UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES

TEMA:

Análisis de datos para la evaluación del rendimiento y optimización de las ventas de carros de marca Peugeot en el Ecuador.

AUTORES:

Jiménez Polanco, Angie Stephanie
Lindao Barros, Dennise Nicole

Trabajo de titulación previo a la obtención del título de

Licenciada en Negocios Internacionales

TUTOR:

Ing. Carrera Buri, Félix Miguel, Mgs

Guayaquil, Ecuador

7 de febrero del 2025

UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA

CARRERA DE NEGOCIOS INTERNACIONALES

CERTIFICACIÓN

Certificamos que el presente trabajo de titulación, fue realizado en su totalidad por **Jiménez Polanco, Angie Stephanie, Lindao Barros, Dennise Nicole** como requerimiento para la obtención del título de **Licenciadas en Negocios Internacionales**.

TUTOR

f._____
Ing. Carrera Buri, Félix Miguel, Mgs

DIRECTOR DE LA CARRERA

f._____
Ing. Hurtado Cevallos, Gabriela Elizabeth, Mgs

Guayaquil, a los 7 del mes de febrero del año 2025

UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES

DECLARACIÓN DE RESPONSABILIDAD

Yo, **Jiménez Polanco, Angie Stephanie**
Lindao Barros, Dennise Nicole

DECLARO QUE:

El Trabajo de Titulación, **Análisis de datos para la evaluación del rendimiento y optimización de las ventas de carros de marca Peugeot en el Ecuador,** previo a la obtención del título de **Licenciada en Negocios Internacionales**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, a los 7 del mes de febrero del año 2025

LAS AUTORAS:

Jiménez Polanco, Angie Stephanie

Lindao Barros, Dennise Nicole

UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES

AUTORIZACIÓN

Yo, **Jiménez Polanco, Angie Stephanie**
Lindao Barros, Dennise Nicole

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación, **Análisis de datos para la evaluación del rendimiento y optimización de las ventas de carros de marca Peugeot en el Ecuador**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los 7 del mes de febrero del año 2025

LAS AUTORAS:

Jiménez Polanco, Angie Stephanie

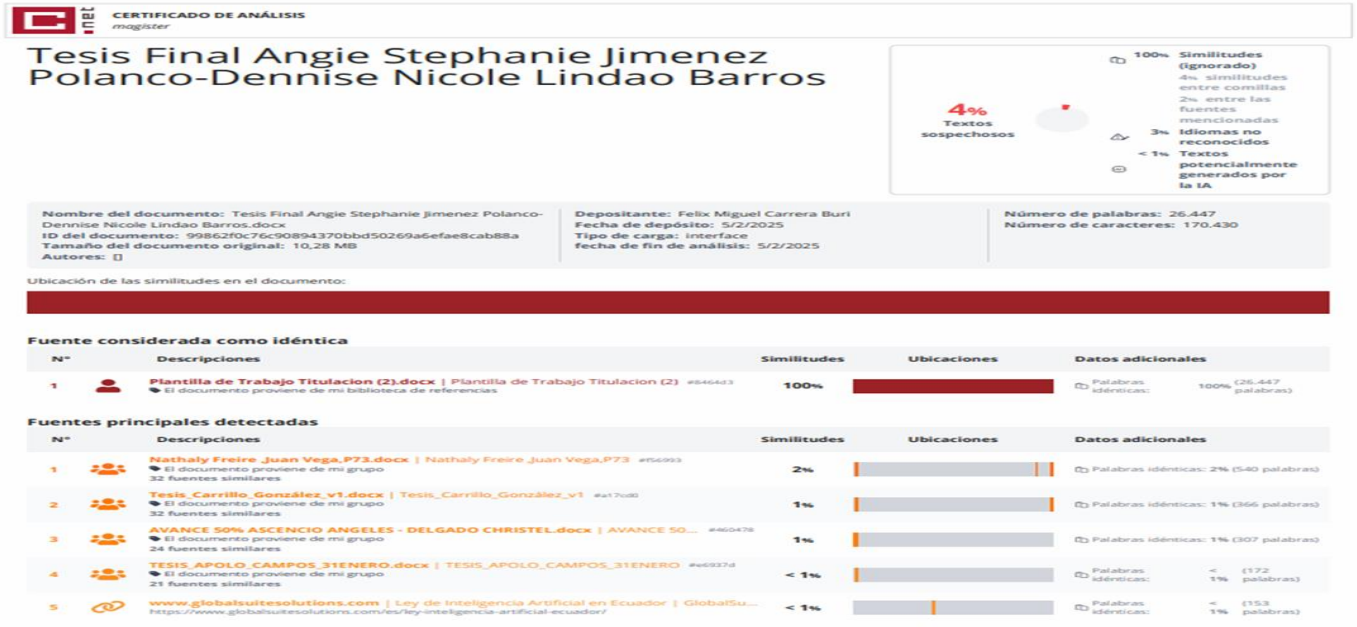
Lindao Barros, Dennise Nicole

UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA

CARRERA DE NEGOCIOS INTERNACIONALES

Reporte Compilatio



CERTIFICADO DE ANÁLISIS
magister

Tesis Final Angie Stephanie Jimenez Polanco-Dennise Nicole Lindao Barros

4% Textos sospechosos

100% Similitudes (ignorado)
4% similitudes entre comillas
2% entre las fuentes mencionadas
3% Idiomas no reconocidos
< 1% Textos potencialmente generados por la IA

Nombre del documento: Tesis Final Angie Stephanie Jimenez Polanco-Dennise Nicole Lindao Barros.docx
ID del documento: 99862f0c76c90894370bbd50269a6efae8cab88a
Tamaño del documento original: 10,28 MB
Autores: []

Depositante: Felix Miguel Carrera Buri
Fecha de depósito: 5/2/2025
Tipo de carga: interface
fecha de fin de análisis: 5/2/2025

Número de palabras: 26.447
Número de caracteres: 170.430

Ubicación de las similitudes en el documento:

Fuente considerada como idéntica

N°	Descripciones	Similitudes	Ubicaciones	Datos adicionales
1	Plantilla de Trabajo Titulacion (2).docx | Plantilla de Trabajo Titulacion (2) #846403 — El documento proviene de mi biblioteca de referencias	100%		Palabras idénticas: 100% (26.447 palabras)

Fuentes principales detectadas

N°	Descripciones	Similitudes	Ubicaciones	Datos adicionales
1	Nathaly Freire_Juan Vega_P73.docx | Nathaly Freire_Juan Vega_P73 #f56993 — El documento proviene de mi grupo — 32 fuentes similares	2%		Palabras idénticas: 2% (540 palabras)
2	Tesis_Carrillo_González_v1.docx | Tesis_Carrillo_González_v1 #a17cd0 — El documento proviene de mi grupo — 32 fuentes similares	1%		Palabras idénticas: 1% (366 palabras)
3	AVANCE 50% ASCENCIO ANGELES - DELGADO CHRISTEL.docx | AVANCE 50... #460478 — El documento proviene de mi grupo — 24 fuentes similares	1%		Palabras idénticas: 1% (307 palabras)
4	TESIS_APOLO_CAMPOS_31ENERO.docx | TESIS_APOLO_CAMPOS_31ENERO #e69374 — El documento proviene de mi grupo — 21 fuentes similares	< 1%		Palabras idénticas: < 1% (172 palabras)
5	www.globalsuitesolutions.com | Ley de Inteligencia Artificial en Ecuador | GlobalSu... — https://www.globalsuitesolutions.com/es/ley-inteligencia-artificial-ecuador/	< 1%		Palabras idénticas: < 1% (153 palabras)

Ing. Carrera Buri, Félix Miguel, Mgs.

Tutor

Ing. Hurtado Cevallos, Gabriela Elizabet, Mgs

Director de carrera

Agradecimientos

Agradezco a dios principalmente por guiarme y permitirme culminar con éxito mi etapa universitaria, agradezco también a mis padres Carlos y Rossy por apoyarme, guiarme, no dejarme sola no solo en esta etapa sino a lo largo de toda mi vida y por ser mi motivación día a día, son lo más importante y valioso que tengo. A mi familia, en especial a mis abuelos, tías y tíos por ser un pilar fundamental en mi vida y estar para mí acompañándome tanto en mis triunfos como derrotas. A mis hermanos, por estar a mi lado siempre, apoyarme y estar pendiente de mí a la distancia.

También a mis compañeros, con los cuales pasé grandes momentos dentro de mi etapa universitaria, que me ayudaron y acompañaron en mis largas noches de estudio y que hicieron que estos últimos años de mi vida sean más entretenidos y fáciles de sobrellevar a pesar de las dificultades, a mi compañera de tesis que más allá de ser mi compañera se convirtió en una amiga para mi, por escucharme siempre y apoyarme en todo aspecto, por la paciencia que me tuvo durante toda mi vida universitaria y por confiar en mí muchas veces más de lo que yo lo hacía.

Y por último, agradezco a mi tutor de tesis Félix Carrera por su guía y apoyo durante todo este proceso, además por ser un gran profesor de los que enseñan con dedicación y te hacen interesar más por la materia.

- Angie Stephanie Jiménez Polanco

Dedicatoria

Quiero dedicar este logro a dios por ser mi guía durante todo el camino y a mis padres por su apoyo incondicional, sin ellos nada de esto hubiera sido posible. Se lo dedico también a mi familia que son mi ejemplo a seguir. Y por último a mí misma por todo el esfuerzo y dedicación puesto en este trabajo.

- Angie Stephanie Jiménez Polanco

Agradecimientos

Quiero comenzar agradeciendo a una mujer, una con un corazón humilde pero con un carácter fuerte, aquella que un día se convirtió en madre y que desde ese momento ha dedicado y sigue dedicando su vida a mí. Mami, este logro mío es tuyo, tú que antes de ser mujer y esposa, has sido mi mami Lorena, la que me levantaba temprano para llevarme al kínder con mis dos trencitas y ha estado en cada uno de mis momentos, buenos y malos, la que aparte de ser mi madre, ha sido mi amiga y compañera de aventuras. No te imaginas lo agradecida que estoy con Dios por haberme hecho tu hija.

Papi Marcelo, eres el hombre de mi vida, el que se levantaba en las noches a hacerme dormir cuando no podía, tenía 3 años pero claro que me acuerdo, el hombre que se hacía tiempo en el trabajo para irme a recoger todos los días a la escuelita, no sabes lo feliz que me hacía verte llegar en tu carro blanco con tu uniforme de trabajo, estresado, cansado o con mil pendientes pero siempre tuviste y has tenido tiempo para mí. Cada vez que veo y recuerdo lo que haces por mí, no me cabe la duda de que sí existe un hombre que me ama infinitamente en este mundo y eres tú.

Gracias a Dios y a la vida por tener la bendición de criarme en un hogar lleno de amor y paciencia, porque lo más bonito es poder caminar por este sendero llamado vida de la mano de ustedes dos.

Es necesario también agradecer a mis amigos de universidad, sin ustedes los momentos difíciles no hubiesen sido tan llevaderos, porque todos esos momentos de estudio fueron más divertidos por ustedes, siempre tuvieron un chiste, una broma para hacerme reír, mejor grupo de amigos no puede haberme encontrado, unos desde el pre y otros se unieron con el tiempo y que ahora nos vamos a graduar juntos.

Gracias a mis hermanas de otras madres, mis queridas niñas, mi vida no sería lo mismo sin ustedes, cada quien tomó rumbos diferentes pero hemos sido incondicionales, hemos reído, llorado y hecho tantas cosas juntas que ya no somos amigas, somos familia.

Agradecer a mi tutor de tesis, Félix Carrera, por ser mi trauma más personal en la carrera y mi motivación a hacer una profesión en Análisis de datos, gracias por darnos su conocimiento y experiencia, ya que sin su guía este trabajo no hubiese sido posible. Espero poder coincidir algún día en el mundo laboral.

Finalmente quiero agradecer a esa niña asustada, que traía mil preguntas en la cabeza, cada noche sin dormir, cada momento de estrés, frustración y preocupación valió la pena porque mira a donde llegaste. No te rendiste, y por muy difícil que se puso no dejaste de luchar. Te comunico con mucho orgullo que pronto te vas a graduar de blanco, Dennise, ¡lo hiciste bien!

- Lindao Barros, Dennise Nicole.

Dedicatoria

Con mucho orgullo y felicidad quisiera dedicar este trabajo de investigación a las personas que han estado junto en todo momento, comenzando por mi mamá, gracias por siempre acompañarme, educarme con tanto amor y paciencia, por ser mi luz, mi guía, mi hermana y amiga. Algún día quisiera llegar a ser la mitad de noble, fuerte y valiente que tú. A mi papi, gracias por estar conmigo, por protegerme y tenerme tanta paciencia, porque eres mi primer y eterno amor, gracias por hacerme tu princesita y dedicar tu vida a mí para ser amada de esa manera incondicional.

Gracias a las niñas que conocí hace 10 años ya, aquellas que hoy son unas mujeres increíbles, nos hemos visto crecer, cambiar y madurar y seguimos juntas apoyándonos en todo, gracias por mostrarme que la amistad honesta y sin envidias si existe y que después de tanto tiempo seguimos riendo, conversando y recolectando momentos.

A los amigos que hice en la universidad, entré con muchas dudas, pero gracias a ellos esta etapa de mi vida fue inolvidable, gracias por las risas, anécdotas, nuestros planes y esa amistad que fue un rayito de sol cuando las cosas se ponían complejas. Estas personas increíbles que pronto pasarán de ser compañeros a ser colegas.

Gracias a las personas que ya no están, ojalá estuvieran aquí conmigo, pero sé que desde una parte bonita del cielo están sintiéndose orgullosos de la persona que soy y todo lo que he logrado. Esa niña pequeña se convirtió en una mujer valiente, fuerte y decidida a conseguir aquello que se proponga.

Finalmente, quiero agradecerme a mí misma porque a pesar de todo lo que viví, me pregunté y hasta dudé de mí, pude demostrarme que soy lo suficientemente capaz

de convertir inseguridades y miedos en fortalezas y aprendizajes. Las cosas no fueron siempre lindas pero estoy agradecida porque cada situación hizo de mí la persona que soy ahora y realmente me siento muy orgullosa de esta versión mía. Sigue soñando despierta, trabaja por tu futuro y lucha por aquello que tanto quieres, y recuerda lo que un día te dijeron: “algún día serás recompensada por ser la persona que eres”.

Quiero finalizar con una frase que me ha servido de motivación en todos los momentos en los que dudaba: “Hay una niña chiquita que te está mirando orgullosa mientras cumples tus sueños, tiene tu pelo, tus ojos y tu risa”. Pequeña Dennise, sigues igual de sentimental, pero valió la pena niña, todo valió la pena porque llegaste a este momento y aún te falta mucho por lograr y vivir.

- Lindao Barros, Dennise Nicole

UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA

CARRERA DE NEGOCIOS INTERNACIONALES

TRIBUNAL DE SUSTENTACIÓN

Ing. Carrera Buri, Félix Miguel, Mgs

TUTOR

Ing. Hurtado Cevallos, Gabriela Elizabet, Mgs

DECANO O DIRECTOR DE CARRERA

Lic. Freire Quintero, César Enrique

COORDINADOR DEL ÁREA O DOCENTE DE LA CARRERA

Índice

Índice de tablas

Índice de figuras

Resumen

Este trabajo de investigación se basa en analizar el rendimiento y optimización de las ventas de los vehículos Peugeot en Ecuador. Después de un análisis exhaustivo pudimos notar que actualmente aquí en Ecuador no se le da tanta importancia a la inteligencia artificial en las empresas, es por esto que se ve la necesidad de incrementar técnicas de machine learning para analizar el rendimiento de las ventas de los autos Peugeot, y con esta información poder desarrollar recomendaciones para su optimización.

Para poder realizar de manera exitosa este análisis, primero se revisaron conceptos claves para entender más a profundidad los modelos a usar, luego se aplicaron dichos conceptos en la práctica por medio de RStudio donde se realizó primero un modelo de k-means donde se lograron segmentar los clientes en 3 grupos y después en base a esta información se logró predecir las ventas en dólares de los dos años siguientes para cada segmento y de esta manera ver cual era el grupo de clientes que generaba más ingresos para la empresa y cuál era el segmento que menos compraba autos de esta marca.

