toc99033001)

[2.3.4. Resultado: Arquetipo de comprador 46](#_Toc99033002)

[2.4. Pruebas 46](#_Toc99033003)

[2.4.1. Tamaño del mercado 46](#_Toc99033004)

[2.4.2. Análisis de la competencia 47](#_Toc99033005)

[2.4.3. Plan de marketing 47](#_Toc99033006)

[CAPITULO III 49](#_Toc99033007)

[Resultados 49](#_Toc99033008)

[3.1. Validación de resultados 49](#_Toc99033009)

[3.1.1. Valoraciones Iniciales 49](#_Toc99033010)

[3.1.1.1. Género 49](#_Toc99033011)

[3.1.1.2. Estado Civil 50](#_Toc99033012)

[3.1.2.3. Cantidad de productos 51](#_Toc99033013)

[3.1.2.3. Rangos de precios 51](#_Toc99033014)

[3.1.2.3. Marcas 52](#_Toc99033015)

[52](#_Toc99033016)

[3.1.2.3. Categorías 53](#_Toc99033017)

[3.1.2.3. Tipos 53](#_Toc99033018)

[3.2. Interpretación de resultados 54](#_Toc99033019)

[3.3. Análisis de impactos 54](#_Toc99033020)

[3.3.1. COVID-19 54](#_Toc99033021)

[3.3.2. Suministro 55](#_Toc99033022)

[Conclusiones y Recomendaciones 56](#_Toc99033023)

[Conclusiones 56](#_Toc99033024)

[Recomendaciones 56](#_Toc99033025)

[Bibliografía 57](#_Toc99033026)

**Índice de Figuras**

[**Figura 1**](file:///C:\lenovo\Desktop\MarcoTeorico-BrayanP%25C3%25A9rez-2.docx#_Toc94824865) Diagrama de Ishikawa 2

[**Figura 2**](file:///C:\lenovo\Desktop\MarcoTeorico-BrayanP%25C3%25A9rez-2.docx#_Toc94824866) Metodología 4

[**Figura 3** Diferencia entre comercio tradicional y el electrónico 8](#_Toc94824867)

[**Figura 4** Diferencia entre mercado virtual y físico 9](#_Toc94824868)

[**Figura 5** Relación entre dato, información y conocimiento 14](#_Toc94824869)

[**Figura 6** Etapas de Data Mining 16](#_Toc94824870)

[**Figura 7** Proceso de Extracción de conocimiento en Data Mining 17](#_Toc94824871)

[**Figura 8** Técnicas de Data Mining 18](#_Toc94824872)

[**Figura 9** Metodología Scrum 22](#_Toc94824873)

[**Figura 10** Procedo de la metodología Scrum 23](#_Toc94824874)

[**Figura 11** Sprint para el desarrollo del prototipo. 27](#_Toc94824875)

[**Figura 12** Población por ocupación 29](#_Toc94824876)

[**Figura 13** Población por estado civil 30](#_Toc94824877)

[**Figura 14** Distribución de ciclistas por recorrido 31](#_Toc94824878)

[**Figura 15** Compras en el almacén por categoría 32](#_Toc94824879)

[**Figura 16** Compras en el almacén por rango de precios 32](#_Toc94824880)

[**Figura 17** Compras en el almacén por marcas 33](#_Toc94824881)

[**Figura 18** Ingresos por venta de bicicletas 2016-2020 34](#_Toc94824882)

[**Figura 19** Flujo de efectivo 2016-2020 34](#_Toc94824883)

[**Figura 20** Utilidad 2016-2020 34](#_Toc94824884)

[**Figura 21** Gráfica de porcentajes para cada una de las columnas. 36](#_Toc94824885)

[**Figura 22** Ejemplo de grafica de Detalle/Diagnostico. 37](#_Toc94824886)

[**Figura 23** Matriz de Correlación de las variables del set de datos. 40](#_Toc94824887)

[**Figura 24** Inicialización "Cao" 42](#_Toc94824888)

[**Figura 25** Inicialización "Huang" 42](#_Toc94824889)

[**Figura 26** Tamaño del mercado con respecto a la población total. 44](#_Toc94824890)

[**Figura 27** Género de compradores 47](#_Toc94824891)

[**Figura 29** Estado civil de compradores 48](#_Toc94824892)

[**Figura 28**](file:///C:\lenovo\Desktop\MarcoTeorico-BrayanP%25C3%25A9rez-2.docx#_Toc94824893) Estado civil de los compradores 48

[**Figura 30** Cantidad de compras 48](#_Toc94824894)

[**Figura 31** Cantidad de compras por rango de precios 49](#_Toc94824895)

[**Figura 32** Compras por marca y genero 49](#_Toc94824896)

[**Figura 33** Compras por marca y genero 50](#_Toc94824897)

[**Figura 34** Compras por tipo de producto 50](#_Toc94824898)

**Índice de Tablas**

**Tabla 1** Ventajas del Comercio Electrónico 12

**Tabla 2** Desventajas del Comercio Electrónico 13

**Tabla 3** Técnicas predictiva 19

**Tabla 4** Técnicas descriptiva 19

**Tabla 5** Resultados de la base de datos 38

**Tabla 6** Moda de cada uno de los clústeres 43

# 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.

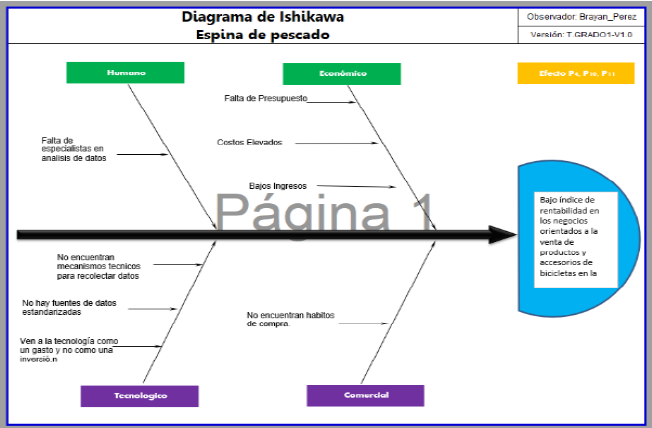
**Talento Humano.** – No encuentran especialistas que puedan hacer análisis de Datos

**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

* Realizar el marco teórico sobre el uso y aplicación de Reglas de Asociación y Clustering en Data Mining con Comercio Electrónico.
* Construir un Data Mining aplicando reglas de asociación y clustering para descubrir hábitos de compra de productos y accesorios de bicicletas.
* 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 de servicios Rest, Python y Qlik Sense, 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.

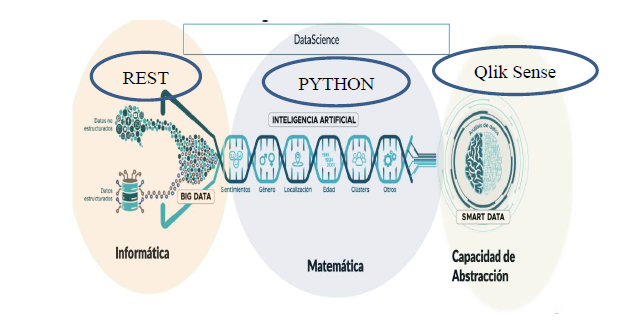
* Con el manejo de servicios Rest se realizan las peticiones de los datos a ser analizados los mismos que serán obtenidos de un dataset de la empresa GOLDEN.
* Posteriormente se realiza el procesamiento de texto utilizando Python.
* Por último, mostraremos los resultados obtenidos en el prototipo desarrollado.

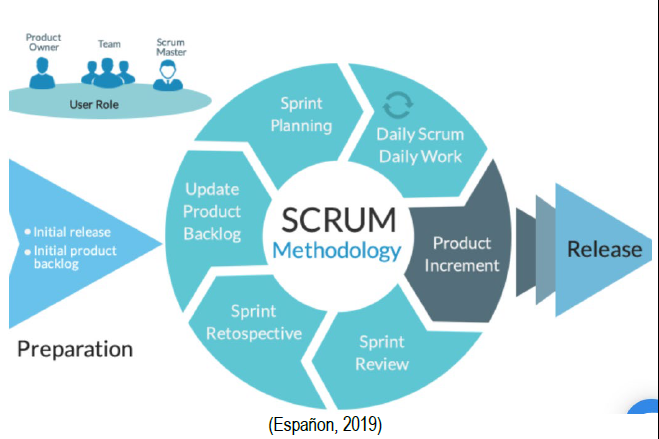
Las tecnologías que se usarán para el análisis de datos aplicando la técnica Reglas de asociación en Data Mining serán:

* Manejo de servicios REST.
* Lenguaje de programación Python.
* Software Qlik Sense.
* 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

## 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 en Qlik Sense.

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.