Al final como resultados obtuvimos que el segmento que más autos Peugeot compraba era el segmento 3 de clase alta, pero decidimos dar recomendaciones distintas para incrementar las ventas de cada uno de estos grupos distintos de clientes, enfocándonos un poco más en el segmento con más ventas y en mantener la calidad y esencia de la marca.

Palabras claves: Optimización, Análisis de datos, Machine Learning, Forecasting, K-means, Segmentación de mercados, Peugeot, Rendimiento de ventas.

Abstract

This research work is based on analyzing the performance and optimization of sales of Peugeot vehicles in Ecuador. After an exhaustive analysis we could notice that currently here in Ecuador is not given so much importance to artificial intelligence in companies, this is why there is a need to increase machine learning techniques to analyze the performance of sales of Peugeot cars, and with this information can develop recommendations for their optimization.

To be able to successfully perform this analysis, we first reviewed key concepts to understand in more depth the models to be used, then applied these concepts in practice by RStudio where a k-means model was first performed where we managed to segment customers into 3 groups and then based on this information we were able to predict the sales in dollars of the next two years for each segment and thus see which group of customers was generating more revenue for the company and which was the segment that bought the least cars of this brand.

In the end as results we obtained that the segment that most cars Peugeot bought was the high class 3 segment, but we decided to give different recommendations to increase the sales of each of these different groups of customers, focusing a little more on the segment with more sales and maintaining the quality and essence of the brand.

Keywords: Optimization, Data Analysis, Machine Learning, Forecasting, K-means, Market segmentation, Peugeot, Sales performance.

Résumé

Ce travail de recherche est basé sur l'analyse des performances et l'optimisation des ventes des véhicules Peugeot en Équateur. Après une analyse approfondie, nous avons pu constater que l'intelligence artificielle n'est pas aussi importante dans les entreprises en Équateur, c'est pourquoi il est nécessaire d'augmenter les techniques de machine learning pour analyser la performance des ventes de voitures Peugeot, et avec cette information peut développer des recommandations pour son optimisation.

Pour être en mesure de réaliser cette analyse avec succès, nous avons d'abord examiné les concepts clés pour comprendre plus en profondeur les modèles à utiliser, puis nous avons appliqué ces concepts dans la pratique au moyen de RStudio où nous avons d'abord réalisé un modèle de K-means où ils ont réussi à segmenter les clients en 3 groupes et puis sur la base de cette information, nous avons pu prédire les ventes en dollars des deux années suivantes pour chaque segment et ainsi voir quel groupe de clients générait le plus de revenus pour l'entreprise et ce qui était le segment qui achetait moins de voitures de cette marque.

Finalement, nous avons obtenu que le segment qui achetait le plus de voitures Peugeot était le segment 3 classe supérieure, mais nous avons décidé de donner des recommandations différentes pour augmenter les ventes de chacun de ces différents groupes de clients, nous concentrer un peu plus sur le segment de vente et sur le maintien de la qualité et de l'essence de la marque.

Mots-clés: Optimisation, Analyse de données, Machine Learning, Forecasting, K-means, Segmentation du marché, Peugeot, Performance des ventes

Introducción

Peugeot tiene su origen en un molinillo de granos, tan pequeño que puede llegar a ser curioso cómo se convirtió en una de las marcas francesas de automóviles más reconocidas del mundo. La marca empezó como un pequeño negocio familiar en el antiguo Franco Condado allá por 1810 en Francia, la historia se remonta a una molinería de harina que tomó el apellido Peugeot como su marca.

Los hermanos Juan Pedro y Juan Federico Peugeot fundaron, junto a su cuñado Jaime Maillard-Salins, la Casa Peugeot en 1810. Su primera labor fue fundir acero, pero luego otros integrantes de la familia diversificaron la producción. Uno de los miembros de la familia, Armando Peugeot, dirigía con gran éxito la fabricación de bicicletas cuando se apasionó por el vehículo autopropulsado. En 1896, Armando fundó Automóviles Peugeot S.A. (Hierro, 2006, p. 66)

La pregunta en cuestión es cómo una empresa familiar de molinillos pasó a ser una marca de automóviles y la respuesta radica en Armand Peugeot, tataranieto de Jean-Pierre Peugeot que junto a su hermano Jean-Georges Peugeot pondrían los cimientos de la empresa que hoy se conoce. Armand se había interesado en las bicicletas, instrumento que eran novedosos para sus tiempos, lo que hizo que la empresa se dividiera en dos, la que se encaminaba por la fabricación de molinillos y la novedosa empresa de bicicletas, quedando La Société des Automobiles Peugeot a cargo de Armand.

No fue hasta el 1889 cuando el visionario Armand Peugeot se sintió muy atraído por el futuro que se encontraba emergiendo en la industria de los automóviles, se atreve a construir el Peugeot Type 1, un triciclo con motor a vapor, asimismo las motocicletas Peugeot tuvieron lugar en años posteriores alrededor de 1898 y el primer prototipo de un auto Peugeot, llamado Type 2 que tuvo su nacimiento por 1890.

El famoso león que adorna los carros Peugeot fue creado en 1847 por Justin Blazer, y fue registrado en 1858. “Emile Peugeot le pidió a Justin Blazer que creara un emblema. El orfebre se inspiró en las cualidades del producto estrella de Peugeot en ese momento: la sierra. Velocidad, flexibilidad y mordida: es el león la figura emblemática de la marca” (Peugeot, 2023). En la década de 1920, comenzó a ser incorporado en los vehículos de dicha marca. Sin embargo el logo ha experimentado múltiples cambios y transformaciones a lo largo de este tiempo. Inicialmente, variaba

dependiendo del producto: un león en lucha representaba las bicicletas, mientras que la cabeza de un león con un escudo era usada en los automóviles. La marca parecía carecer de una dirección específica en lo que respecta a su logo. Habían ocurrido numerosos cambios en menos de 50 años, el regreso del león heráldico tuvo lugar en 1970, aunque en esta ocasión no estaba ubicado dentro de un escudo. El león erguido sobre sus patas traseras y observando hacia la izquierda, se convertiría en una imagen icónica, contribuyendo a la consistencia de la marca y facilitando su recuerdo. Otra renovación del logo tuvo lugar en 1998, y posteriormente se retocó mínimamente en 2002 y en enero 2010 se volvió a actualizar, quedando como está actualmente.

Peugeot se incorporó al grupo Stellantis en 2021 tras la fusión entre el PSA Group (que incluía a Peugeot, Citroën, DS Automobiles y Opel/Vauxhall) y Fiat Chrysler Automobiles (FCA), conglomerado que contaba con marcas como Fiat, Chrysler, Jeep, Dodge, Alfa Romeo y Maserati.

Lo que la fusión buscaba era unir las fuerzas con la finalidad de enfrentar los difíciles desafíos que presentaba la industria automotriz, es decir la adaptación de los vehículos tradicionales hacia vehículos eléctricos y la creciente exigencia de los consumidores por la sostenibilidad. La compañía Stellantis subió al cuarto nivel de los mayores fabricantes de automóviles a nivel mundial con respecto a los volúmenes de producción que manejan, por esto el nombre de “Stellantis” proviene del antiguo verbo latino stello, que puede traducirse como “iluminar con estrellas”. Al unirse ambas empresas se fusionaron recursos, infraestructuras y tecnologías con la finalidad de potenciar su ventaja competitiva a nivel mundial y optimizar costos. Esto puede reflejarse en la investigación y desarrollo de las nuevas tecnologías correspondientes a la propulsión, así mismo como la creación de plataformas compartidas para los nuevos modelos de carros tanto eléctricos como autónomos.

Peugeot, como parte del grupo Stellantis, ha experimentado un destacado crecimiento a nivel mundial. En Europa, ha consolidado su presencia en el mercado automotriz.

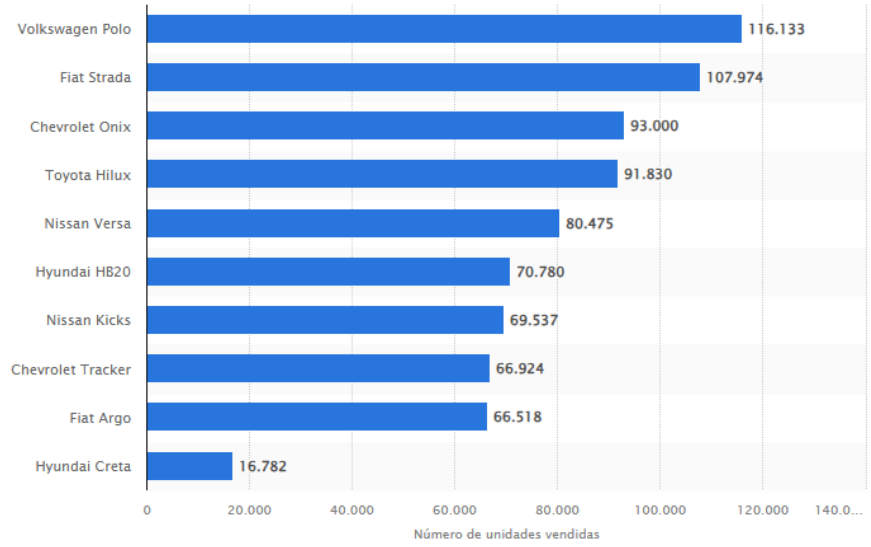
España representó el quinto lugar por volumen de ventas para Peugeot, por detrás de Turquía (78.632 unidades), y del top 3 del ranking compuesto por Reino Unido (88.467 unidades), Italia (91.319 unidades) y Francia, el mercado más potente para la marca con 305.295 unidades (Bacorelle, 2024).

Durante los últimos años Peugeot centró su atención en la transición de sus vehículos hacia la electrificación, de esta manera amplió sus cartera de productos no solo en coches eléctricos sino también híbridos con el objetivo de cumplir con las normativas

medioambientales que establece la Unión Europea y a su vez disminuye las emisiones de carbono fuera de Europa. Por otro lado, Peugeot ha aumentado su presencia en nuevos mercados emergentes tanto en América Latina, África y Asia, sin embargo dentro de los mismos existe una fuerte competencia. Debido a esto Stellantis ha estado invirtiendo en plataformas para la fabricación y distribución ubicadas en regiones estratégicas para poder de esta forma minimizar costos y adaptarse con facilidad a las preferencias locales. Esto ha permitido a Peugeot optimizar su cadena de suministro y reducir su huella de carbono.

Figura 1

Ranking de los modelos de automóviles con mayores ventas en América Latina en 2024



Nota: Marcas de autos que representan la mayor competencia para Peugeot. Fuente: Statista, 2024

Peugeot ha mantenido una presencia sólida en América Latina. Sin embargo, como vemos en la figura 1 de modelos de autos más vendidos en América Latina en los primeros meses del 2024, Peugeot no figura con ninguno de sus modelos dentro de esa lista, esto quiere decir que su cuota de mercado es más baja en comparación con marcas locales y otras compañías internacionales como Volkswagen, Chevrolet, Fiat, afronta también la competencia de marcas chinas que están ganando terreno en el mercado automotriz.

En cuanto a la situación de la venta de autos en latinoamérica:

Según la Asociación Latinoamericana de Distribuidores de Automotores (ALADDA), en el segmento de vehículos livianos, Venezuela, Bolivia, Ecuador y Argentina reportaron las mayores caídas interanuales en junio (35,6%, 30,1% y 25,2% respectivamente). En cambio, Venezuela, Brasil y México fueron los paises que tuvieron mayores incrementos en sus ventas interanuales del 92,0%, 12,6% y 8,3% respectivamente, al compararlo con el año pasado (2024, p.1).

A pesar de la dura competencia y de la clara reducción en las ventas en ciertos países Peugeot está bien establecido y en algunos de ellos figura en el top de las marcas de autos que más ventas tiene.

Figura 2

Venta de vehículos en Ecuador del 2023 y primeros meses del 2024

Nota: Comparación de ventas de vehículos entre los meses de enero y febrero del 2023 y del 2024. Fuente: AEADE, 2024.

Peugeot llegó a Ecuador hace varias décadas, estableciendo su presencia en el mercado a través de distribuidores autorizados y concesionarios en las principales ciudades. La marca ha logrado establecerse con éxito en el país, ofreciendo vehículos que sobresalen por su diseño europeo y una combinación de durabilidad, eficiencia y tecnología que resulta atractiva para los consumidores locales, a pesar de, como

podemos apreciar en la figura 2 las ventas de vehículos nuevos en Ecuador ha ido disminuyendo en comparación a los primeros meses del 2023, esto quiere decir que Peugeot se enfrenta actualmente a un panorama difícil.

En la actualidad Peugeot se ha tomado en serio el fortalecimiento de su red de distribución por medio de la apertura de concesionarios en ciudades con relevancia como lo son Quito, Guayaquil y Cuenca, ciudades en las que se ha visto mayor demanda por los vehículos de la marca. Entre los modelos más populares de la marca en el mercado ecuatoriano se encuentran los SUVs Peugeot 2008 y 3008 que han sido los predilectos por su estética contemporánea y sus características enfocadas en la comodidad y la tecnología.

Asimismo, la Peugeot Partner, una furgoneta que ha destacado en el ámbito comercial, siendo la preferida de pequeñas empresas y emprendedores por su practicidad y eficiencia para adaptarse a las necesidades de los usuarios ya que ofrece comodidad y confort, además de que se puede elegir entre un motor a gasolina o diesel.

En Guayaquil, Peugeot ha logrado ganar terreno, destacándose particularmente con sus SUVs y vehículos comerciales ligeros. El mercado guayaquileño constituye un eje crucial en la actividad comercial y logística por su cercanía a los puertos marítimos y aéreos que se encuentran en dicha ciudad. Por otro lado, en Cuenca, la marca ha aumentado su popularidad debido a sus vehículos compactos que son propicios para moverse por los diferentes sectores de la ciudad, ya que su demanda principal son de carros que sean accesibles y con menor tamaño.

La estrategia de Peugeot ha desplegado en Ecuador también se apega a la sostenibilidad y el desarrollo tecnológico, ya que la marca presenta interés en la incorporación de autos eléctricos e híbridos en un futuro cercano a medida que el cambio sea vea más consolidado en el país. Por otro lado, la empresa está buscando maneras de ajustarse a la creciente demanda de vehículos más eficientes y menos contaminantes ya que en Ecuador se ha observado un interés cada vez mayor por alternativas de transporte más sostenibles.

Problemática

La demanda automotriz en Ecuador ha venido en descenso sobre todo en el 2024, las ventas de vehículos en Ecuador según el reporte de la Asociación de Empresas Automotrices del Ecuador (AEADE), basado en las cifras de ventas del Servicio de Rentas Internas (SRI) cayeron en un 18,6% en el periodo de enero a agosto del 2024

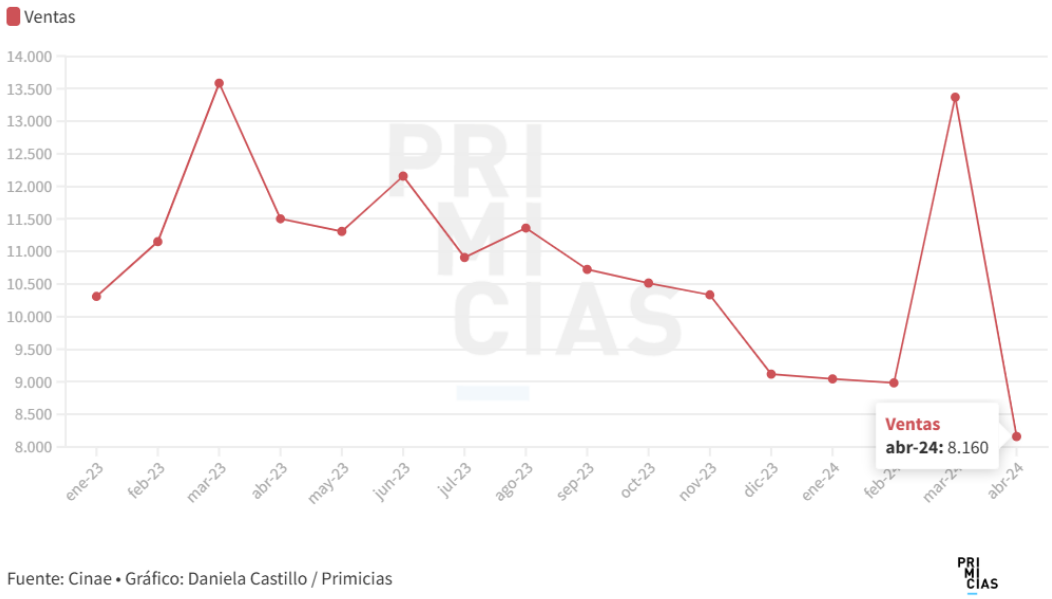
en comparación con el mismo periodo del año 2023. Según la Cámara de la Industria Automotriz del Ecuador (2024), en septiembre de 2024 se vendieron 8.150 carros en Ecuador, siendo la cifra más baja en lo que va del año.

El mes en el que se presentaron las ventas más bajas fue en abril, esto se debe al cambio en la tarifa del Impuesto al Valor Agregado (IVA), que pasó a ser del 15%

En abril de 2024, el mercado automotor en Ecuador vendió 8.160 carros, una caída de 39% frente a marzo de 2024, según estadísticas de la Cámara de la Industria Automotriz Ecuatoriana (Cinae). Esa caída tiene relación con el alza del Impuesto al Valor Agregado (IVA) del 12% al 15%, a partir del 1 de abril, lo que tuvo un impacto importante en el precio de los autos. (González, 2024)

Figura 3

Diagrama de puntos de las ventas durante el primer semestre del 2024



Nota: Baja en las ventas de marzo y abril 2024 debido al incremento del IVA.

Fuente: Cinae/Primicias, 2024.

Por otro lado, el año 2024 ha representado un obstáculo a las ventas de vehículos en Ecuador no solo por el aumento de IVA sino porque también subió el ICE (Impuesto a los consumos especiales), esto se debe a que si aumenta el IVA también aumentará el precio de venta al público y el ICE se calcula a partir de ese PVP.

Esto ocurre porque el incremento del IVA hará que aumente el precio de venta al público (PVP) del bien y el ICE se calcula con base en ese PVP, explica la abogada tributaria Yael Fierro. "Al tener un precio de venta al público más alto, por el alza del IVA, también aumenta la base imponible sobre la que se calcula la tarifa de ICE y, por lo tanto, el consumidor final pagará más por el vehículo" (Fierro, 2024).

El ICE para los carros tiene una tarifa que varía entre el 5% y el 35% y va aumentando conforme al precio del producto, como se explica en el gráfico a continuación.

Figura 4

Tarifa del ICE para vehículos

Tarifa del ICE para vehículos

Vehículos motorizados de transporte terrestre hasta 3,5 toneladas de carga.

Detalle	Tarifa
Vehículos motorizados cuyo precio de venta al público sea de hasta USD 20.000	5%
Camionetas, furgonetas, camiones y vehículos de rescate cuyo precio de venta al público sea de hasta USD 30.000	5%
Vehículos motorizados, excepto camionetas, furgonetas, camiones y vehículos de rescate cuyo precio de venta al público sea superior a USD 20.000 y de hasta USD 30.000	10%
Vehículos motorizados cuyo precio de venta al público sea superior a USD 30.000 y de hasta USD 40.000	15%
Vehículos motorizados cuyo precio de venta al público sea superior a USD 40.000 y de hasta USD 50.000	20%
Vehículos motorizados cuyo precio de venta al público sea superior a USD 50.000 y de hasta USD 60.000	25%
Vehículos motorizados cuyo precio de venta al público sea superior a USD 60.000 y de hasta USD 70.000	30%
Vehículos motorizados cuyo precio de venta al público sea superior a USD 70.000	35%

Tabla: Primicias • Fuente: SRI • Descargar los datos • Creado con Datawrapper

Nota: Porcentaje de ICE que paga un vehículo debido a sus características. Fuente: SRI/Primicias, 2024.

Según CINAE, debido a los cambios en el IVA algunos modelos de carros van a tener una alza en sus precios, dependiendo de sus características, por lo que los carros que más van a aumentar sus precios con aquellos que tengan más cambio de banda tarifaria de ICE, a continuación se explica ejemplos de manera general el cambio según el tipo de carro.

Figura 5

Alza de los precios de los vehículos

Ejemplos del alza de los precios de los vehículos

Precio de venta en USD, según la tarifa de IVA

Vehículo	IVA 12%	IVA 13%	IVA 15%
Camioneta	29 990	33 139	33 726
Auto	19 990	21 129	21 503
SUV	39 990	46 110	46 927

**Los precios incluyen todos los impuestos*

Tabla: PRIMICIAS • Fuente: CINAE • Descargar los datos • Creado con Datawrapper

Nota: Precio de venta como ejemplo según la tarifa de IVA correspondiente. Fuente: CINAE/Primicias, 2024.

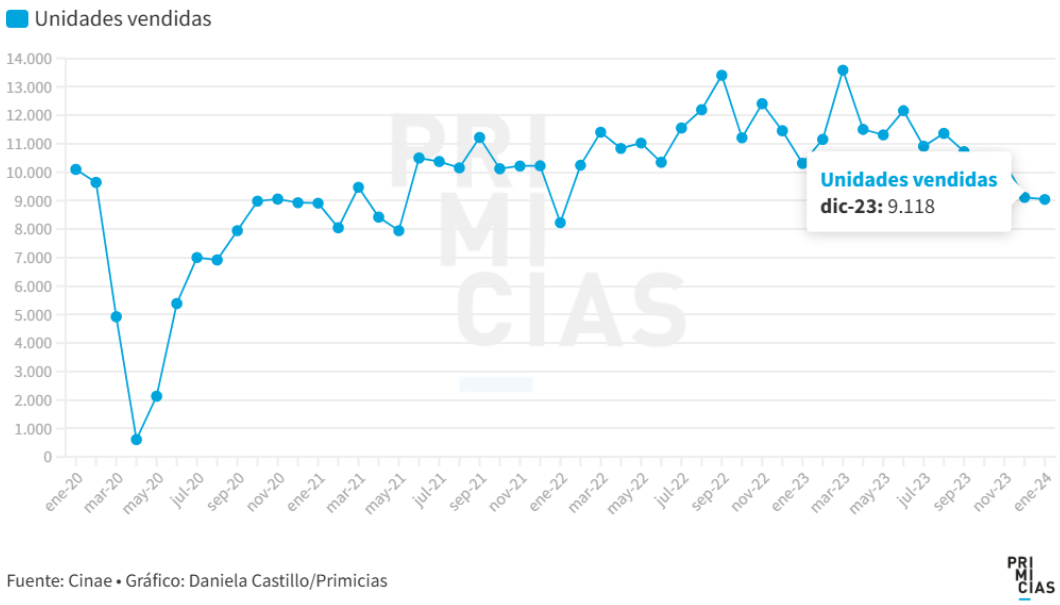
Sin embargo, en el segmento de vehículos híbridos, eléctricos y las motocicletas las ventas continúan en ascenso. En el caso de los carros híbridos y eléctricos, estos siguen marcando una tendencia al alza, se vendieron aproximadamente 7.276 de enero a agosto de 2023, y por el otro lado 9.655 en el mismo periodo pero de este año 2024. Y a pesar de que los vehículos a combustión siguen representando casi el 90 % de las ventas totales, los carros y motos electrificados siguen aumentando en ventas. Se conoce que entre híbridos y eléctricos, este segmento representa el 12, 8 % en este año 2024. Hablando por otro lado de las motocicletas, estas han aumentado en ventas en un 11% este año en comparación a los ocho primeros meses del 2023. En el Ecuador hay actualmente en lo que va del año 112 marcas automotrices, de estas existen marcas que han destacado a lo largo de los años, actualmente en el primer puesto en ventas se encuentra Chevrolet, seguida por Kia en el segundo lugar y Toyota en tercer puesto, Entre las tres abarcan el 40,5% del mercado automotriz. Dentro de este ranking, Peugeot se encuentra en el puesto número 18. ¨Según el Boletín de Ventas de marzo de 2024, Peugeot se ubicó en el puesto 18 en ventas de automóviles y SUV durante enero y febrero de 2024¨. Asociación de Empresas Automotrices del Ecuador (AEADE). (2024).

Durante el 2023 las ventas de vehículos nuevos en Ecuador experimentó un declive en el número de carros que se vendieron ya que pasaron de 132402 a 135250 carros

vendidos durante el 2022, estas cifras fueron incluidas por la Asociación de Empresas Automotrices del Ecuador (AEADE) en su reporte anual.

Figura 6

Gráfico de puntos de unidades de vehículos vendidas



Nota: Unidades de vehículos vendidas desde enero del 2020 al 2024. Fuente: Cinae/Primicias, 2024.

Por otro lado, se ha registrado una caída en las ventas por provincia especialmente en Esmeraldas con el 50%, seguido de Napo, Santo Domingo, Carchi y El Oro. Sin embargo, también se ha visto un crecimiento en el sector vehicular en Zamora Chinchipe, Galápagos, Azuay y Pichincha. Tomando como referencia Azuay las ventas de los automóviles nuevos fueron de 11767 unidades en 2022, en contraste con el 2023 que fueron de 12167 aproximadamente.

Juan Carlos Roldán, presidente ejecutivo del Grupo Roldán, explicó que las cifras que se muestran no son específicamente de ventas sino que se refieren a la matriculación de los vehículos en dicha provincia. También explicó que esto se debe a que los vehículos que en su mayoría se venden en El Oro al menos el 50% se matriculan en

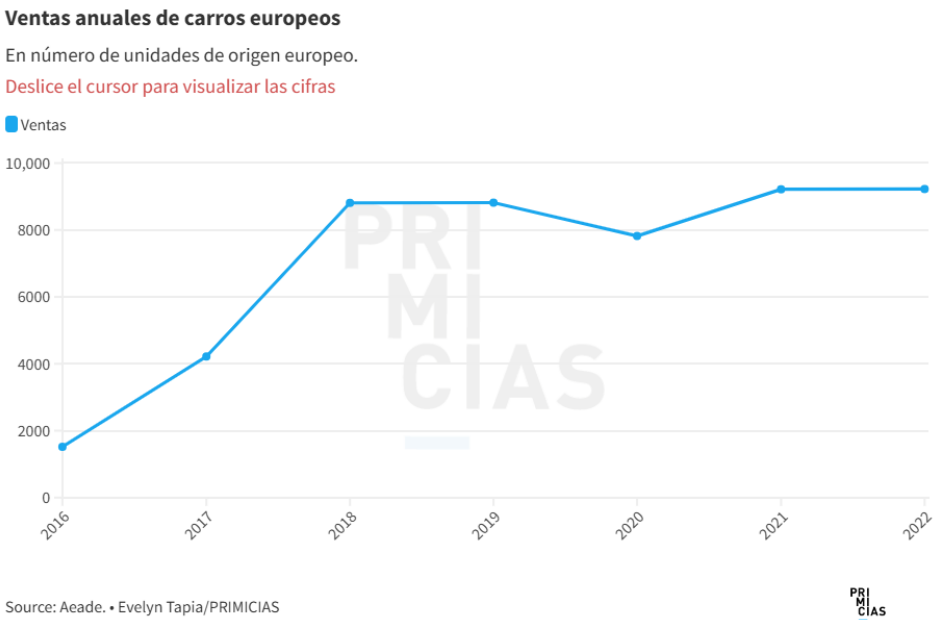
Azuay y eso hace que dichas cifras se reflejen en el portal de Servicio de Rentas Internas.

La explicación a esta situación es que los consumidores prefieren matricular sus vehículos en Azuay debido a la depreciación de un carro matriculado en El Oro, al momento de venderlo. También dio a conocer que las bajas ventas de los vehículos se debe a la mala situación económica en los hogares y empresas ecuatorianas y que para el 2024 la caída de las ventas de los carros se mantenga. Además de que durante el 2024 se ha visto como un año competitivo debido a la inserción de nuevas marcas híbridas en el mercado automovilístico, ya que este tipo de carros no pagan impuestos y se vuelven los preferidos para los consumidores.

En el caso específico de Peugeot en el 2018 se notó un incremento notable en las ventas, aproximadamente las ventas en ese año crecieron hasta un 200% debido al acuerdo comercial con la unión europea que permitió una baja en los aranceles, este acuerdo consistía en que cada año se debía bajar un 5% a los aranceles hasta llegar al 35%.

Figura 7

Ventas anuales de carros europeos en Ecuador



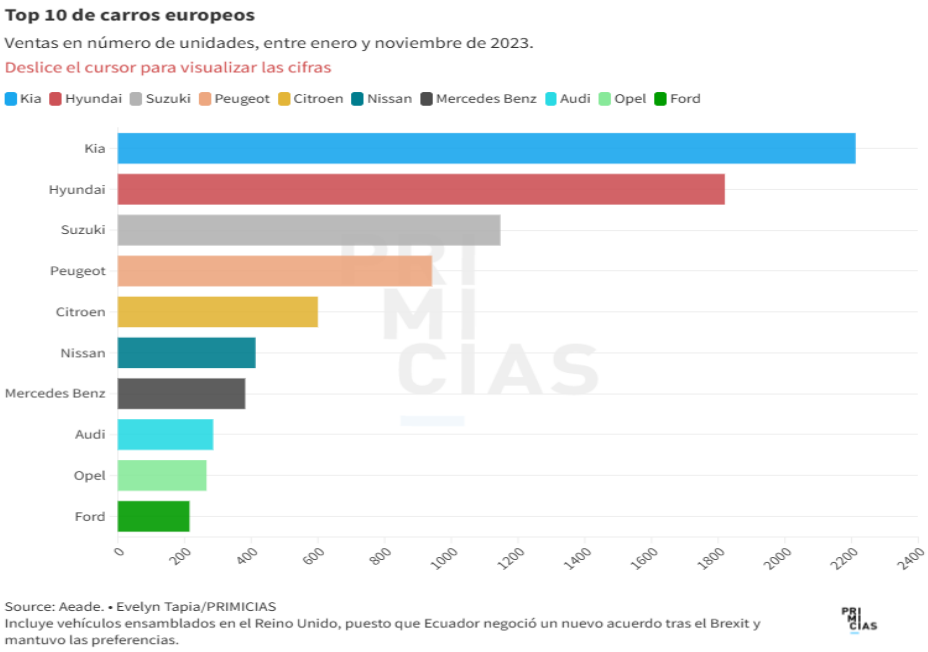
Nota: Ventas de carros europeos en número de unidades desde el 2016 al 2022. Fuente: Aeade/Primicias, 2023.

La marca hizo una negociación con la fábrica que les permitió adelantarse a la bajada del arancel que hubieran tenido en 7 años, gracias a esta negociación Peugeot pudo tener tanto éxito en el 2018, este éxito no solo se debió a sus precios competitivos en el mercado ecuatoriano sino que a su vez la marca de carros aplicó una estrategia comercial agresiva y además introdujo una gama de vehículos diésel, que incluyó modelos como el Peugeot 301, 208, 3008 y la minivan Partner Tepee.

Ya para el 2019 decidieron no enfocarse tanto en los precios debido a que estos ya eran competitivos sino que decidieron enfocarse en ofrecer vehículos altamente equipados e ir subiendo las gamas, que ha sido una característica fija de Peugeot, la calidad de sus vehículos y poder ofrecer vehículos de alta gama. Con esto las ventas en el 2019 volvieron a subir considerablemente, con vehículos como el sedán Peugeot 301 y el SUV Peugeot 3008 que ganaron más popularidad en el mercado en este año, la marca logró seguir creciendo en ventas nuevamente con sus modelos a diesel que resultan altamente atractivos sobre todo en el mercado ecuatoriano debido a sus precios competitivos, logrando posicionarse como el cuarto productor de vehículos europeos presentes en el mercado nacional, como se detalla en el cuadro a continuación.

Figura 8

Top 10 carros europeos en Ecuador



Nota: Top 10 marcas líderes de vehículos europeos en el 2023 entre enero y noviembre. Fuente: Aeade/Primicias, 2023.