# CAPITULO I

# 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).

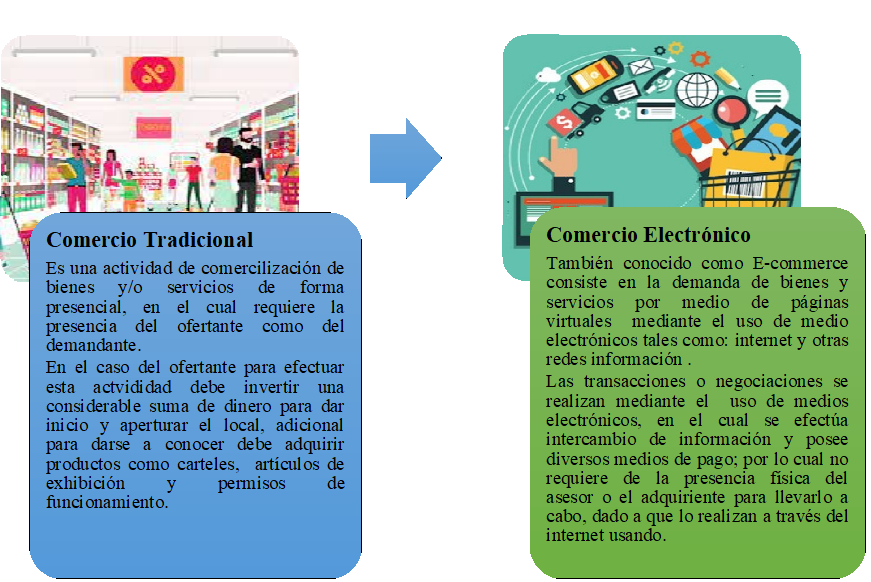
Por ello se puede manifestar que el comercio electrónico ha transformado el mercado, las funciones tradicionales de la intermediación, dado a que ha sido reemplazada y han aparecido nuevos productos y servicios en el mismo. De esta forma se ha modificado la organización e­n el trabajo, aumentando la 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).

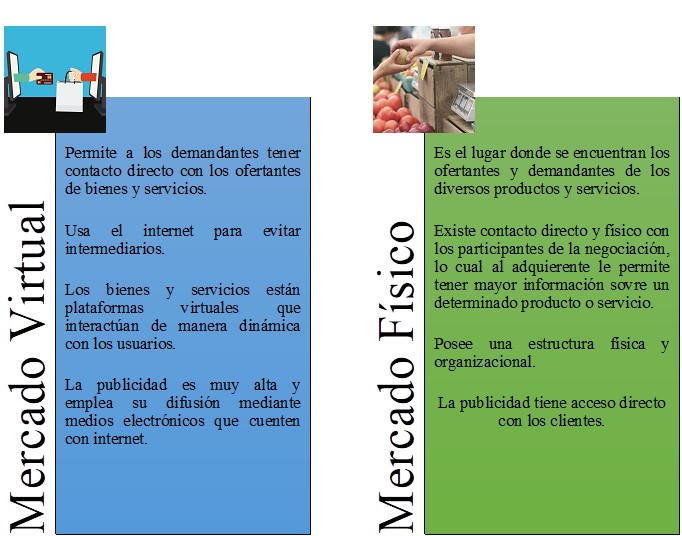
Se puede manifestar que el comercio electrónico se ha encargado de cambiar la ideología del comercio tradicional por uno método innovador, en el cual los bienes, productos y/o servicios son auspiciados en un portal web o aplicaciones móviles de forma dinámica y atractiva para el usuario, además que facilitan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

1. **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).

1. **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.

1. **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).

1. **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.

1. **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).

1. **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.

1. **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 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. | Permite encontrar un portafolio dinámico 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

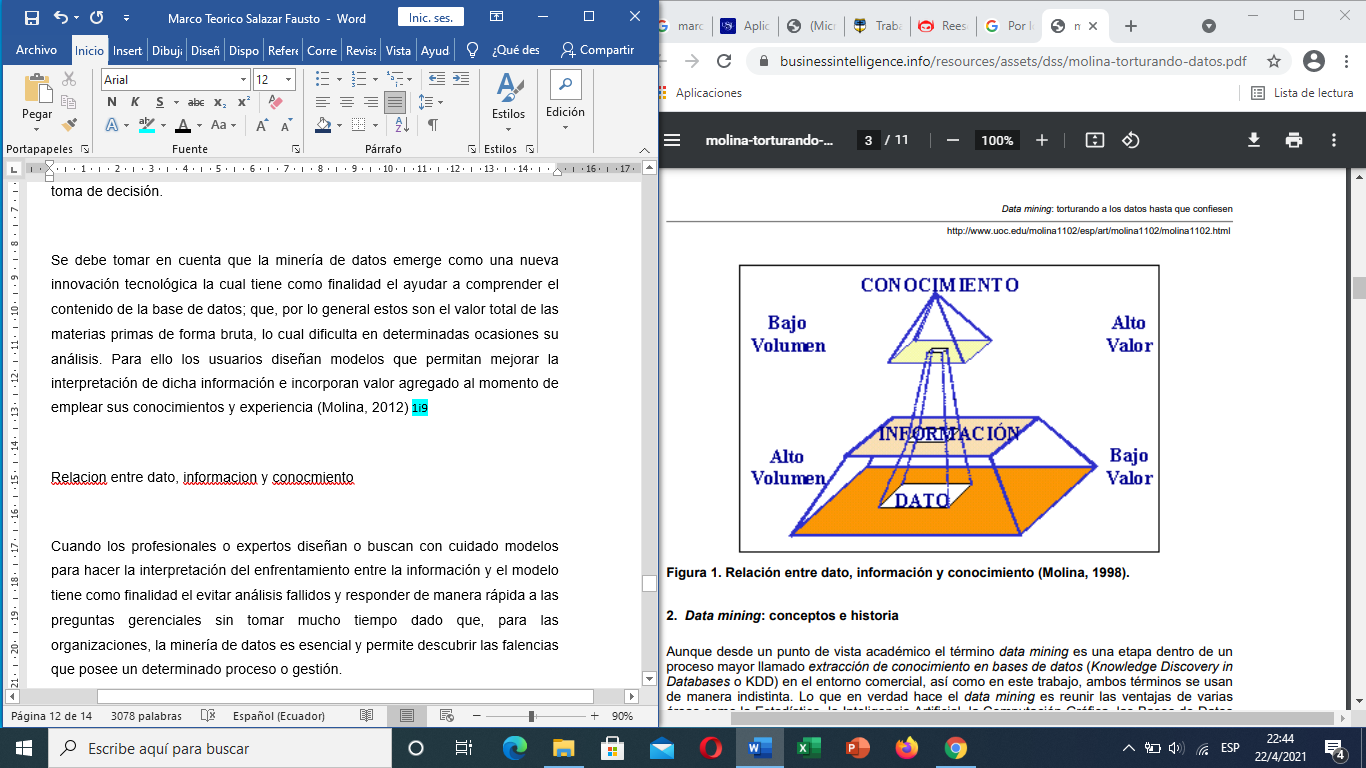
| **EMPRESA (OFERTANTE)** | **COMPRADOR (DEMANDANTE)** |
| --- | --- |
| Requiere de acuerdo comerciales internacionales para armonizar el proceso de comercio electrónico en la nación.  Periodos cortos de validez de contratos y negociaciones sin papeles  Mayor control de las transacciones internacionales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

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

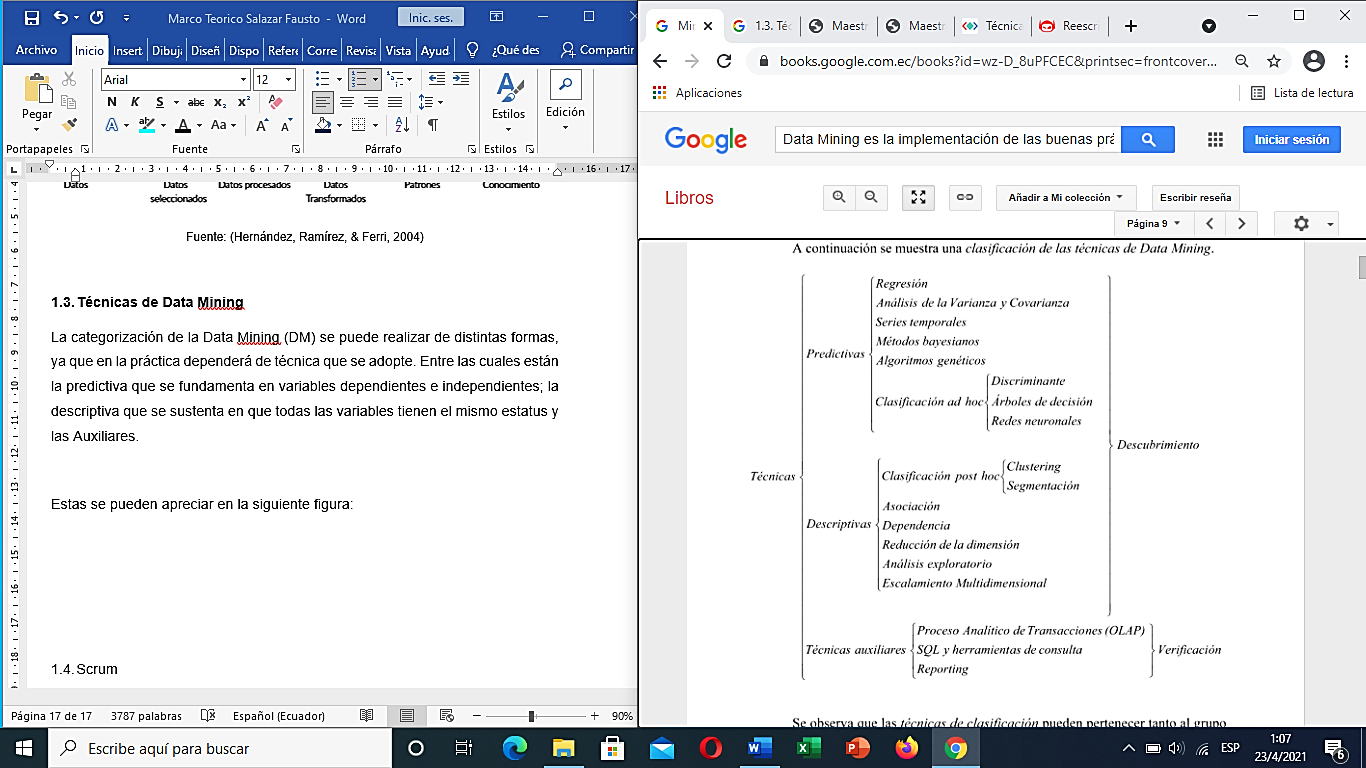
Fuente: Hernández, Ramírez y Ferri (2004)

## 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 dimensión** | Es un requisito importante para un sistema de aprendizaje, 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).

### 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).

**Métodos de Clustering**

* **Algoritmo de k-medias**

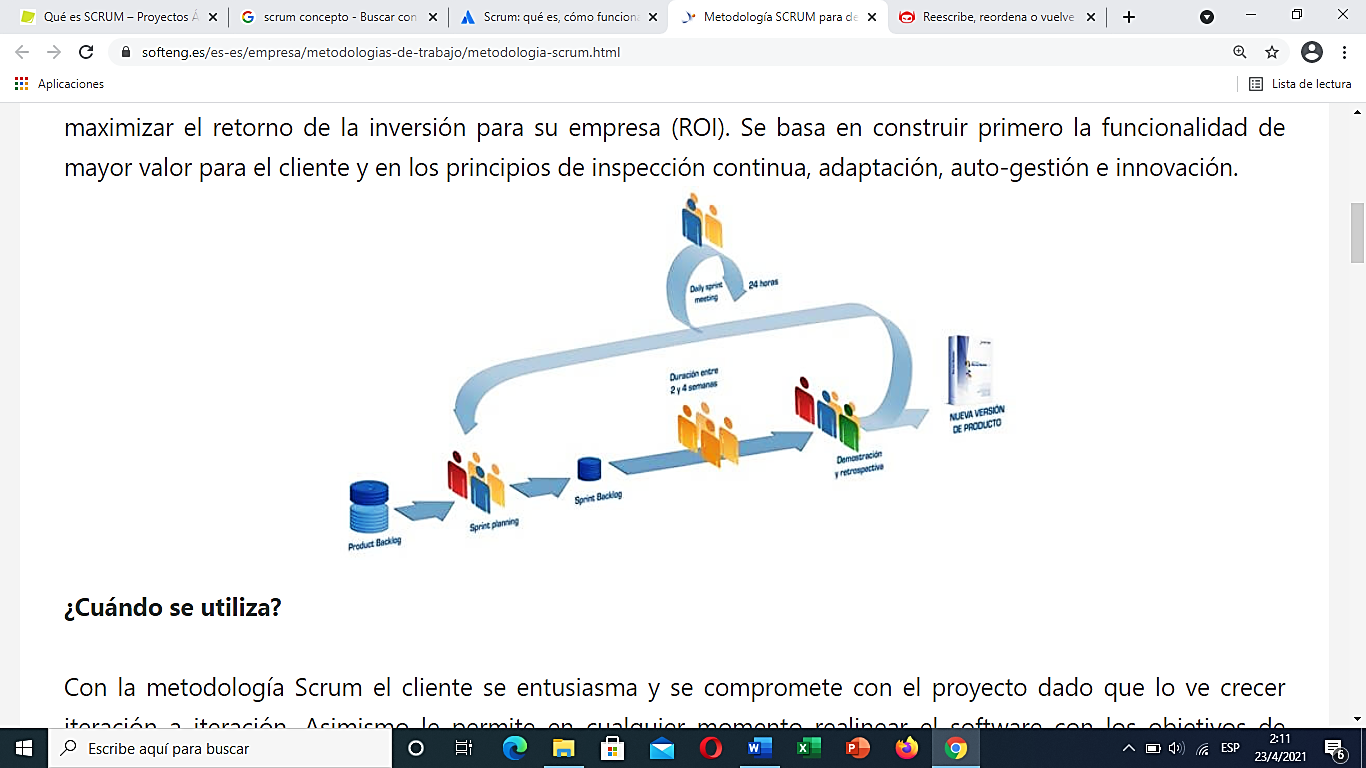
Es un método clásico en el que se aplica un proceso de agrupamiento, en este caso el algoritmo busca los mejores centroides para realizar la correlación de manera que los mejores centroides estén lo más cerca (Monrroy, 2016).

* **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).

## 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 9** 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).

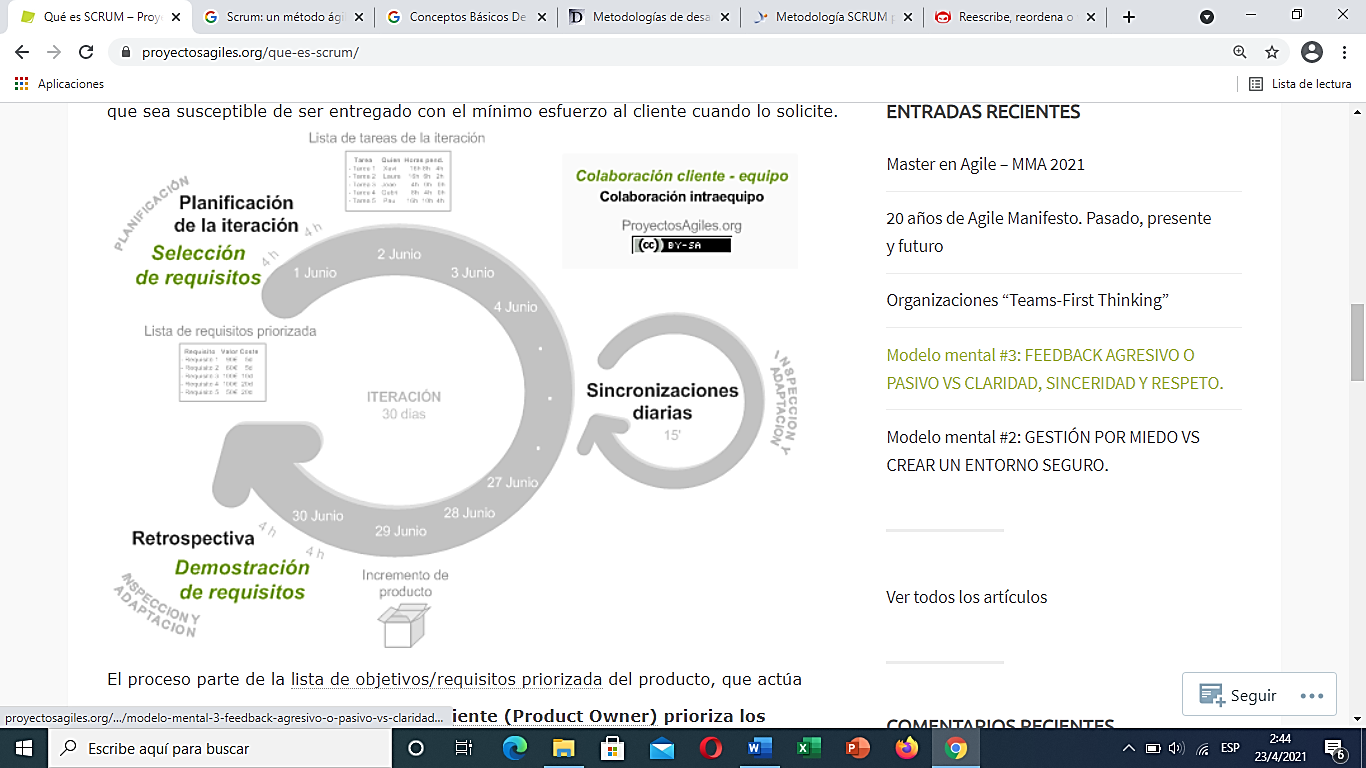
En otras palabras, se puede decir expresar que el Scrum es un instrumento que gestiona el desarrollo de in software cuya finalidad es maximizar el retorno de inversión de la entidad, además de acrecentar la productividad y competitividad de la organización.