En general en el 2019 Peugeot continuó con su incremento en ventas, este año logró ser un año sólido para la marca especialmente en el segmento de SUVs, con modelos como el Peugeot 3008 que lograron destacarse. A pesar de todas estas mejoras y de cada vez aumentar en ventas, Peugeot en estos dos años no logró entrar en el ranking de las marcas líderes en venta de autos en Ecuador.

En el 2020 debido a la pandemia en general las ventas de vehículos tuvieron un decrecimiento siendo los buses y automóviles los que representan una mayor caída con -67% y -48%, respectivamente.

“En el 2020 el sector automotriz evidenció una contracción del 35 por ciento en la venta de vehículos. Cifras de la Asociación de Empresas Automotrices del Ecuador (Aeade) muestran que en este periodo se comercializaron 85.818 unidades, frente a las 132.208 que se vendieron en 2019” (Armijos, 2021).

A pesar de todo esto Peugeot especialmente con el segmento SUVs y camionetas se mantuvo competitivo y a pesar de que en Ecuador sus ventas si disminuyeron, esta disminución no fue tan drástica.

Ya para el 2021 las ventas aumentaron a comparación del 2020, debido a la pandemia del COVID-19 sabemos que las ventas en general bajaron en niveles nunca antes vistos, para el 2021 como ya sabemos la economía empezó a reactivarse, pero a pesar de esto a partir de este año hasta la actualidad las ventas de Peugeot en Ecuador han venido en descenso, se conoce que una de las principales razones es la llegada de los vehículos de origen chino y sus bajos precios. “Como sucede en otros países de América Latina, los vehículos chinos han inundado el mercado ecuatoriano. Según datos de la primera mitad del año, ahora ocupan un 38 % de participación en las ventas” (Aguilera, 2024).

La llegada de los vehículos chinos ha afectado no solo a Peugeot sino a distintas marcas del mercado automotriz, el presidente ejecutivo de la sociedad quien es representante de 130 empresas del sector en el país dijo que actualmente al menos hay 60 marcas de autos chinos presentes en Ecuador lo cual demuestra su diversidad e innovación tecnológica. Se cree que el éxito de estos automóviles se debe además de sus bajos costos a su red de distribución y a la cobertura del 99% que tienen dentro del país. Además de este otros factores que han influido para que las ventas no solo de Peugeot sino de algunas marcas de automóviles disminuyan en Ecuador han sido factores como: la crisis económica y la inestabilidad política, todo esto ha influido negativamente a la marca. Han existido también factores positivos como acuerdos comerciales y baja en aranceles, asimismo como esfuerzos de la marca por aumentar sus ventas, pero todo esto no ha sido suficiente para contrarrestar la disminución de la demanda de vehículos.

Como antes se había mencionado muchos fueron los factores que incidieron en el declive de las ventas de automóviles en el mercado ecuatoriano, y uno de ellos fue el cambio en los impuestos que se pagan por los carros importados y el incremento del IVA, una medida que fue tomada a principios del año 2024. En este contexto, el sector del automóvil también se muestra cauteloso con el aumento del impuesto del 12% al 15%. David Molina, director ejecutivo de la Cámara Ecuatoriana de la Industria Automotriz (Cinae), dijo que el aumento de impuestos tendrá un mayor impacto en el sector.

Además de esta existen otras razones por la cual la marca se ha visto afectada, por ejemplo en el 2023 existieron problemas y fallas en motores PureTech en algunas marcas de automóviles entre estas Peugeot, en este año varias personas presentaron problemas significativos con este tipo de motor, entre estos la corrosión de la cadena de distribución que se daba gracias a la mezcla de combustible y aceite, expertos en la materia señalan que este tipo de problemas pueden llevar a fallas graves en el motor, debido a esto existieron demandas masivas en contra de esta y otras marcas que usaban este tipo de motor.

Como estos se han presentado varios problemas que afectan la reputación de la marca y esto a su vez hace que las ventas bajen, por ejemplo en páginas donde los usuarios presentan sus experiencias como propietarios de distintos tipos de autos, el modelo de Peugeot 2008 a pesar de ser uno de los más populares y más vendidos ha presentado algunos problemas según relatan los clientes, problemas con la caja de cambios o el consumo excesivo de aceite. Opiniones como estas también influyen en la disminución de ventas y afectan el prestigio de Peugeot.

Justificación

Hoy en día la inteligencia artificial está presente en muchos aspectos de la vida cotidiana, desde las actividades más sencillas como hacer tareas escolares hasta la recopilación de datos de cualquier tipo.

Actualmente la Inteligencia Artificial es un área de la ciencia de gran interés por ser un área multidisciplinaria donde se realizan sistemas que tratan de hacer tareas y resolver problemas como lo hace un humano, asimismo se trata de simular de manera

artificial las formas del pensamiento y cómo trabaja el cerebro para tomar decisiones (Ponce et al., 2014, p. 11).

No obstante, a diferencia de los humanos, los aparatos basados en Inteligencia Artificial no requieren descanso y pueden examinar grandes cantidades de información simultáneamente. Igualmente, la tasa de errores es considerablemente inferior en las máquinas que llevan a cabo las mismas funciones que sus homólogos humanos.

Dentro de la inteligencia artificial encontramos al aprendizaje automático (machine learning) el cual se ha vuelto muy importante para la sociedad hoy en día. Por esta razón, el machine learning se ha vuelto un término necesario de entender para poder solucionar problemas que se presentan dentro de las empresas o simplemente para ayudar con el aumento de las ventas de la misma. Bajo este contexto el Machine Learning se define como: ¨un método de análisis de datos que se nutre del Big Data. Es una de las ramas de la Inteligencia Artificial, ya que su propósito es crear modelos que aprenden automáticamente¨ (Salvador, 2019).

Este trabajo de investigación se está desarrollando con la finalidad de crear un modelo que ayude a aumentar las ventas de la empresa Peugeot por medio de la inteligencia artificial, específicamente con algoritmos usados en machine learning como K-Nearest Neighbors (KNN) y Forecasting, cada uno de estos definidos de la siguiente manera: ¨K-means es una técnica de aprendizaje no supervisada, donde el objetivo de la técnica es determinar la definición de la respuesta correcta mediante la búsqueda de agrupaciones o clusterización de los datos¨ (L. Yu et al., 2017). Y por otro lado el forecasting implica ¨el uso de datos pasados para predecir valores futuros¨ (Hyndman et al., 2018). También se podría definir como:
El proceso de estimar, calcular o predecir el valor futuro de una variable de interés, utilizando información pasada y presente, así como la identificación de patrones y tendencias en los datos disponibles. Este enfoque combina métodos cualitativos y cuantitativos para proporcionar estimaciones precisas y confiables, permitiendo a las organizaciones anticipar cambios y planificar estrategias efectivas (Makridakis et al., 1998).

Después de realizar un análisis exhaustivo de las ventas y proyectarlas a futuro para ver el comportamiento de estas en los próximos años, además de haber hecho la correspondiente segmentación de clientes con la ayuda del algoritmo K-means neighbors, se usará esta información para poder buscar las correctas técnicas de marketing que ayuden a incrementar las ventas de Peugeot Ecuador.

El marketing se ocupa del proceso de planificar las acciones de la empresa en términos de precio, promoción, distribución y venta de productos y servicios que brinda, con la ayuda de una buena segmentación de mercado que garantice la aceptación de sus productos, para satisfacer las necesidades y anhelos de los clientes y observar el desarrollo de la empresa para que resulte muy lucrativa. Es por esto que el marketing se considera un componente esencial dentro de la empresa.Solo a través de una correcta interpretación del mercado es posible lograr un producto que sea promocionado de manera efectiva, que satisfaga las necesidades de los clientes a un precio esté al alcance del consumidor en el momento y el lugar adecuado (Álvarez et al., 2020).

Dentro de esta investigación el problema central que existe es la baja en las ventas de autos en Ecuador, esta industria está enfrentando varios desafíos debido a la crisis por la que atraviesa el país y no solo a la crisis sino que también año tras año incrementa la competencia en esta industria, fue por esto que es notable la necesidad de abordar este tema para poder optimizar los procesos y las ventas de Peugeot para así poder en un futuro contribuir a mejorar este problema en esta empresa en específico.

Actualmente, se ha determinado que las empresas no se dedican a hacer un análisis profundo de sus ventas, debido a esto muchas de ellas presentan bajas en las mismas, como se ha observado esto es un problema muy común en este país, por esta razón se decidió por medio de este trabajo de investigación poder demostrar que si se hace un buen análisis y una correcta segmentación de clientes, junto con las correctas estrategias de marketing dado el análisis previo, se podría mejorar las ventas de una empresa.

Se busca también responder a la necesidad actual de poder entender de mejor manera los patrones de compra de los ecuatorianos específicamente en la compra de autos Peugeot, para así poder identificar qué características específicas tienen sus clientes

potenciales y con esto poder atraer potenciales consumidores de este segmento y también poder trabajar en ampliar la cartera de clientes fieles a la marca.

"En el caso del departamento de mercadeo, no podría generar estrategias sin los datos de los clientes, las características de los productos o servicios, y más aún, sin los precios" (Peña, 2017, p. 20). A manera de resumen, el análisis de datos como ya se mencionó anteriormente no solo es importante sino que también se considera indispensable, tanto que expertos dicen que no se podrían generar correctas estrategias sin los datos de los clientes y un análisis previo de estos, con este trabajo de investigación se podrá resaltar la importancia del machine learning y del marketing dentro de las empresas.

Por otro lado, en un mundo globalizado es necesario poder ser competitivo en un mundo donde las relaciones comerciales están interconectadas, dejando de lado algunos procesos que se pueden considerar ambiguos y que puede traducirse en pérdidas económicas, de clientes y de competitividad, por lo que con esta investigación se pretende demostrar que la necesidad de aplicar machine learning en las empresas es imperativo para lograr obtener una ventaja competitiva frente a las demás marcas presentes en el mismo mercado.

"El uso de machine learning en el campo de las ventas ya no es una opción, es una necesidad. Las empresas que se dedican al análisis de datos y que emplean herramientas de machine learning para predecir las necesidades de los consumidores, automatizar la generación de leads o determinar el momento óptimo para hacer una oferta, se encuentran en una posición mucho más competitiva que aquellas que siguen utilizando métodos tradicionales" aclara Thomas H. Davenport, Profesor en la Escuela de Negocios de la Universidad de Harvard y experto en Analytic, en donde se aclara que es una necesidad poder usar el análisis de datos para desarrollar un ambiente competitivo en la empresa.

Alcance

Este trabajo de investigación se está haciendo con el propósito de buscar distintas soluciones para la empresa Peugeot haciendo uso del machine learning, ya que se ha visto que las ventas de Peugeot en Ecuador han ido disminuyendo a pesar de que la empresa no ha sido por lo general una de las marcas líderes en ventas de carros en el país.

Es por esto que este trabajo de investigación va dirigido a los directivos de la empresa Peugeot usada para el presente análisis y a distintas áreas de esta misma empresa que se pueden beneficiar de la información, como lo son: el área de marketing, ventas y operaciones. Asimismo va dirigido a estudiantes, maestros o profesionales interesados en la investigación de factores influyentes en las ventas de autos, no solo de autos Peugeot sino también de otras marcas de vehículos existentes en Ecuador.

Dentro de esta investigación se ha podido evidenciar un gran problema dentro de las empresas y es que no se le da correcto uso a los distintos recursos que ofrece la avanzada tecnología hoy en día como lo es el machine learning, que ayuda a más de un departamento de la empresa, como ya se mencionó gracias al análisis de datos se podrían optimizar los procesos de las empresas automovilísticas y como en el presente caso de investigación optimizar las ventas, pensando en técnicas para cubrir necesidades y para robustecer ventajas competitivas de la empresa.

Existen varios recursos bastante útiles dentro de esta investigación, por ejemplo por medio de este análisis los expertos encargados de ventas de la empresa podrían ver cuáles son los segmentos de clientes que más ingresos le aportan a la empresa, cuál es el segmento más predominante para Peugeot y con esto hacer estrategías adecuadas dirigidas a estos grupos específicos y asimismo formar campañas dirigidas a los segmentos donde Peugeot no tiene tanta acogida.

El análisis de datos es fundamental para la optimización de los procesos empresariales, mejoras en rendimiento de los diferentes departamentos que a su vez se ve traducido en el aumento de la satisfacción de los clientes y del posicionamiento de las marcas. La incorporación de las metodologías del Machine Learning se ha vuelto vital para el funcionamiento óptimo de las compañías ya que puede ser aplicable a múltiples

contextos y con diferentes enfoques, aumentar calidad, mejorar la productividad, desarrollar una buena reputación en el mercado, por esta razón este artículo resulta provechoso para los empresarios y ejecutivos que estén búsqueda de información sobre la implementación de estas técnicas. Este proyecto de investigación también puede servir de ejemplo y modelo para la resolución de los problemas empresariales que se pueden presentar en el día a día.

Por otro lado, aunque el uso adecuado de estos metodologías de Machine Learning pueden ser un poco complejas de entender y también representan un costo significativo para las empresas, la incorporación del mismo se podría considerar como una inversión ya que se ha vuelto necesario para la innovación de las empresas debido a que gracias a estas técnicas se pueden reducir la probabilidad de errores organizacionales que pueden conllevar a una mala administración, además de que esta aplicación es un paso importante hacia la automatización de los procesos empresariales que se han visto afectados por la ola de la tecnología y globalización.

Objetivos

- **Objetivo General**

- Desarrollar un modelo de clasificación y de predicción para la evaluación del rendimiento y optimización de las ventas de autos de marca Peugeot en el Ecuador.

- **Objetivos Específicos**

- Exponer, explicar y entender los términos con respecto a las teorías, conceptos y el contexto legal del Machine Learning aplicado en un entorno empresarial.
- Analizar y evidenciar el uso de los diferentes algoritmos de Machine Learning para la clasificación de segmentos de mercados y pronóstico de ventas.
- Explicar el modelo de K-means y Forecasting usados en las ventas de Peugeot para la obtención de resultados aplicables en un contexto empresarial real.

Marco teórico

- **Machine learning**

"El Machine Learning o aprendizaje automático es un campo científico y, más particularmente, una subcategoría de inteligencia artificial. Consiste en dejar que los algoritmos descubran «patterns», es decir, patrones recurrentes, en conjuntos de datos." (DataScientest, 2022)

"El machine learning, a menudo abreviado como ML, es un subconjunto de la inteligencia artificial (IA) que se centra en el desarrollo de algoritmos informáticos que mejoran automáticamente mediante la experiencia y el uso de datos". (Datacamp, 2024)

- **Teorías de la Inteligencia Artificial**

Teoría de la computabilidad

Esta teoría estudia qué problemas pueden ser resueltos por una computadora. Entre sus elementos se incluyen:

- Máquina de Turing: es un modelo que define que es computable
- Problemas decidibles e indecidibles: clasifica los problemas según su forma de resolución.
- Límites computacionales: restricciones fundamentales en la capacidad de cálculo.

Teoría de la complejidad

Esta teoría analiza la eficiencia de los algoritmos y clasifica los problemas según cuales son los requisitos para ser computacionales

- Clasificación de problemas
- Eficiencia algorítmica
- Optimización computacional

Teoría de la comunicación

Esta teoría establece las bases para la codificación y transmisión de datos.

- Cuantifica la información
- Métodos para representación eficaz de datos
- Límites en la transmisión de información

Teoría de las probabilidades

Esta teoría es fundamental para manejar la incertidumbre de los modelos de IA.

- Probabilidad bayesiana: actualiza teorías basadas en nueva evidencia.
- Redes bayesianas: representa dependencia probabilísticas
- Procesos estocásticos: modela sistemas que evolucionan en el tiempo

Lógica matemática

Esta rama es esencial para el razonamiento artificial y la representación de conocimiento.

- Lógica proposicional: representa y manipula proposiciones mediante operadores lógicos
- Lógica de predicados: rama de la lógica proposicional que permite expresar relaciones y cuantificadores.
- Lógica temporal: razone sobre secuencias de eventos y estados de tiempo.