### 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 10** 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?

1. **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.

1. **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.

1. **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.

1. **Lanzamiento**

Volverá a los resultados del trabajo y la entrega del producto. Aquí debe realizar 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).

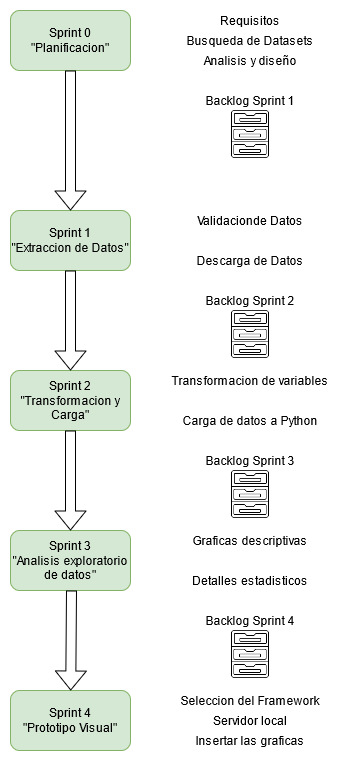
# CAPITULO II

# 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 11** 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.

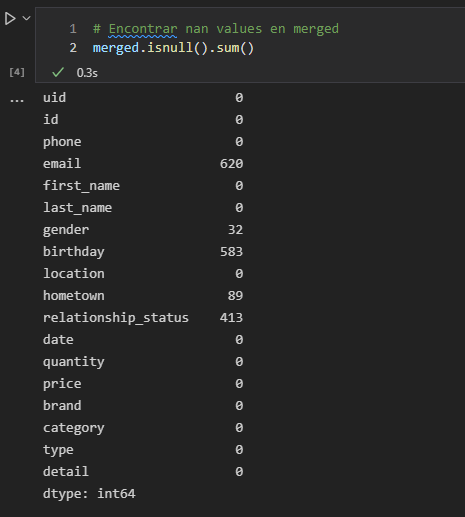
Se utilizan varias técnicas para la implementación de la metodología Agile Scrum, que se incluyen en este trabajo. Para implementar estas técnicas, podemos utilizar diversas herramientas que permitirán que todo el trabajo sea transparente y visible para todos los participantes de un proyecto.

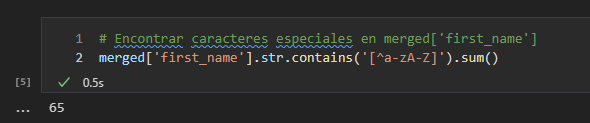
Para crear las historias de usuario y medir así la cantidad de trabajo realizada en cada Sprint se utilizó la plataforma Trello online.

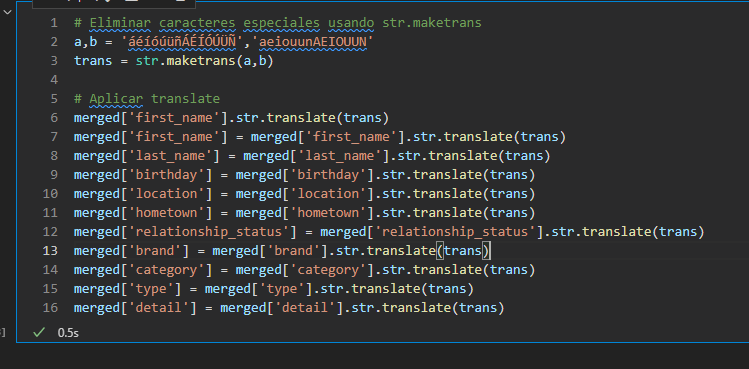
Trello es una herramienta de colaboración visual que le permite organizar y priorizar proyectos de una manera divertida, flexible y gratificante. Un tablero de Trello es una serie de listas, con un montón de tarjetas adjuntas y repletas de potentes funciones y automatización.

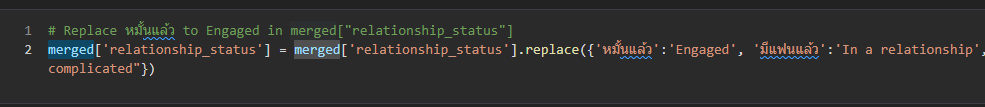
### 2.1.2. Limpieza de datos

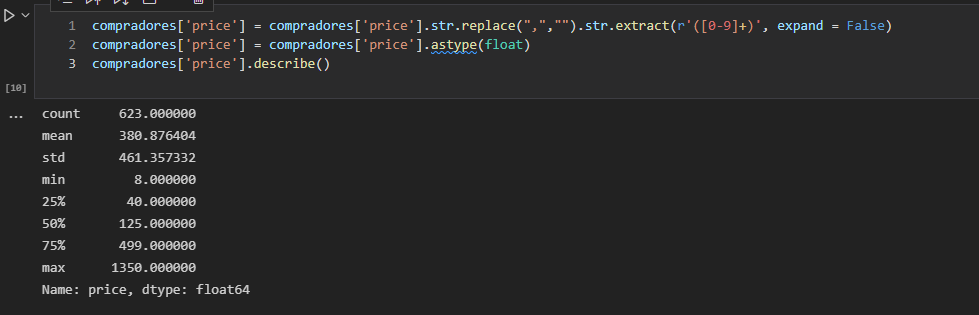
Para realizar la limpieza de datos primero encontramos los valores vacios en el set de datos

Luego vimos si existen caracteres especiales en los nombre

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

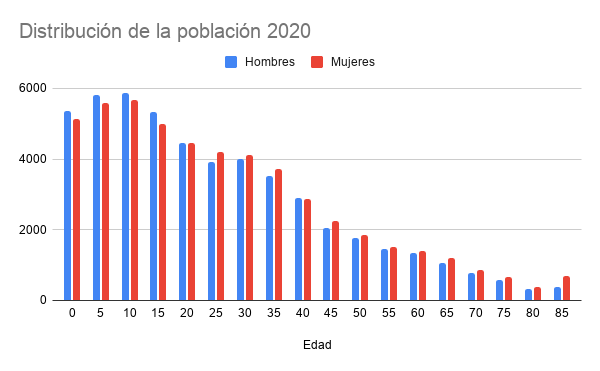
Luego ya que el programa por defecto tiene el idioma indones, se reemplazaron estas palabras por su equivalente al inglés.

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



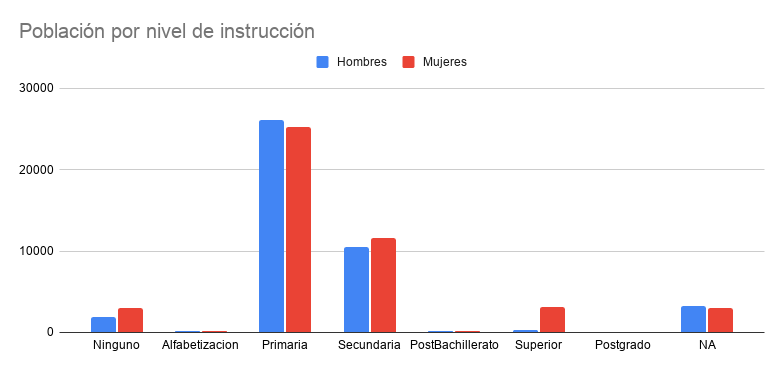
### 2.1.3. Datos demográficos de Tulcán

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).