- **Sets de entrenamiento, validación y prueba**

Al iniciar un proyecto que incluya Machine Learning se necesita tener que elegir entre múltiples modelos y tomar la decisión de cuál de ellos va a tener resultados que proporcionen predicciones mucho más precisas. Este procedimiento implica el entrenamiento de los datos y posteriormente evaluarlos con la finalidad de identificar cuál de ellos muestra un mejor rendimiento y que se puede asegurar que el modelo que se haya escogido sea el adecuado para los tipos de datos que se están utilizando. Para hacer este procedimiento, se hace uso de tres conjuntos de datos, los datos de entrenamiento, validación y prueba.

Encontrar los parámetros de cada modelo

Si se utiliza todo el conjunto de datos, podría surgir un inconveniente, ya que el objetivo no es solo entrenar el modelo, sino también evaluarlo en un entorno real. En otras palabras, una vez que el modelo ha sido entrenado, es imprescindible que se le presenten datos que no haya visto anteriormente para poder determinar si es capaz de clasificar correctamente a los sujetos. Por lo tanto, en lugar de entrenar los modelos candidatos con todos los datos disponibles, se procede a realizar una primera partición: por ejemplo, se selecciona aleatoriamente un 70% de los datos y se les ofrece a cada modelo durante el entrenamiento para que pueda calcular sus parámetros, mientras que el 30% restante de la base de datos se reserva para su uso posterior. A este conjunto de datos utilizado para estimar los parámetros durante el entrenamiento de cada modelo se le denominará, precisamente, el conjunto de entrenamiento.

Encontrar los hiperparámetros de cada modelo y elegir el mejor modelo

En primer lugar, si se ajustan los hiperparámetros de cada modelo utilizando el mismo conjunto de entrenamiento, podríamos enfrentar una situación de sobreajuste (o overfitting): cada modelo tenderá a "memorizar" los datos de entrenamiento, lo que podría dar la impresión de un buen rendimiento, pero en realidad no lograría realizar buenas clasificaciones cuando se le presenten datos que nunca ha visto antes. Por lo tanto, la manera adecuada de optimizar los hiperparámetros de cada modelo y de seleccionar el modelo más apropiado es utilizando un conjunto de datos que no haya sido revisado previamente por ninguno de los modelos. Siguiendo con el ejemplo, se había reservado el 30% de los datos. Así que tomaremos este subconjunto, lo mezclaremos aleatoriamente y lo dividiremos en dos partes. Finalmente, tomaremos la mitad de estos datos (es decir, el 15% del conjunto de datos total) que se utilizará para afinar los hiperparámetros y para seleccionar el modelo más adecuado. Este subconjunto se conoce precisamente como el conjunto de validación y permitirá obtener los mejores hiperparámetros de cada modelo, así como seleccionar el mejor entre todos los que se hayan entrenado, de una manera objetiva.

Poner a prueba el modelo seleccionado

La parte final de este proceso es la de poner a prueba el modelo con los datos que no ha visto posteriormente para poder determinar así cual sería su desempeño en situaciones reales. Este conjunto de datos se conoce como datos de prueba. Es necesario hacer énfasis que dicho conjunto de datos debe permanecer oculto en todo momento durante el proceso de entrenamiento y solo pueden ser revelados al final del proceso, cuando ya la fase entrenamiento haya finalizado con el modelo que se adapte mejor a las necesidades y se hallan ajustados los hiperparámetros. Caso contrario, el resultado que se obtenga puede ser poco confiable al momento de poner a prueba en un contexto real, lo que podría afectar la veracidad del modelo al hacer predicciones con datos nuevos.

- **Modelo Supervisado**

Modelos supervisados: un modelo supervisado se caracteriza principalmente por el entrenamiento de los datos ya que estos tienen una "etiqueta" conocida. Estos modelos tienen dos variables, las variables de entrada, la X, y una variable de salida, es decir la Y. El objetivo principal de los modelos supervisados es entrenar los datos para que aprendan una relación entre ambas variables.

"El aprendizaje supervisado, también conocido como aprendizaje automático supervisado, es una subcategoría del aprendizaje automático y la inteligencia artificial". (IBM, s.f.)

"Se define por su uso de conjuntos de datos etiquetados para entrenar algoritmos que clasifiquen datos o predigan resultados con precisión". (IBM, s.f.)

"El machine learning o aprendizaje automático es una rama de la inteligencia artificial que, a partir de algoritmos matemáticos, logra que las máquinas puedan aprender de una forma similar a la que lo hacemos los humanos y de realizar análisis sin que hayan sido explícitamente programadas para ello". (Emilio, s.f.)

- **Ejemplos de modelos supervisados:**

1. **Regresión:**
   - Regresión Lineal
   - Regresión Logística
2. **Clasificación:**
   - Árboles de Decisión (Decision Trees)
   - Bosques Aleatorios (Random Forest)
   - Máquina de soporte vectorial (SVM)
   - k-Nearest Neighbors (k-NN)
   - Redes Neuronales (en contextos supervisados)

- **Modelo no supervisado:**

En los modelos no supervisados, los datos se entrenan usando datos sin "etiqueta", es decir que no tienen una variable de salida como en los modelos supervisados. Las principales características de estos modelos es que solo utilizan una variable de entrada, la X, que no tiene una relación establecida. El objetivo principal es el de encontrar patrones, agrupaciones o estructuras en los datos del tratamiento.

- *"Es una técnica de machine learning en la que los modelos no aprenden a partir de los llamados datos de entrenamiento. Son los propios modelos sin supervisión los que encuentran los patrones subyacentes en los datos a analizar." (INESDI, 2022)*

- *"Los algoritmos de Aprendizaje no Supervisados se utilizan para agrupar los datos no estructurados según sus similitudes y patrones distintos en el conjunto de datos. El término no supervisado se refiere al hecho de que el algoritmo no está guiado como el algoritmo de Aprendizaje Supervisado" (Gonzalez, 2020).*

-

- **Ejemplos de modelos no supervisados:**

1. **Clustering (Agrupación):**
   - K-means
   - Clustering jerárquico

- DBSCAN (Density-Based Spatial Clustering)

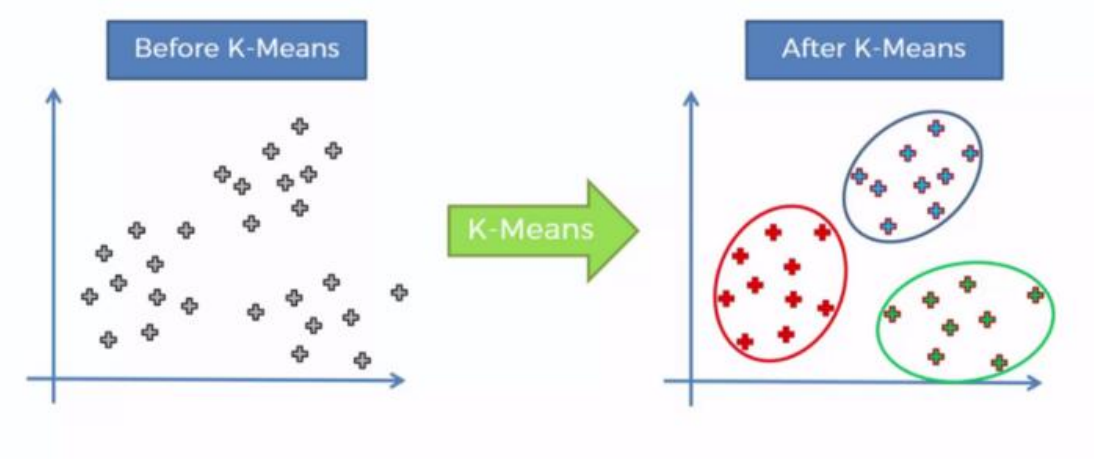
2. **Reducción de Dimensionalidad:**
   - PCA (Principal Component Analysis)
   - t-SNE (t-Distributed Stochastic Neighbor Embedding)
   - Autoencoders (cuando no tienen supervisión)
3. **Asociación:**
   - Algoritmo Apriori
   - Algoritmo ECLAT

- **Concepto de K-means (Clustering)**

K-means es un algoritmo de aprendizaje no supervisado, cuyo objetivo principal es la agrupación de datos en un número predefinido de grupos o clusters. Su uso principal es organizar los datos en clusters de forma que los elementos que corresponde a un grupo en específico tengan características similares entre sí y que se diferencien de los demás clusters.

Figura 9

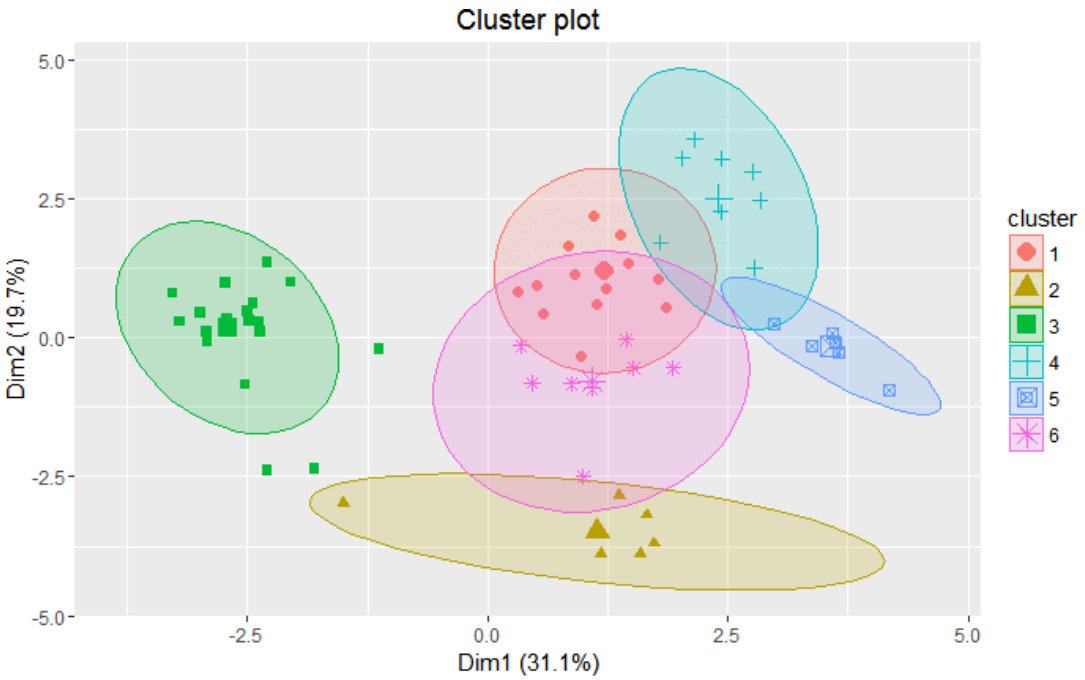
Representación gráfica del algoritmo de K-means



Nota: Clasificación de los datos antes y después de usar K-Means. Fuente: Exponentis, 2019.

Figura 10

Representación gráfica de K-means utilizando R Studio



Nota: Ejemplo de resultados de K-means en el programa de R Studio. Fuente: ResearchGate, 2016.

- **Pasos principales del algoritmo:**

1. Elegir el número de clusters **k**.
2. Asignar aleatoriamente **k centroides** (puntos representativos de cada cluster).
3. Asignar cada punto de datos al cluster cuyo centroide esté más cercano (minimizando la distancia, por lo general se usa la distancia euclidiana).
4. Calcular los centroides como el promedio de los puntos asignados a cada cluster.
5. Repetir los pasos 3 y 4 hasta que la distancia entre los centroides no cambie significativamente o se alcance un número máximo de iteraciones.

- **Concepto de Forecast (Pronóstico)**

El modelo Forecast es un método de aprendizaje supervisado o de series de tiempo que se usa para hacer pronósticos, lo que busca es predecir valores futuros basándose en datos anteriores. El objetivo es identificar patrones, tendencias y estacionalidades en los datos para poder hacer estimaciones más precisas sobre lo que se quiere predecir.

Tipos de modelos predictivos más comunes:

1. **Modelos estadísticos:**
   - ARIMA (AutoRegressive Integrated Moving Average)
   - SARIMA (ARIMA con componente estacional)
   - Suavización exponencial

Para poder predecir dentro de estos modelos hay varios métodos que se pueden usar cuando se presenten ciertos casos específicos, algunos de estos son:

- **Promedio simple**

En este método todas las demandas de los precios anteriores tienen el mismo precio relativo. El promedio hace que las demandas elevadas tiendan a ser equilibradas por las demandas bajas de otros periodos, reduciendo las posibilidades de error que se podrían cometer al dejarse llevar por fluctuaciones aleatorias que pueden ocurrir en un periodo (Hernández et al., 2004).

- **Promedios móviles**

Cuando las series no tienen componente estacional, y consisten únicamente en los componentes de tendencia e irregularidad, la descomposición de la serie involucra únicamente la estimación de ese componente de tendencia, éste se puede estimar por medio de la suavización reduciendo la variación aleatoria, el modelo más básico de suavización se denomina promedios móviles (Hernández et al., 2004).

2. **Modelos basados en Machine Learning:**
   - Redes Neuronales Recurrentes (RNNs)
   - Long Short-Term Memory (LSTM)
   - Gradient Boosting Machines (e.g., XGBoost, LightGBM) en datos temporales.

- **Series de tiempo**

“Una serie de tiempo es una secuencia de N observaciones ordenadas y equidistantes cronológicamente sobre una característica denominada serie univariante o sobre varias características denominada serie multivariante de una unidad observable en diferentes momentos” (Mauricio, 2007). Estas series de tiempo pueden ser de dos tipos, ya sean estacionarias, lo que quiere decir que es estable a lo largo del tiempo o no estacionaria.

■ **Clasificación de series de tiempo:**

- Estacionarias.- "Una serie es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo" (Villavicencio, 2022).
- No estacionarias.- "Son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante" (Villavicencio, 2022).

Asimismo las series de tiempo tienen los siguientes componentes:

- Tendencia: "Representa el comportamiento predominante de la serie. Esta puede ser definida vagamente como el cambio de la media a lo largo de un extenso período de tiempo" (Rios et al., 2008).
- Ciclo: "Caracterizado por oscilaciones alrededor de la tendencia con una larga duración, y sus factores no son claros. Por ejemplo, fenómenos climáticos, que tienen ciclos que duran varios años" (Rios et al., 2008).
- Estacionalidad: "Es un movimiento periódico que se produce dentro de un periodo corto y conocido. Este componente está determinado, por ejemplo, por factores institucionales y climáticos" (Rios et al., 2008).
- Aleatorio (A): "Son movimientos erráticos que no siguen un patrón específico y que obedecen a causas diversas. Este componente es prácticamente impredecible. Este comportamiento representa todos los tipos de movimientos de una serie de tiempo que no son tendencia, variaciones estacionales ni fluctuaciones cíclicas" (Rios et al., 2008).

■ **Medición del error:**

Para poder evaluar el desempeño de este modelo de predicción se usan diversos indicadores, estos tienen la función de ver que tan próximos están los valores pronosticados de las series originales. Uno de los métodos más utilizados es la Raíz Cuadrática Media del Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (Y_t^s - Y_t^a)}^2 \quad (1)$$

Y_t^a:valor pronosticado de Yt.

Y_t^s:valor real de Yt.

T: número de periodos.

- **Procesos lineales estacionarios:**

  - **Modelos autorregresivos AR(p)**

"Los modelos autorregresivos se basan en la idea de que el valor actual de la serie, X_t, puede explicarse en función de p valores pasados X_{t-1}, X_{t-2},....., X_{t-p}, donde p determina el número de rezagos necesarios para pronosticar un valor actual"(Villavicencio, 2010).

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots + \phi_p X_{t-p} + \varepsilon_t \quad (2)$$

ε_t: Ruido blanco

  - **Modelos de media móvil MA(q)**

"En los modelos de media móvil, el proceso se representa como una suma ponderada de errores actuales y anteriores. El número de rezagos del error considerados (q) determina el orden del modelo de media móvil" (Rios et al., 2008).

$$X_t = \theta_0 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots\ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

  - **Proceso Autorregresivo de Medias Móviles (ARMA) (p,q)**

"En estos modelos, el proceso se representa en función de observaciones pasadas de la variable y de los valores actuales y rezagados del error. El número de rezagos de la variable de interés (p) y el número de rezagos del error (q) determinan el orden del modelo mixto"(Rios et al., 2008).

$$X_t = c + \phi_1 X_{t-1} + \ldots\ldots + \phi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4)$$

Los modelos ARMA siempre van a compartir las características de los modelos AR(p) y MA(q), esto es porque contiene a ambas estructuras a la vez. El modelo tiene media cero, varianza constante y finita y una función de autocorrelación infinita. La función de autocorrelación es infinita decreciendo rápidamente hacia cero (Villavicencio, 2010).