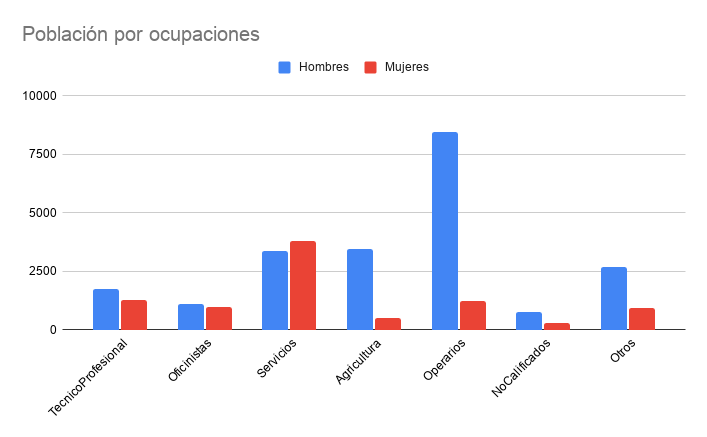
**Figura 12** Distribución por edad y género de la población de Tulcán

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



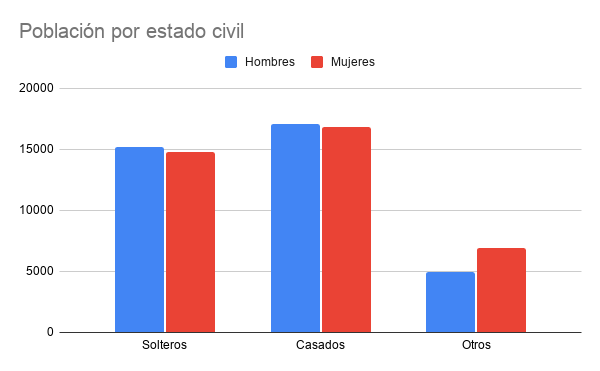
**Figura 13** Población por género y nivel de instrucción

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



**Figura 14** Población por ocupación

Y por último, para el estado civil.

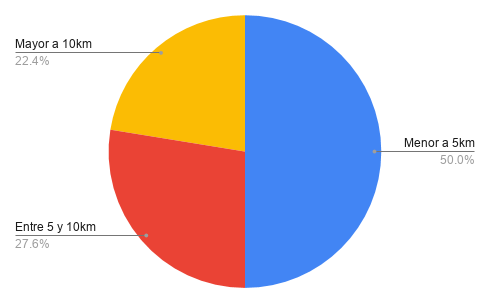


**Figura 15** Población por estado civil

### 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).



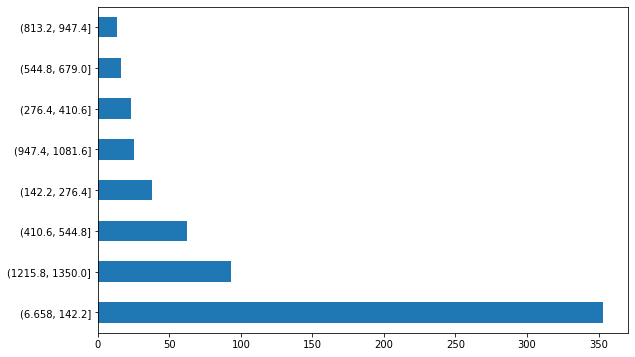
**Figura 16** Distribución de ciclistas por recorrido

### 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 1712** Compras en el almacén por categoría

**Figura 18** Compras en el almacén por rango de precios

**Figura 19** 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 20** Ingresos por venta de bicicletas 2016-2020

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

**Figura 21** Flujo de efectivo 2016-2020

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

**Figura 22** 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.

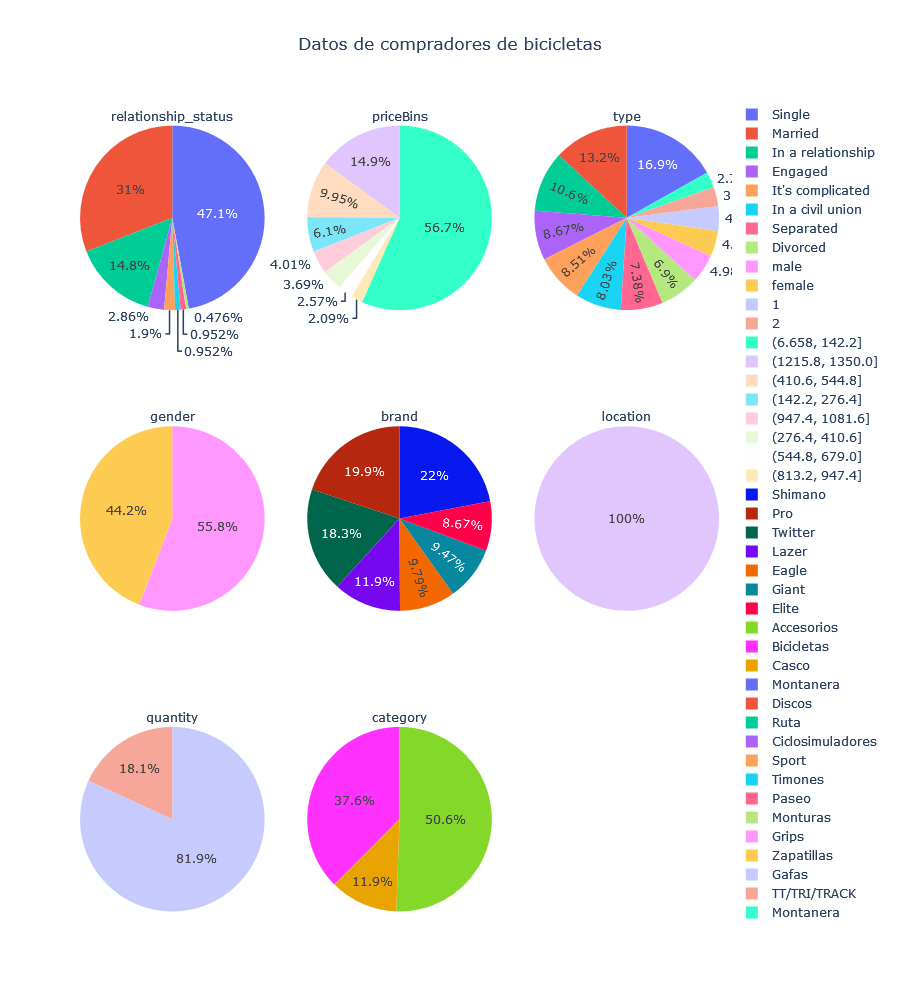
Una variable importante es la utilidad que tienen las empresas y se observa que en el periodo 2018-2020 cae abruptamente. Esto se pude explicar por la situación de la emergencia sanitaria que se desató desde principios de 2020 y que ha tenido repercusiones socioeconómicas fuertes en el territorio ecuatoriano 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

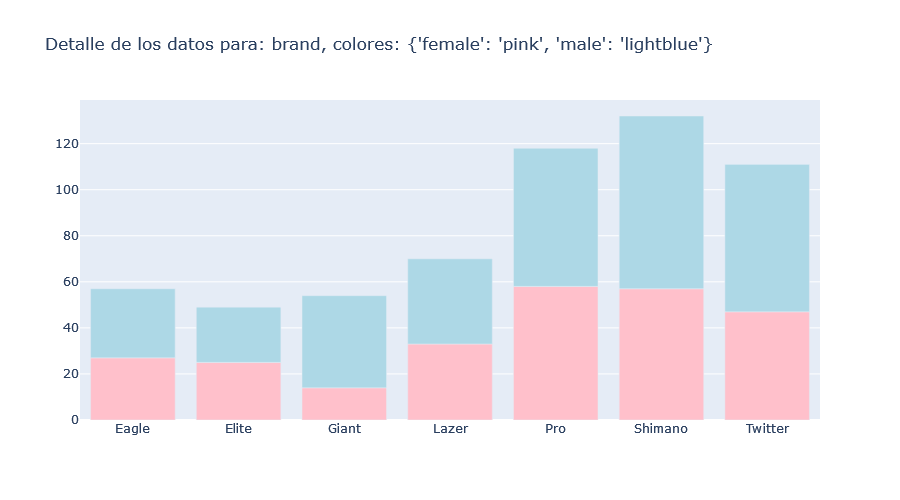
Para la visualización del prototipo se ha usado el entorno de trabajo Dash. Dash es un entorno de trabajo para Python que nos ayuda a crear aplicaciones web analíticas. Dash está escrito sobre Flask, Plotly y React JS. Es ideal para construir visualizaciones con interfaces de usuario personalizadas. 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.

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.

**Figura 23** 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 24** Ejemplo de grafica de Detalle/Diagnostico.

## 2.2. Reglas de asociación

**2.2.1. Metodología Apriori**

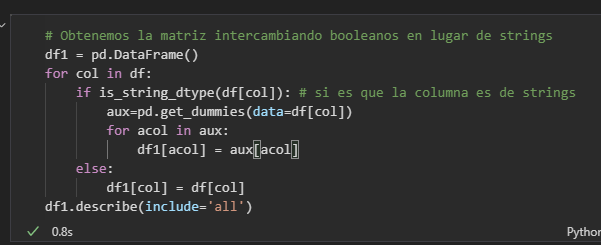
El funcionamiento del algoritmo Apriori empieza con la obtención de los llamados “conjuntos de ítems frecuentes”, los cuales son aquellos conjuntos cuyos ítems superan un umbral que define un valor mínimo para la medida de soporte. Debido al amplio uso del algoritmo Apriori, desde que se formalizó la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

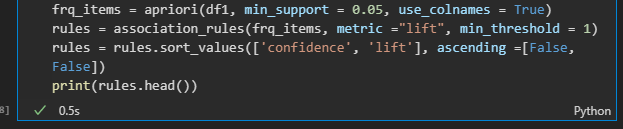
El algoritmo a priori ayuda a relacionar los datos, a partir de reglas de asociación, las cuales son basadas en las variables definidas en el pre-procesamiento. Este algoritmo es muy usado en la minería de datos por su confianza y certeza.

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.

**2.2.2. Implementación de Apriori**

- Primero cambiamos los datos categóricos por datos booleanos para poder utilizar cualquier método de asociación, lo hacemos con el método str.get\_dummies():

Luego, utilizamos apriori en python:

Los 15 primeros resultados para la base de datos de Compradores son los siguientes:

**Tabla 5** Resultados de la base de datos

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
| ({'Accesorios'}) | ({'Pro', 'Monturas'}) | 0.50561797752809 | 0.069020866773676 | 0.069020866773676 | 0.136507936507936 | 1.97777777777778 | 0.034122675708334 | 1.07815598149372 |
| ({'Accesorios'}) | ({'Shimano', 'female', 'Tulcan'}) | 0.50561797752809 | 0.069020866773676 | 0.069020866773676 | 0.136507936507936 | 1.97777777777778 | 0.034122675708334 | 1.07815598149372 |
| ({'male'}) | ({'Discos'}) | 0.529695024077047 | 0.131621187800963 | 0.07223113964687 | 0.136363636363636 | 1.0360310421286 | 0.002512051405589 | 1.00549125623046 |
| ({'male'}) | ({'Shimano', 'Discos'}) | 0.529695024077047 | 0.131621187800963 | 0.07223113964687 | 0.136363636363636 | 1.0360310421286 | 0.002512051405589 | 1.00549125623046 |
| ({'male'}) | ({'Accesorios', 'Discos'}) | 0.529695024077047 | 0.131621187800963 | 0.07223113964687 | 0.136363636363636 | 1.0360310421286 | 0.002512051405589 | 1.00549125623046 |
| ({'male'}) | ({'Shimano', 'Accesorios', 'Discos'}) | 0.529695024077047 | 0.131621187800963 | 0.07223113964687 | 0.136363636363636 | 1.0360310421286 | 0.002512051405589 | 1.00549125623046 |
| ({'female'}) | ({'Discos'}) | 0.418940609951846 | 0.131621187800963 | 0.056179775280899 | 0.134099616858238 | 1.01883001588637 | 0.001038314580977 | 1.00286225656614 |
| ({'female'}) | ({'Shimano', 'Discos'}) | 0.418940609951846 | 0.131621187800963 | 0.056179775280899 | 0.134099616858238 | 1.01883001588637 | 0.001038314580977 | 1.00286225656614 |
| ({'female'}) | ({'Accesorios', 'Discos'}) | 0.418940609951846 | 0.131621187800963 | 0.056179775280899 | 0.134099616858238 | 1.01883001588637 | 0.001038314580977 | 1.00286225656614 |
| ({'female'}) | ({'Shimano', 'Accesorios', 'Discos'}) | 0.418940609951846 | 0.131621187800963 | 0.056179775280899 | 0.134099616858238 | 1.01883001588637 | 0.001038314580977 | 1.00286225656614 |
| ({'Tulcan'}) | ({'Married'}) | 0.626003210272873 | 0.104333868378812 | 0.080256821829856 | 0.128205128205128 | 1.22879684418146 | 0.014943485284532 | 1.02738173921254 |
| ({'female'}) | ({'Lazer'}) | 0.418940609951846 | 0.118780096308186 | 0.052969502407705 | 0.126436781609195 | 1.06446101273687 | 0.003207696410214 | 1.00876488975247 |
| ({'female'}) | ({'Casco'}) | 0.418940609951846 | 0.118780096308186 | 0.052969502407705 | 0.126436781609195 | 1.06446101273687 | 0.003207696410214 | 1.00876488975247 |
| ({'female'}) | ({'Casco', 'Lazer'}) | 0.418940609951846 | 0.118780096308186 | 0.052969502407705 | 0.126436781609195 | 1.06446101273687 | 0.003207696410214 | 1.00876488975247 |

Tabla 1. Primeros resultados de Apriori

Como se puede observar en la tabla anterior, se tiene un gran número de asociaciones muy certeras debido a que pares de columnas como marca-modelo y modelo-descripción tienen una relación directa, así que se ha hecho una nueva iteración utilizando solamente el género del comprador y el tipo de artículo de compra que fue el par de columnas con los que se obtuvieron los mejores resultados, que son los siguientes:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
| ({'Ruta'}) | ({'male'}) | 0.105939004815409 | 0.529695024077047 | 0.062600321027287 | 0.590909090909091 | 1.11556473829201 | 0.006484957320891 | 1.14963438558944 |
| ({'Montanera'}) | ({'male'}) | 0.168539325842697 | 0.529695024077047 | 0.099518459069021 | 0.59047619047619 | 1.11474747474747 | 0.010244016808845 | 1.14841912725372 |
| ({'Discos'}) | ({'male'}) | 0.131621187800963 | 0.529695024077047 | 0.07223113964687 | 0.548780487804878 | 1.0360310421286 | 0.002512051405589 | 1.04229751420763 |
| ({'Discos'}) | ({'female'}) | 0.131621187800963 | 0.418940609951846 | 0.056179775280899 | 0.426829268292683 | 1.01883001588637 | 0.001038314580977 | 1.01376319114784 |
| ({'male'}) | ({'Montanera'}) | 0.529695024077047 | 0.168539325842697 | 0.099518459069021 | 0.187878787878788 | 1.11474747474747 | 0.010244016808845 | 1.02381351668623 |
| ({'male'}) | ({'Discos'}) | 0.529695024077047 | 0.131621187800963 | 0.07223113964687 | 0.136363636363636 | 1.0360310421286 | 0.002512051405589 | 1.00549125623046 |
| ({'female'}) | ({'Discos'}) | 0.418940609951846 | 0.131621187800963 | 0.056179775280899 | 0.134099616858238 | 1.01883001588637 | 0.001038314580977 | 1.00286225656614 |
| ({'male'}) | ({'Ruta'}) | 0.529695024077047 | 0.105939004815409 | 0.062600321027287 | 0.118181818181818 | 1.11556473829201 | 0.006484957320891 | 1.01388360278665 |

**2.2.3 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 indica de manera porcentual cuantas veces se repite una especificación de antecedente → consecuencia, en este caso vemos que no llega a ser un 10% en ningún caso, lo que quiere decir que son casos específicos pero no aislados ya que la suma de estos 4 casos (tomando en consideracion los repetidos) representan un 30% del total de ventas.

**Confianza**

Muestra cuan probable es que si sucede A, B también suceda, se calcula como el soporte de A, B dividido entre el soporte de A, en este apartado vemos una diferencia muy grande con respecto a las columnas repetidas ya que por ejemplo si un artículo es del tupo “Ruta” es un 60% probable que haya sido comprado por u hombre, pero al revés si un hombre compra un artículo, solo tenemos un 11.8% de probabilidad de que este sea de tipo “Ruta”.

**Lift**

Muestra cuan probable es que si sucede A, B también suceda, si es un valor cercano a 1 quiere decir que no están asociados, si es mayor es más grande su asociación y si es menor que 1 están asociados negativamente, en este caso todos son superiores a 1 por muy poco.

**2.2.4 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 25** 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.