- **Proceso Autorregresivo Integrado y de Media Móvil ARIMA (p,d,q)**

"Muchas series de tiempo no son estacionarias, por ejemplo el Producto Nacional Bruto o la Producción Industrial. Un tipo especial de series no estacionarias, son las no estacionarias homogéneas que se caracterizan porque, al ser diferenciadas una o más veces, se vuelven estacionarias" (Rios et al., 2008).

$$X_t^d = c + \phi_1 X_{t-1}^d + \ldots\ldots + \phi_p X_{t-p}^d + \theta_1 \varepsilon_{t-1}^d + \theta_2 \varepsilon_{t-2}^d + \ldots\ldots + \theta_q \varepsilon_{t-q}^d + \varepsilon_t^d \qquad (5)$$

- **Algoritmo y Ecuaciones de k-Means**

El algoritmo de **k-Means** se centra en minimizar la distancia intra-cluster (dentro de cada grupo) y maximizar la distancia inter-cluster (entre grupos). Pasos del Algoritmo de k-Means

1. **Iniciación**: Elegir K centroides iniciales, ya sea aleatoriamente o mediante una heurística.
2. **Asignación de Clusters**: Cada punto de datos Xi se asigna al cluster con el centroide más cercano:

$$Ci = arg\ min_j \left| x_i - \mu_j \right|^2 \qquad (6)$$

Ci: es el cluster asignado al punto Xi

Ui: el centroide del cluster j

$\|Xi-Ui\|^2$= la distancia euclidiana al cuadrado entre Xi y Ui

3. Actualización de Centroides: Calcular el nuevo centroide para cada cluster como el promedio de los puntos asignados al mismo:

$$\mu_j = \frac{1}{|Cj|} \sum_{Xi \epsilon Cj} X_i \quad (7)$$

μ_(j):el nuevo centroide cluster j

Cj= el conjunto de puntos en el cluster

|C_J |:el número de puntos en el cluster j

4. Convergencia: Repetir los pasos 2 y 3 hasta que los centroides no cambien significativamente o se alcance un número máximo de iteraciones.

- ***Función Objetivo (Criterio de Optimización):***

K-Means busca minimizar la suma de distancias cuadradas dentro de los clusters:

$$J = \sum_{j=1}^{k} \sum_{Xi \epsilon Cj} \left|\left| \chi_i - \mu_j \right|\right|^2 \quad (8)$$

Métodos para minimizar la distancia de los centroides:

- **Distancia Euclídea**

"La distancia euclídea entre dos puntos p y q se define como la longitud del segmento que une ambos puntos. En coordenadas cartesianas, se calcula empleando el teorema de Pitágoras." (Valladolid, C., 2020)

Ecuaciones:

$$\mathbf{d\ (A,B) = \sqrt{(X_B - X_A)^2 + (Y_B - Y_A)^2}} \quad (9)$$

- **Distancia Manhattan:**

"A diferencia de la distancia euclidiana, que mide la línea más corta posible entre dos puntos, la distancia de Manhattan mide la suma de las diferencias absolutas entre las coordenadas de los puntos."(Gower,J.C. 2017)

La distancia Manhattan entre dos puntos $P_1 = (x_1, y_1)$ y $P_2 = (x_2, y_2)$ en un espacio bidimensional se calcula como:

Distancia Manhattan $= |x_1 - x_2| + |y_1 - y_2|$ (10)

Si se trata de un espacio de más dimensiones entre dos puntos p = (p1,p2, . . . ,pn) y q=(q1,q2, . . . ,qn) en un espacio de n-dimensional.

Distancia Manhattan $= \sum_{=!} |pi - qi|$ (11)

- **Escalamiento de datos:**

El objetivo del escalamiento de datos es lograr la estandarización de los mismos, lo que se busca es que todos los datos pertenecientes a un conjunto determinado tengan valores similares con la finalidad de que cada uno de los valores ayuden a calcular la distancia de manera equitativa evitando así que los datos con mayor valores predominen en el cálculo y pertenezcan a un rango en específico.

- **Normalización de datos:**

El objetivo principal de la normalización de datos es tratar los datos pertenecientes al conjunto para que sigan una distribución en específico, este tratamiento puede ser hecho para que siga una distribución gaussiana u otra, esto con la finalidad de que los datos puedan ser interpretados con mayor facilidad por el investigador.

- **Algoritmos y Ecuaciones de Forecasting (Pronóstico)**

El pronóstico puede usar métodos estadísticos o modelos más avanzados de machine learning. Aquí te detallo algunos métodos populares y sus ecuaciones:

- **1. ARIMA (AutoRegressive Integrated Moving Average)**

ARIMA modela datos temporales considerando tres componentes:

- **AR**: Componente autorregresivo (Autoregressive).
- **I**: Diferenciación para hacer los datos estacionarios (Integrated).
- **MA**: Componente de promedio móvil (Moving Average)

$$t=c+\phi 1ydt-1+\phi 2ydt-p+...+\phi pydt-p+\theta 1et-1+\theta 2et-2+...+\theta qet-q$$

(12)

Donde: p es la cantidad de términos autorregresivos, d la cantidad de veces que la serie tiene que ser diferenciada para que sea estacionaria y q el número de términos de la media móvil.

■ 2. Modelos de Suavización Exponencial

Los métodos de Holt-Winters son comunes para datos con tendencia y estacionalidad.

El modelo de Holt-Winters es un método de suavización exponencial diseñado para realizar pronósticos de series temporales que presentan tanto tendencia como estacionalidad. Es una extensión del modelo de suavización exponencial simple.

■ Componentes del Modelo:

1. **Nivel (Lt)**: Representa el valor promedio de la serie temporal en el tiempo T.
2. **Tendencia (Bt)**: Captura la tasa de cambio o la dirección general (crecimiento o decrecimiento) en la serie.
3. **Estacionalidad (St)**: Mide las fluctuaciones recurrentes que ocurren en intervalos regulares (diaria, mensual, anual, etc.).
4. **Pronóstico (Yt+h)**: Es el valor proyectado para un horizonte futuro de h períodos, basado en la combinación de los componentes anteriores.

Ecuación general del modelo Holt-Winters:

La fórmula general del pronóstico es:

D t, t+1 = (at +T bt) + F t+T-P (13)

Dónde:

D = Demanda o variable a estimar

a = Promedio de ventas

b = Representa la tendencia

F = Factor de estacionalidad

t = Período actual

T = Número de períodos en adelante que se quiere proyectar

- **Usos de k-Means y Forecast**

  - **1. k-Means (Clustering):**

k-means se utiliza para hacer agrupaciones o para identificar patrones en los datos, en especial cuando dichos datos no cuentan con categorías en específico. Ayuda a dividir un solo conjunto de datos en un número k de grupos en donde hay similitud entre los elementos pertenecientes a un mismo grupo.

"En el proceso de K-means, el analista debe especificar previamente el número de grupos (k) que se desea obtener. El algoritmo clasifica los objetos en diferentes grupos, de manera que los objetos dentro de un mismo grupo sean lo más parecidos entre sí." (Tejada, 2023)

K-means se usa para segmentar los datos y descubrir relaciones no tan fáciles de visualizar entre los datos cuando hay grandes volúmenes de información por tratar.

- **Aplicaciones del K-means en los negocios**

La finalidad principal del uso de k-means dentro de un contexto empresarial es buscar grupos dentro de los clientes, productos o procesos que pueden ser optimizados mediante estrategias de marketing, operaciones e investigación y desarrollo. Algunos de sus usos en este contexto son los siguientes:

- **Segmentación de clientes:** el algoritmo permite la agrupación de clientes de acuerdo a sus gustos, sus preferencias, poder adquisitivo, entre otros.

- **Análisis de mercados**: busca ayudar a identificar mercados emergentes identificando características similares en términos de consumo, ingresos o preferencias culturales.
- **Optimización de inventarios**: clasifica productos en función de la demanda, la frecuencia de ventas o el margen de las ganancias.

○ **2. Forecast (Pronóstico):**

Gracias a que un modelo de forecast ayuda a predecir valores futuros tomando como referencia datos históricos, se usa para predecir el clima, la evolución de pandemias como la reciente del COVID, o la demanda de productos.

■ **Aplicación de Forecast en los negocios:**

Su objetivo principal es prever tendencias futuras para mejorar la toma de decisiones y maximizar la eficiencia.

- **Pronóstico de demanda**: predecir la cantidad de productos necesarios para satisfacer la demanda a su vez que minimiza los costos de inventarios.
- **Análisis financiero**: hacer estimaciones de ingresos futuros, costos y rentabilidad. En los bancos se usa el forecast para identificar tendencias de tasas de interés o los comportamientos de los mercados bursátiles.
- **Marketing y ventas**: predecir el impacto de campañas publicitarias en los mercados objetivos o identificar cual es la mejor temporada para lanzar un producto en específico.

Marco conceptual

Dentro de esta investigación, se pretende analizar una parte del mercado automovilístico ecuatoriano específicamente de la marca Peugeot, para poder analizar esto se debe definir los distintos tipos de autos que la marca comercializa.

○ **Autos urbanos o de ciudad:**

Un coche urbano es aquel que tiene su mayor ámbito de uso pensado para dentro de la ciudad y alrededores. Es más pequeño que los utilitarios, con una longitud de, en torno a, 3,5 metros de largo. Con este reducido tamaño, son ágiles en ciudad y fáciles de aparcar. Suelen tener buen radio de giro, para que las maniobras sean sencillas, y

cuentan con direcciones asistidas en las que prima la suavidad de manejo del volante (Autocasión, 2016, párr. 5).

Dentro de este tipo de autos Peugeot tiene varios modelos alrededor del mundo, uno de los más conocidos aquí en Ecuador es el Peugeot 208, siendo este un vehículo urbano de aspecto deportivo.

-

- **Autos Hatchbacks:**

Este tipo de automóvil tiene un alto nivel de practicidad y se clasifica como compacto. Un automóvil hatchback es un automóvil con base sedán con una parte trasera más corta y la puerta del maletero está diseñada para abrirse hacia arriba. Los compartimentos de carga y pasajeros se combinan para un uso simultáneo. El acceso a la puerta trasera se puede utilizar a través de la tercera o quinta puerta, con bisagras hacia arriba. La función de asiento trasero plegable permite flexibilidad para usar los compartimentos de carga y pasajeros simultáneamente (Yamin et al., 2023, p. 89).

Asimismo, dentro del mercado ecuatoriano un ejemplo de este tipo de autos también sería el Peugeot 208 y además el 308, siendo un auto de tipo urbano, cómodo y de buen precio.

- **Autos SUVs:**

Como sabemos los autos SUVs son los más populares dentro de los modelos que comercializa Peugeot, no solo en Ecuador sino en América Latina en general e incluso Europa. Como definición podemos decir que:

Los SUV, o Sport Utility Vehicles, son coches elevados con la apariencia de un todoterreno pero que no están diseñados para sortear obstáculos fuera de la carretera. Los SUV suelen tener entre 5 y 7 plazas. Por lo tanto, los SUV son más utilizados por las familias. Compactos, híbridos o eléctricos, los SUV Peugeot ofrecen un gran confort y numerosas tecnologías de asistencia a la conducción, como el Grip Control, para vivir una experiencia única (Peugeot, 2023).

Debido a las características que tiene, la practicidad y lo económicos que son hacen que este modelo de autos sean de los más vendidos en Ecuador, algunos de los autos que Peugeot comercializa dentro de este modelo son: Peugeot 2008, 3008, SUV 5008, entre otros.

- **Vehículos utilitarios o comerciales:**

Y por último tenemos la gama de vehículos comerciales, de estos podemos encontrar algunos tipos pero solo se definirán los comerciales ligeros, estos se definen como:

Los vehículos comerciales ligeros (VUL, por sus siglas en inglés) son aquellos que tienen un peso máximo autorizado de hasta 3,5 toneladas. Son ideales para transportar pequeñas cargas y realizar trabajos en entornos urbanos. Algunos ejemplos de VUL incluyen: Furgonetas: vehículos cerrados utilizados para transportar mercancías o herramientas. Pick-ups: vehículos con una cabina y una plataforma abierta en la parte trasera para transportar cargas. Camiones ligeros: vehículos con una capacidad de carga mayor que las furgonetas y pick-ups, pero menor que los camiones pesados (Alarcón, 2023).

Como ejemplo de este tipo de vehículos en Peugeot Ecuador tenemos la camioneta Landtrek.

También se debería tener claro otros conceptos técnicos para poder entender de mejor manera el rumbo de la investigación.

- **Segmentación de mercado:**

A la tarea de dividir el mercado en grupos con características homogéneas, se le conoce con el nombre de "segmentación del mercado"; el cual, se constituye en una herramienta estratégica de la mercadotecnia para dirigir con mayor precisión los esfuerzos, además de optimizar los recursos y lograr mejores resultados (Thompson, 2005).

- **Inteligencia artificial:**

La IA es la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano. Sin embargo, a diferencia de las personas, los dispositivos basados en IA no necesitan descansar y pueden analizar grandes volúmenes de información a la vez. Asimismo, la proporción de errores es significativamente menor en las máquinas que realizan las mismas tareas que sus contrapartes humanas (Rouhiainen, 2018, p.17).

Marco Legal

Una norma es una regla o un conjunto de las mismas, una ley, una pauta o un principio que se adopta, se impone y se debe seguir para el desarrollo correcto de las acciones interpersonales, también las normas son medios para guiar, dirigir o ajustar la conducta y comportamiento de los individuos.

- **Normas técnicas INEN**

El INEN es una entidad nacional que se encarga de las Normas Técnicas Ecuatorianas que fue creada el 28 de agosto de 1970, el cual pretende satisfacer las necesidades nacionales y facilitar el comercio dentro del país así mismo que a nivel internacional, mediante la imposición de normas que los productos deben cumplir con carácter de obligatorio. En el sector automotriz hay algunas normas INEN que se deben cumplir como el Reglamento Técnico Ecuatoriano RTE INEN 034, que establece los requisitos mínimos de seguridad en un vehículo, estas normas técnicas abarcan los frenos, luces, cinturones de seguridad, etc., además de que los automóviles deben adherirse a normas ambientales relacionadas con emisiones de gases.

- **Marco legal en la IA**

A nivel nacional, la IA aún se encuentra en desarrollo y no solo en ámbitos como la seguridad, informática, procesamiento y tratamiento de datos y educación sin embargo su utilización también está presente en actividades diarias de la población ecuatoriana.

El 36 % para automatizar tareas repetitivas, otro 36 % para hacer consultas ante inquietudes diarias; el 28 % para mejorar la interacción con entornos digitales; y el 27 % para resolver problemas cotidianos. A pesar de que solo el 36 % de los encuestados en Ecuador hizo uso de la IA en su trabajo, para el 56 % es útil la incorporación en sus tareas diarias. La tendencia se repite en el resto de los países de la región. (Primicias, 2024)

El pasado 20 de junio del 2024, Ecuador presentó a la Asamblea el Proyecto de Ley Orgánica de Regulación y Promoción de la Inteligencia Artificial. Organismos internacionales que han profundizado en el estudio de la IA, la han definido como:

Un sistema de IA es un sistema basado en máquinas que, para unos objetivos explícitos o implícitos, infiere, a partir de la entrada que recibe, cómo generar salidas tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los distintos sistemas de IA varían en sus niveles de autonomía y adaptabilidad tras su despliegue (Fernandez, 2024).

Algunos de los artículos más relevantes de este proyecto se desarrollan a continuación:
El artículo 2 del nuevo proyecto de ley establece como uno de los principales objetivos crear un marco jurídico y dinámico centrado en las personas para poder establecer límites de los sistemas de inteligencia artificial, lo que se busca es prevenir y mitigar los posibles impactos negativos de la IA en los derechos y deberes de los ciudadanos, con énfasis en la privacidad, igualdad, libertad de expresión, autonomía y dignidad humana.

El artículo 3 establece que la ley se aplicará a todas las actividades de investigación, desarrollo, comercialización y uso de sistemas de IA realizadas por entidades públicas o privadas, nacionales o extranjeras que operen en territorio ecuatoriano o que generen efectos jurídicos sobre personas naturales o jurídicas domiciliadas en el país, independientemente del lugar donde ocurra dicho tratamiento.

Clasificación de los sistemas de inteligencia artificial por niveles de riesgo
En el proyecto de ley que se entregó se dictamina que los sistemas pertenecientes a la inteligencia artificial se clasificarían de acuerdo a su nivel de riesgo, ya sea bajo, moderado, alto y extremo. Las parte involucradas deben aplicar un enfoque basado en estos riegos con la finalidad de identificar, evaluar, prevenir y de ser el caso mitigar el impacto negativo que el uso de estos sistemas puedan tener sobre los derechos intrínsecos de las personas, la seguridad pública y el bienestar de los ciudadanos en general, teniendo el cuenta que existe una probabilidad de un evento fortuito y la gravedad de su impacto. Esto incluye el hecho de poder tomar medidas razonables y

que sean proporcionales a la naturaleza, el alcance, contexto y propósito de la IA en diferentes contextos.

Prohibiciones y responsabilidades de la Ley de Inteligencia Artificial en Ecuador

Este proyecto de ley incluye la prohibición del diseño, desarrollo y uso de algunos sistemas que puedan perjudicar la dignidad humana, los derechos fundamentales o el estado de derecho. Esto incluye los sistemas de calificación social masiva, técnicas subconscientes o que sean imperceptibles que influyan de manera significativa y sin el consentimiento previo informado en las decisiones, emociones o comportamientos inconscientes de las personas, tecnologías de vigilancia que no están regularizadas, así como los modelos de reconocimiento facial que puedan discriminar sin justificación por motivos de raza, color de piel, orientación sexual o identidad de género, entre otros.

Metodología

Tabla 1

Cuadro de operacionalización de una variable

Variable	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Instrumentos	Tipo de variable
País	Territorio con características y ubicación propia en el cual fue fabricado el vehículo	Lugar del cual el automovil fue fabricado y exportado a Ecuador	Ubicación geográfica	Lugar de origen del vehículo	Registro de automoviles matriculados de acuerdo a datos del SRI	Categórica nominal
Clase	Denominación del automovil dependiendo de sus características generales	Clasificación por la función, carrocería, tamaño y sistema de tracción	Función, carrocería, tamaño, capacidad y sistema de tracción	Función y tamaño del vehículo	Registro de automoviles matriculados de acuerdo a datos del SRI	Categórica nominal
Subclase	Denominación del automovil dependiendo de sus características más específicas	Clasificación más específica tomando en cuenta adicionalmente la capacidad, tamaño, altura, tracción, capacidad del motor.	Capacidad, tamaño, altura, tracción, capacidad del motor	Tipo del vehículo específico dentro de la clase	Registro de automoviles matriculados de acuerdo a datos del SRI	Categórica nominal
Avalúo	Valor o precio del vehículo	Precio en dólares dependiendo del valor comercial	Características del automovil	Kilometraje, modelo y estado del vehículo	Registro de automoviles matriculados de acuerdo a datos del SRI	Numérica contínua
Cilindraje	Capacidad del cilindro del motor	Volumen total de los cilindros el cual es resultado de la multiplicación del volumen de cada cilindro por el número de cilindros del motor	Volumen total de los cilindros medido en centímetros cúbicos o litros	Volumen en cc o litros	Registro de automoviles matriculados de acuerdo a datos del SRI	Numérica contínua
Tipo de combustible	Combustible utilizado para generar energía mecanica para el funcionamiento del vehículo	Tipo de combustible apropiado para cada tipo de vehículo dependiendo de sus necesidades	Tipo de combustible	Gasolina, diesel	Registro de automoviles matriculados de acuerdo a datos del SRI	Categórica binaria
Fecha de compra	Fecha en la que se compró el automovil	Fecha exacta con día, mes y año de la compra del vehículo	Día, mes y año	Fecha específica de compra	Registro de automoviles matriculados de acuerdo a datos del SRI	Numérica contínua

○ Algoritmo de K-means

K-means lo que busca es optimizar, es decir minimizar la suma de las distancias cuadráticas de los datos pertenecientes a un mismo grupo en relación con el centroide de su clúster. Las observaciones se representan con vectores que tienen una N dimensión y lo que se busca es minimizar la distancia que hay entre los datos y el centro de su grupo.

La ecuación se puede formular de la siguiente manera:

$$Min\ E(\mu_i) = min_s \sum_{i=1}^{k} \quad \sum_{X_j \varepsilon S_i} \quad \left|\left|X_j - \mu_i\right|\right|^2 \qquad (14)$$

Donde S es el grupo de datos cuyos elementos son los objetos de Xj representados por vectores, y cada elemento representa un atributo. Cada grupo K tiene su centroide μ_i

En cada cálculo del centroide, existe una condición necesaria para la función E(μ_i), cuya función cuadrática es la siguiente:

$$\frac{\partial E}{\partial \mu_i} = 0 \Rightarrow \mu_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{X_j \epsilon S_i^{(t)}} \quad X_j \qquad (15)$$

En este caso se toma el promedio de los elementos que tiene cada grupo de datos como el nuevo centroide.

- **Asignación de clústeres:**

Cada punto de datos Xi se asigna al cluster con el centroide más cercano:

$$Ci = arg\ min_j \left|x_i - \mu_j\right|^2 \qquad (16)$$

Ci: es el cluster asignado al punto Xi

μ_j: el centroide del cluster j

‖Xi-Ui‖²= la distancia euclidiana al cuadrado entre Xi y Ui

- **Cálculo del centroide:**

Calcular el nuevo centroide para cada cluster como el promedio de los puntos asignados al mismo:

$$\mu_j = \frac{1}{|Cj|} \sum_{Xi \epsilon Cj} \quad X_i \qquad (17)$$

μ_j:el nuevo centroide cluster j

C_J= el conjunto de puntos en el cluster

$\left|C_j\right|$:el número de puntos en el cluster j

- **Algoritmos de Forecast**

ARIMA (AutoRegressive Integrated Moving Average)

1. Preprocesamiento

Primero es necesario saber si la serie es estacionaria, es decir si tiene o no tendencia y estacionalidad). En el caso de ser estacionaria se debe aplicar la diferencia hasta que llegue a la estacionalidad.

2. Modelado

Se determinan los parámetros autoregresivos, diferencias y medias móviles usando algún tipo de criterio como el AIC o BIC.

Los criterios de AIC y BIC con el Criterio de Información de Akaike y Criterio de Información Bayesiano respectivamente, son métricas que son usadas principalmente para la evaluación de la calidad de modelos estadísticos y de machine learning, y son de gran utilidad para la elección de algún tipo de modelo como las series temporales ya sea ARIMA, regresión, entre otras.

3. Entrenamiento

Se ajusta el modelo a los datos precedentes para poder estimar los coeficientes.

4. Predicción

Se utiliza el modelo ajustado para poder hacer una predicción de los valores futuros.

5. Evaluación

Se comparan los resultados de las predicciones con los datos reales del modelo, utilizando métricas como el Error cuadrático medio o el Error absoluto medio.

Criterio de Información de Akaike (AIC)

Este criterio evalúa el equilibrio que existe entre la bondad del ajuste del modelo y su complejidad. Su fórmula se expresa de la siguiente:

$$AIC = -2ln(L) + 2k \qquad (18)$$

L: es la función de verosimilitud del modelo, es decidir explica que tan bien el modelo se ajusta a los datos.

k: es el número de parámetros estimados del modelo.

Criterio de Información Bayesiano (BIC)

Tiene una similitud con el AIC pero tiene una diferencia, si el tamaño de los datos (n) es grande tiene una penalización mayor para la complejidad del modelo. Su ecuación es:

$$BIC = -2ln(L) + kln(n) \qquad (19)$$

n: número de observaciones en los datos

L: función de verosimilitud del modelo

k: es el número de parámetros estimados

Medias Móviles Simple (SMA)

El promedio Móvil Simple es una técnica de análisis de datos cuyo uso es principalmente en series temporales para suavizar las fluctuaciones que se pueden hacer en el corto plazo y observar tendencias a largo plazo. Su resultado es el promedio de un conjunto de los valores en un intervalo de tiempo en específico, y que siempre se mueve hacia adelante con cada paso.

Pasos:

1. Para cada tiempo (T), se calcula el promedio de los últimos X valores.

$$\widehat{Yt} = \frac{1}{n}\sum_{i=t-n}^{t-1} Yi \qquad (20)$$

2. Se usa el resultado de la ecuación anterior como la predicción de Yt.

- **Introducción a R Studio**

RStudio es un herramienta de programación sumamente importante en análisis estadísticos y de machine learning gracias a su utilidad en el tratamiento de datos y la creación de gráficos. R tiene su origen en el lenguaje de programación S, diseñado en los laboratorios Bell en los años de 1970. En el 1993, Ross Ihaka y Robert Gentleman crearon R como una implementación del código S, y desde ese momento R ha tenido grandes avances que lo han convertido en uno de los lenguajes de programación más usados en el campo de Big Data y estadístico. En el presente trabajo de titulación se ha utilizado la versión R 4.3.3 de febrero del 2024.

- **Ventajas de RStudio**

RStudio tiene múltiples ventajas, una de las principales es su facilidad de uso, esto hace que sea amigable con los usuarios que recién están aprendiendo el análisis de datos. El proceso de carga de datos es relativamente sencillo, por lo que se facilita la administración de los datos a tratar al escribir y desarrollar los códigos. Por otro lado, también cuenta con la opción de generar gráficos que permiten una mejor interpretación de los resultados de los modelos.

La instalación y desinstalación de paquetes se hacen directamente desde RStudio lo cual es beneficioso ya que se ahorra el aprender a usar la terminal de interfaz de los comandos o los símbolos del sistema que puede ser un poco más complejos de usar.

Asimismo, la plataforma cuenta con un autocompletado de códigos, por lo que se facilita el proceso de codificación, y es necesario destacar que también posee una comunidad de investigadores activa donde los usuarios pueden consultar y compartir sus conocimientos, resolver alguna inquietud y aumenta la posibilidad de intercambiar ideas, por lo cual es de gran ayuda en el desarrollo de modelos y algoritmos que permitan el tratamiento efectivo de datos para la obtención de resultados mejorados.

- **Información sobre la base de datos**

La base de datos que se va a utilizar fue sacada de la página web del Servicio de Rentas Internas (SRI) y muestran las ventas de carros en el Ecuador, específicamente de marca Peugeot, durante los años del 2018 al 2023, en donde se encuentran diferentes variables que se van a utilizar para hacer los modelos de K-means y Forecast

Librerías de RStudio

R cuenta con una variedad de librerías que permite al usuario hacer tareas en específico y que se clasifican de acuerdo a su funcionalidad. A continuación se presentan las librerías que se van a utilizar en el trabajo:

Librerías de K-means

Caret: es un paquete que contiene herramientas para el entrenamiento de datos y la evaluación de modelos de aprendizaje automático. Esta librería incluye métodos para la clasificación y regresión,

Ggplot2: es una librería para la elaboración de gráficos que sirvan de apoyo visual a los resultados obtenidos del desarrollo de los modelos.

GridExtra: es una librería para la elaboración de gráficos que ayuda a combinar varios gráficos en una misma cuadrícula para poder hacer una visualización minimizada de las gráficas.

Tidyverse: es una colección de varias librerías que se encapsulan en una que permite hacer el trabajo de manipular, importar y visualizar los datos de manera más fácil.

Algunos de los paquetes que contienen son readr, dplyr, tidyr, forecast, entre otros.

Class: es una librería que contiene múltiples funciones que son útiles para clasificación.

Cluster: Es un paquete utilizado para hacer análisis de clustering, que incluye una gran diversidad de algoritmos y métodos que son usados principalmente para identificar las agrupaciones en la base de datos. Algunas de sus funciones es la realización de análisis de agrupamiento jerárquico aglomerativo, conocido también como AGNES, y el agrupamiento jerárquico divisivo, conocido como DIANA.

Factoextra: este paquete facilita la extracción y visualización de los datos obtenidos en alguna clase de análisis exploratorio de datos de diferente naturaleza, como por ejemplo el Análisis de componentes principales también conocido como PCA.

NbClust: este paquete se utiliza principalmente para determinar el número de clústeres en un conjunto de datos y también le da el usuario el número óptimo de agrupamientos mediante el método del codo.

Librerías de Forecasting

Forecast: este paquete de R proporciona al usuario métodos y herramientos para hacer pronósticos de series de tiempo, incluyendo el método de suavizado exponencial.

Dplyr: se trata de una versión mejorada del paquete plyr, consiste en un conjunto de verbos que permitirán manipular las bases de datos a tratar. Algunos ejemplos de verbos son “filter”, “select”, “arrange”, etc.

Lubridate: se utiliza cuando se está desarrollando un modelo de serie de tiempo y no permite que los datos usen fechas y horas por lo que este paquete facilita el uso de las fechas.

RandomForest: este paquete permite realizar bosques aleatorios para regresión y clasificación. Esta librería ayuda a construir árboles no correlacionados para que mejore su desempeño y los resultados tengan mejor precisión.

Otras librerías

Lattice: paquete usado para la visualización de comparativas entre los diferentes tipos de variables, ya sean estas categóricas, cualitativas y cuantitativas. Se usa en datos multivariables.

VIM: paquete de R que es utilizado para la fase inicial del tratamiento de datos, la limpieza de la base de datos, ya que con esta librería se puede visualizar los datos faltantes, conocidos como los NAs de los datos que se van a utilizar.

Modeest: este paquete tiene utilidad en el data cleaning y es fundamental para sacar las medidas de tendencia central, específicamente la moda, con lo que se brindará información sobre la base de datos que se va a usar en los modelos.

Moments: el paquete es usado para obtener los coeficientes de normalidad, estos son la curtosis y el coeficiente de asimetría

- **Metodología K-means**

El método de K-means es un algoritmo que nos permite clasificar los datos de acuerdo a las características que tienen en común, esto hará posible la clasificación e identificación de clientes potenciales.

Para comenzar se debe importar la base de datos a R para proceder con su análisis.

Peugeot2 <- read.csv2("../Downloads/Compilado ventas peugeot2.csv")

Para este algoritmo y para poder segmentar a los clientes de manera correcta solo se usarán 2 variables, la variable precio y el cilindraje los cuales son los necesarios para definir y clasificar a los clientes dependiendo de su poder adquisitivo, para esto haremos una subclasificación de variables con el siguiente comando:

Peugeot_data_cleaned <- Peugeot2[,c(4,5)]

Luego, se llaman a las librerías a usar en este algoritmo específicamente, las cuales son las siguientes:

```
library(class)
library(caret)
library(tidyverse)
library(cluster)
library(factoextra)
library(NbClust)
```

A continuación, se planta la semilla, dentro de R existen algunos tipos de semilla, pero, en este modelo usaremos la 123, usaremos semilla ya que este modelo incluye componentes aleatorios, con esta semilla se podrá generar reproducibilidad en los resultados.

```
set.seed(123)
```

Luego de esto es muy importante dividir los datos en entrenamiento y prueba, los datos del entrenamiento se usan para que el modelo aprenda las características y relaciones que existen entre las variables, mientras que los datos de prueba son la parte de los datos que el modelo no ha visto en el entrenamiento y se utiliza para evaluar los resultados del modelo.

Para lo antes mencionado esto se tiene el comando siguiente:

```
trainIndex <- createDataPartition(Peugeot_data_cleaned$VENTA,
                                  p=.7, list = FALSE)
```

TrainIndex: es un vector que contiene el conjunto de datos del entrenamiento.

CreateDataPartition: es la función que parte los datos de entrenamiento en subconjuntos de manera aleatoria.