### 2.2.3. Variables a predecir

Para poder hacer nuestro análisis del cantón Tulcán se necesita de las variables del set de datos de compradores, de las que destacamos las columnas más importantes:

* Genero.
* Ocupación.
* Estado civil.
* Cantidad
* Precio
* Fechas

## 2.3. Procesamiento y clústering de datos usando Python

### 2.3.1. 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.

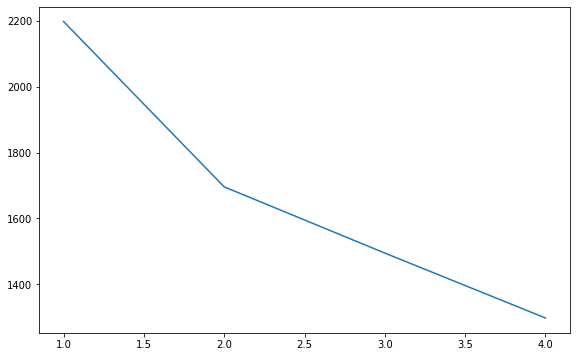
El algoritmo de disimilaciones de k-modes se ha desarrollado inicialmente para la dimensión vertical con el fin de proporcionar una agrupación más robusta que la agrupación de un solo enlace. El algoritmo de k-modos es básicamente una modificación del algoritmo de k-means auto organizado de Hartigan 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.

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.

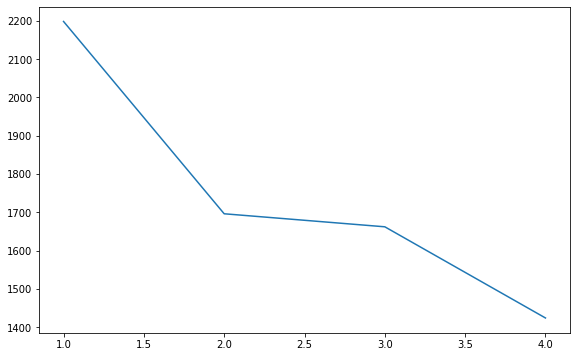
## 2.3.2. Costo de K-Modes

El costo de K-Modes se refiere a la suma del número de diferencias entre grupos. Este está directamente relacionado al número de clústeres que ingresemos a la ecuación.

Además de esto, existen dos tipos de inicialización “Cao” y “Huang”, para elegir cuál de estas vamos a utilizar y el número de clústeres óptimo, se hará una comparativa de ambas ecuaciones utilizando diferentes números de clústeres para determinar la mejor opción para este análisis.



**Figura 26** Inicialización "Cao"



**Figura 137** Inicialización "Huang"

En las Figuras 13 y 14 el eje de las “x” corresponde a el número de clústeres seleccionados el eje de las “y” al costo o las diferencias entre cada clúster, si bien ambas graficas parecen muy similares, la sutil diferencia en los costos del eje de las “y” significan 138 puntos de diferencias que se ganan al aplicar la inicialización “Huang”.

### 2.3.3. Número de clústeres

Una vez elegida la ecuación y la inicialización que vamos a utilizar es hora de determinar el número de clústeres, si bien al ir aumentando este número de clústeres las diferencias entre ellos disminuyen, elegiremos un número tal que nos permita hallar un público objetivo coherente y con ello determinar el arquetipo de “comprador de bicicletas”.

Para ello vamos a buscar el porcentaje de compradores en cada clúster variando el número de clústeres:

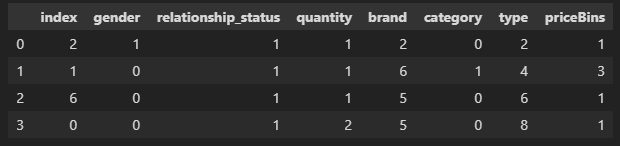
* Para 1 clúster, o sea la totalidad de registros se tiene un 14.92% de compradores para el rango más alto de precio (entre $1215.80 y $1350.00).
* Para 2 clústeres, el mayor porcentaje de compradores es de 24.69%.
* Para 3 clústeres, el mayor porcentaje de compradores es de 29.37%.
* Y para 4 clústeres, el mayor porcentaje de compradores es de 46.039%.

Por lo tanto, se trabajará con 4 clústeres para esta investigación.

### 2.3.4. Resultado: Arquetipo de comprador

Los arquetipos son personajes modelo que representan grupos de clientes en función de sus motivaciones, necesidades, comportamientos y estilos de vida. Son especialmente útiles en marketing y ventas porque ofrecen una plantilla para anticipar cómo responderá un cliente específico.

**Tabla 6** Moda de cada uno de los clústeres

Para determinar el arquetipo de comprador, o mercado objetivo en el que deberíamos centrarnos para el negocio de las bicicletas se utilizara la moda del grupo perteneciente al clúster con mayor porcentaje de compradores en el rango de precios alto.

Observamos que el arquetipo de comprador tiene las siguientes características:

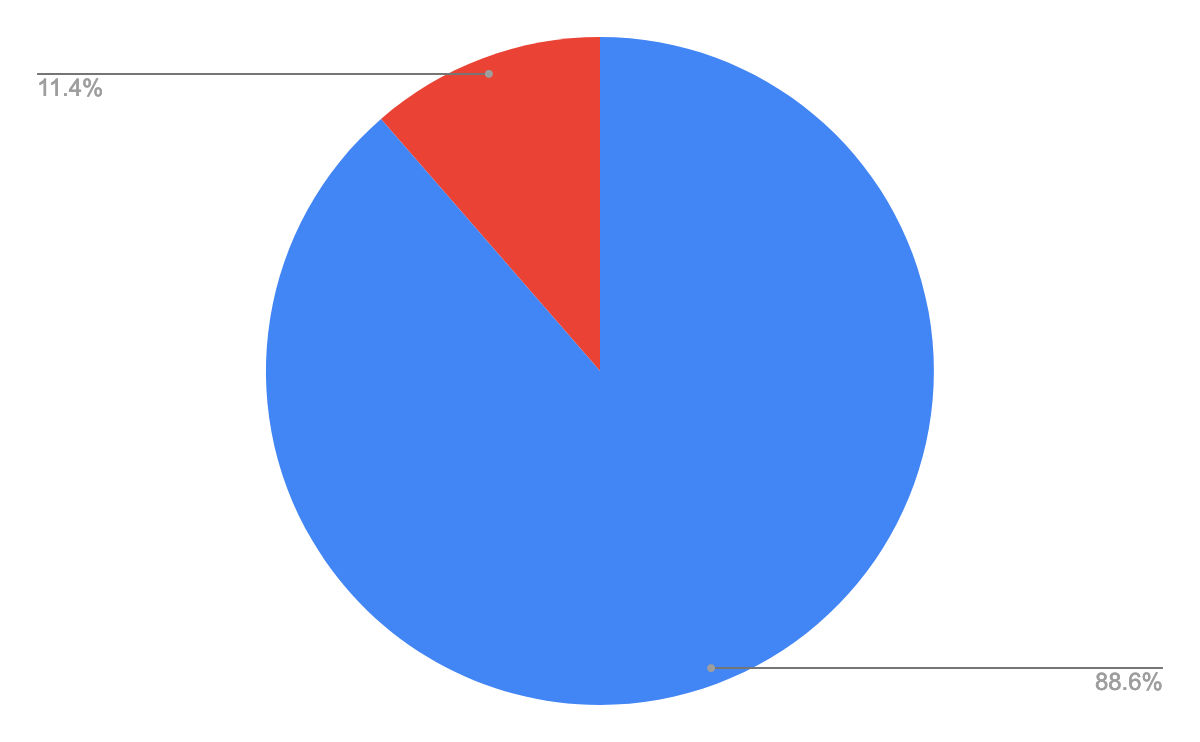
* Género: Masculino.
* Estado Civil: No ingresado
* Ítems comprados:1
* Marca: Twitter.
* Categoría: Bicicletas.
* Tipo: Montañera

## 2.4. Pruebas

### 2.4.1. 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 28** 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.

### 2.4.2. 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.

### 2.4.3. 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.

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.

# CAPITULO III

# 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.

### 3.1.1.1. 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.

****

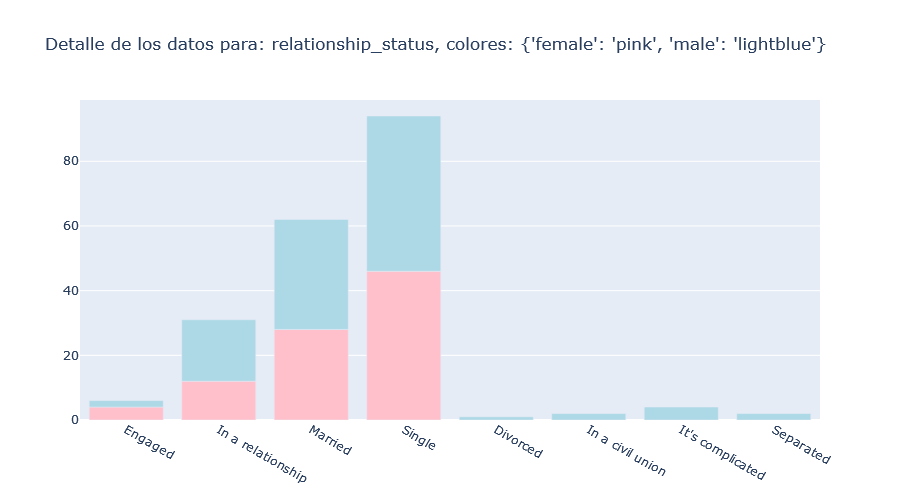
**Figura 29** Género de compradores

### 3.1.1.2. Estado Civil

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

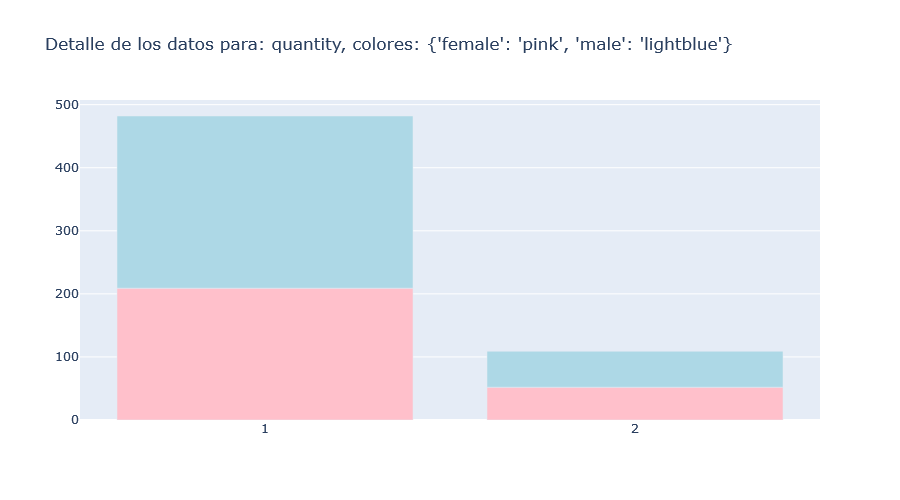
* 0: Solteros
* 1: No especificado
* etc...

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.

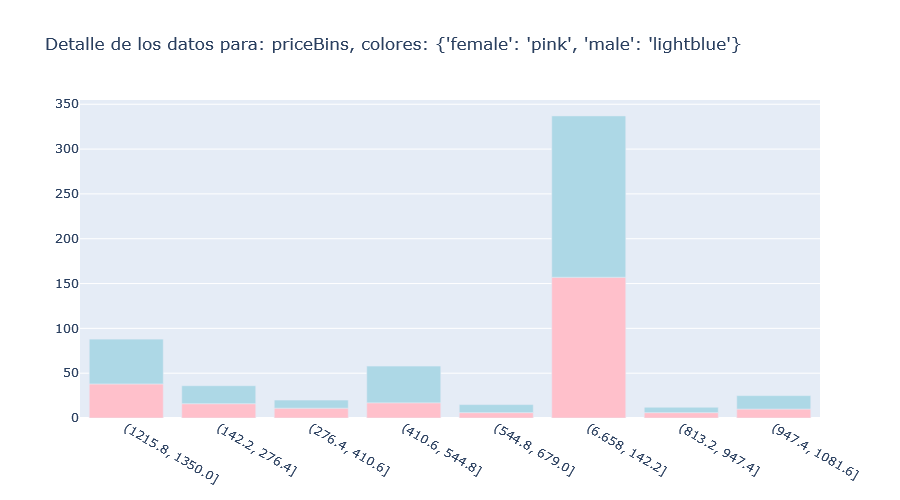
**Figura 30** Estado civil de compradores

**Figura 14** Estado civil de los compradores

### 3.1.2.3. Cantidad de productos

**Figura 31** Cantidad de compras

### 3.1.2.3. Rangos de precios

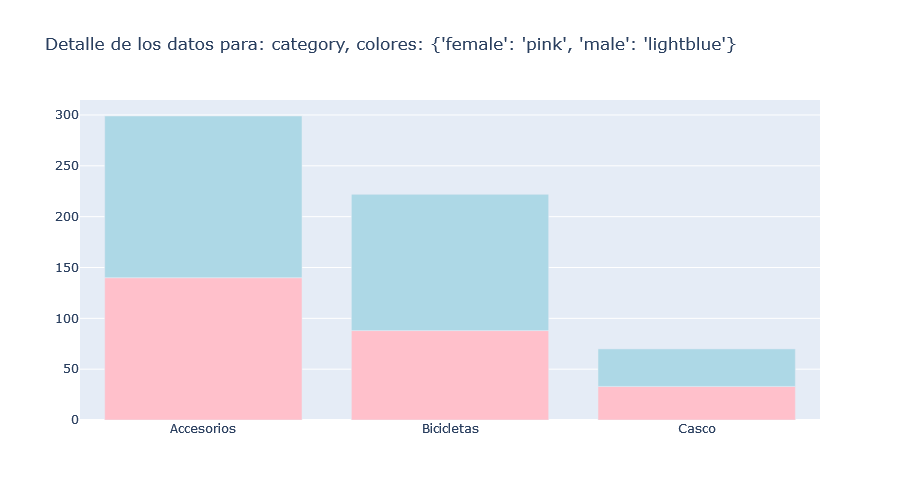
**Figura 32** Cantidad de compras por rango de precios

### 3.1.2.3. Marcas

## 

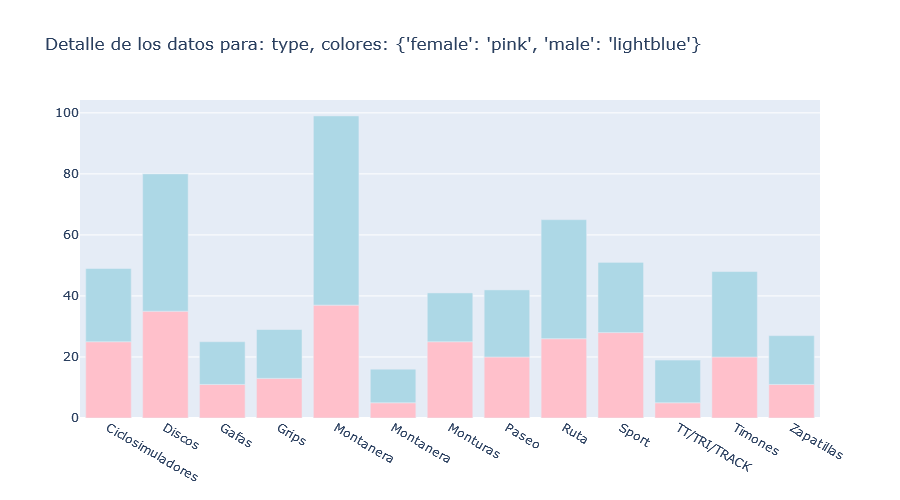
**Figura. 153** Compras por marca y género

### 3.1.2.3. Categorías



**Figura 34** Compras por marca y género

### 3.1.2.3. Tipos



**Figura 35** 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.

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.

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.

### 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.

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.

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 y Recomendaciones

## Conclusiones

Las reglas de asociación nos sirven para comprender las relaciones entre datos que a priori no son evidentes, para así entender la demografía de los compradores, adicionalmente, ayuda también a sacar conclusiones antes de modelar sabiendo datos preliminares.

El algoritmo Kmodes es un digno sustituto de Kmeans creando clústeres para datos categóricos que nos ayudan a determinar la moda de grupos de datos y esto ayuda encontrar arquetipos de compradores.

## Recomendaciones

Se recomienda probar ambos algoritmos de Kmodes y utilizar varias entradas de números de clusters ya que dependiendo el modelo y el procesamiento pueden haber valores más efectivos para resolver un problema, además de definir claramente la variable de salida que se espera para que sea más fácil encontrar un clúster que satisfaga nuestras necesidades.

El algoritmo Apriori nos puede dar una buena aproximación a las relaciones entre datos y demostrarnos si los datos son muy relacionados o incluso sesgos, lo que nos ayuda a comprender de mejor manera los datos con los que estamos trabajando.

La limpieza de datos es fundamental para el correcto funcionamiento de los algoritmos, comprender también los tipos de datos y el tratamiento de los mismos ayuda a tener u n desarrollo más ordenado, conciso y datos organizados de la mejor manera.

Probar varias combinaciones de datos para las gráficas y los modelos, ya que la selección de las mismas puede afectar al desempeño del modelo, función o grafica que estamos realizando, saber también que tipo de dato existe en cada columna con el comando “describe” es de gran ayuda para entender el contexto de los datos.

# 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.*