Peugeot_data_cleaned: es el conjunto de datos con las variables seleccionadas a usar en el modelo.
Venta: es la variable que se va a segmentar
p=7: representa el porcentaje de datos que se van a utilizar en el entrenamiento, es decir que se tomó el 70% de los datos.
list=FALSE: esto en una función de visualización que permitirá ver los resultados no es forma de lista sino como filas.

train_data <- Peugeot_data_cleaned[trainIndex,]
Train_data: es el vector que representa los datos de entrenamiento
Peugeot_data_cleaned: es el conjunto de datos con las variables seleccionadas a usar en el modelo.
TrainIndex: es un vector que contiene el conjunto de datos del entrenamiento, es decir el 70% correspondiente

test_data <- Peugeot_data_cleaned[-trainIndex]
Test_data: es el vector que representa los datos de prueba
Peugeot_data_cleaned: es el conjunto de datos con las variables seleccionadas a usar en el modelo.
-TrainIndex: es un vector que contiene el conjunto de datos de la prueba, es decir el 30% restante.

El siguiente paso es el escalamiento de datos, este proceso se usa para poder estandarizar los datos es decir que se quiere que los datos tiendan a la normalidad que sigan una tendencia normal de los datos en la campana de Gauss.

SKtraindata <- scale(train_data)
SKtraindata: vector representativo del escalamiento de los datos de entrenamiento
Scale: función que va a permitir el escalamiento
train_data: datos de entrenamiento

SKtestdata <- scale(test_data)
SKtestdata: vector representativo del escalamiento de los datos de prueba
Scale: función que va a permitir el escalamiento

test_data: datos de prueba

A continuación, se empieza a trabajar con los datos de entrenamiento

Se trabaja con los datos de entrenamiento para, con la ayuda de 3 métodos ("silhouette", "wss" y "gap_stat") poder encontrar el número óptimo de clusters.

Método silhouette:

fviz_nbclust(SKtraindata, kmeans, method = "silhouette")

fviz_nbclust: Se encuentra dentro de la librería factoextra y es el encargado de determinar y visualizar el número óptimo de clusters, se puede usar en diferentes métodos así como el silhouette.

SKtraindata: vector representativo del escalamiento de los datos de entrenamiento

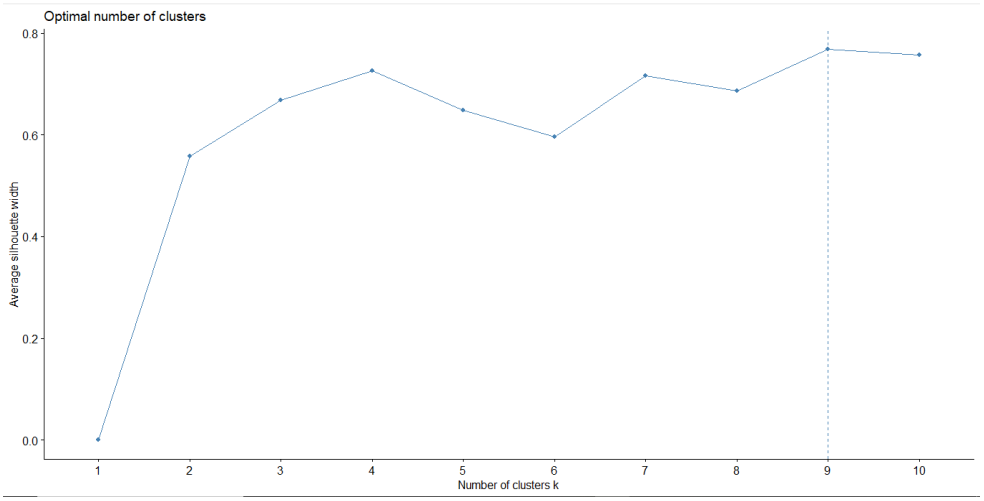
Kmeans: Algoritmo de agrupamiento usado específicamente para formar los clusters.

Method = "silhouette": Primer método usado para buscar el número óptimo de clusters.

Con esta función obtuvimos la siguiente gráfica:

Figura 11

Número óptimo de clusters con el método silhouette



Se usó el método "silhouette" el cual demostró que existe un primer quiebre significativo en 2, lo que quiere decir que al inicio del análisis los datos cuentan con 2 grupos bien definidos, pero el gráfico señala que según este método el número óptimo de clusters es 9 ya que tiene el valor más alto en el eje y.

Método wss:

fviz_nbclust(SKtraindata, kmeans, method = "wss")

fviz_nbclust: Se encuentra dentro de la librería factoextra y es el encargado de determinar y visualizar el número óptimo de clusters, se puede usar en diferentes métodos así como el Within-Cluster Sum of Squares.

SKtraindata: vector representativo del escalamiento de los datos de entrenamiento

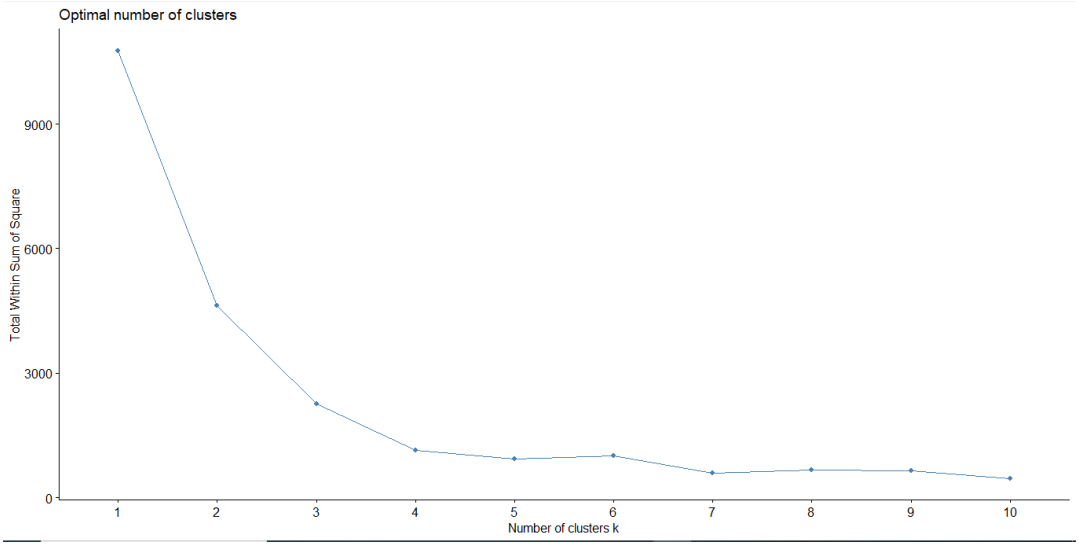
Kmeans: Algoritmo de agrupamiento usado específicamente para formar los clusters.

Method = "wss": Segundo método usado para buscar el número óptimo de clusters.

Con esta función obtuvimos la siguiente gráfica:

Figura 12

Número óptimo de clusters con el método wss



En esta gráfica se puede notar que el wss o la suma total de cuadrados internos va disminuyendo, esto significa que los puntos de cada agrupación están más cerca del centroide, cuando el punto deja de disminuir significativamente o cuando se crea el quiebre o codo en la gráfica es cuando se ha encontrado el número óptimo de clusters, según este método el número óptimo de clusters es 3 ya que es donde se produce el quiebre, a partir de aquí la gráfica deja de decrecer significativamente.

Método gap_stat:

fviz_nbclust(SKtraindata, kmeans, method = "gap_stat")

fviz_nbclust: Se encuentra dentro de la librería factoextra y es el encargado de determinar y visualizar el número óptimo de clusters, se puede usar en diferentes métodos así como el método gap statistic.

SKtraindata: vector representativo del escalamiento de los datos de entrenamiento

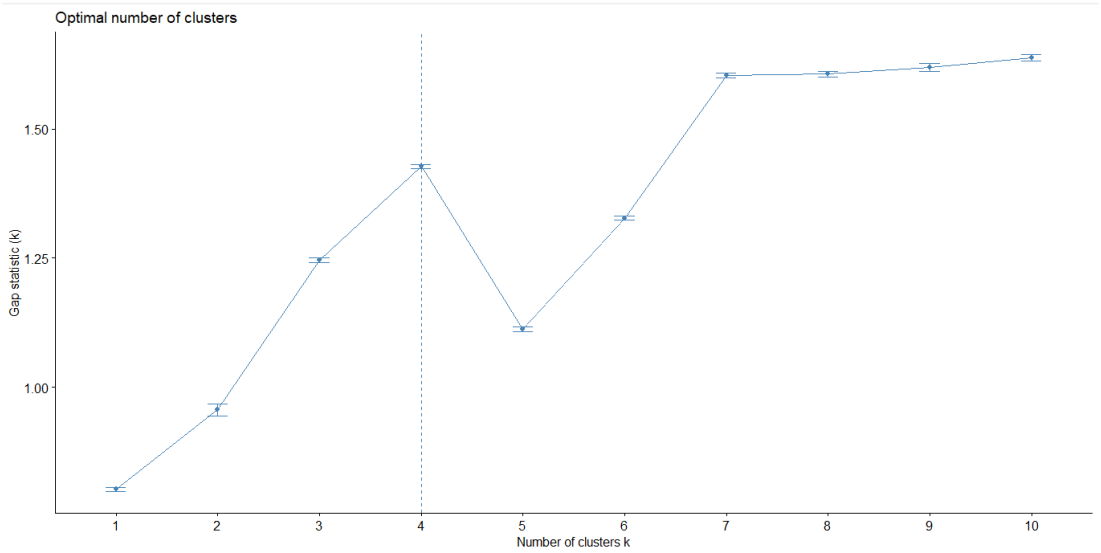
Kmeans: Algoritmo de agrupamiento usado específicamente para formar los clusters.

Method = "gap_stat": tercer método usado para buscar el número óptimo de clusters.

Con esta función obtuvimos la siguiente gráfica:

Figura 13

Número óptimo de clusters con el método gap stat



Según este método el número óptimo de clusters es 4, esto se debe a que el valor de estadística del gap a partir de k=4 deja de aumentar significativamente. Este método se basa en comparar la dispersión que hay entre los datos del conjunto original con conjuntos de datos aleatorios, cuando se logra encontrar el punto donde las diferencias entre ambos valores se maximizan se logra encontrar el número óptimo de clusters.

Comparando las tres gráficas resultantes se llegó a la conclusión de que el número óptimo de clusters sería 3, debido a que en la gráfica del método del codo no existe mayor variación después de que el codo caiga en 3.

A continuación, después de haberse definido que el número óptimo de clusters es 3, se calculan los clusters.

K3 <- kmeans(SKtraindata, centers = 3, nstart = 25)

K3: Es el objeto que va a almacenar los datos de los 3 clusters creados.

kmeans: Algoritmo aplicado

SKtraindata: vector representativo del escalamiento de los datos de entrenamiento.

Centers=3: Especificación de dividir los datos en 3 clusters.

Nstart=25: Número de veces que se ejecutará el algoritmo de kmeans, esto para asegurar resultados más confiables.

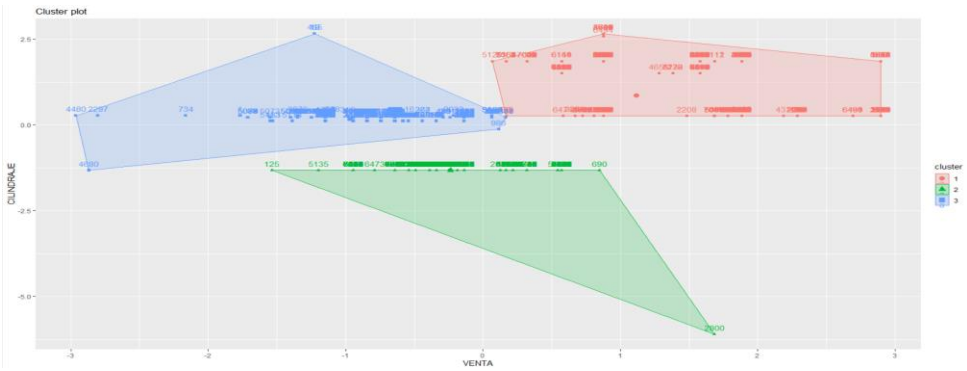
Con los clusters calculados se procede a graficarlos para visualizar los resultados del modelo de clustering, se usó la función fviz_cluster del paquete factoextra en 3 diferentes maneras, la primera se usó para ver de una manera inicial y más simple la distribución de los datos y la asignación de clusters, en la segunda ya se agrega la distancia euclidiana la cual nos permite notar la separación geométrica entre clusters y su distancia entre centroides, la tercera nos permite notar de una mejor manera la dispersión de los datos dentro de cada uno de los clusters para ver si están distribuidos homogéneamente.

PTrainOC <- fviz_cluster(K3, data = SKtraindata,repel = FALSE)

PTrainOC= es el vector que va a representar el primer resultado gráfico de los clusters
Fviz_cluster= función que ayudará a visualizar los resultados de los segmentos que se obtuvieron.
k3= Es el objeto que almacena los datos de los 3 clusters creados.
SKtraindata= son los datos con el que se está desarrollando el gráfico de clusters, en este caso los datos de entrenamiento.
repel=FALSE= es una función para evitar que las etiquetas se sobrepongan y que sean legibles en el gráfico.

Figura 14

Resultados del modelo de clustering



El gráfico muestra el número de clusters que con anterioridad se habían tenido como resultado, en este caso se ven los 3 segmentos bastantes definidos con un color

diferente cada uno, el cluster 1 es el rojo, el cluster 2 es el verde y el cluster 3 en es de color azul. Otra característica de este resultado es que ninguno de los grupos se cruzan.

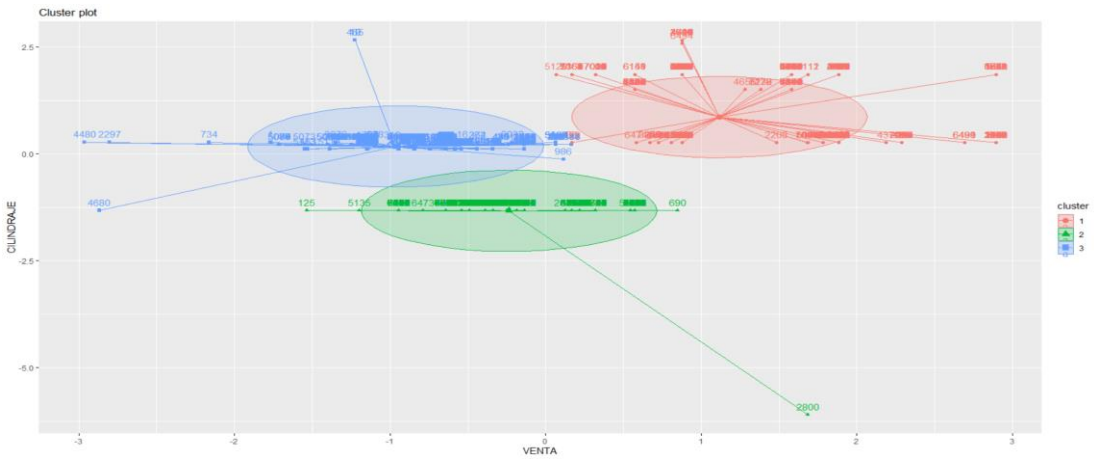
Posteriormente se desarrolla el mismo proceso, pero esta vez se incluye la distancia Euclidiana.

```
PTainMC <- fviz_cluster(K3, data = SKtraindata,
                        ellipse.type = "euclid",
                        repel = FALSE, star.plot= TRUE)
```

PTainMC= es el vector que va a representar el segundo resultado gráfico de los clusters
Fviz_cluster= función que ayudará a visualizar los resultados de los segmentos que se obtuvieron.
k3= Es el objeto que almacena los datos de los 3 clusters creados.
SKtraindata= son los datos con el que se está desarrollando el gráfico de clusters, en este caso los datos de entrenamiento.
ellipse.type= "euclid" es una función que dibuja elipses alrededor de cada punto perteneciente a cada cluster. El "euclid" hace que se dibujen los elipses en el centroide de cada cluster que ya han sido calculados con la distancia euclidiana.
repel=FALSE= es una función para evitar que las etiquetas se sobrepongan y que sean legibles en el gráfico.
start.plot=TRUE= es una función que dibuja "estrellas" que unen cada uno de los puntos pertenecientes a un cluster con su centroide respectivo.

Figura 15

Resultado del modelo de clustering incluyendo distancia euclidiana



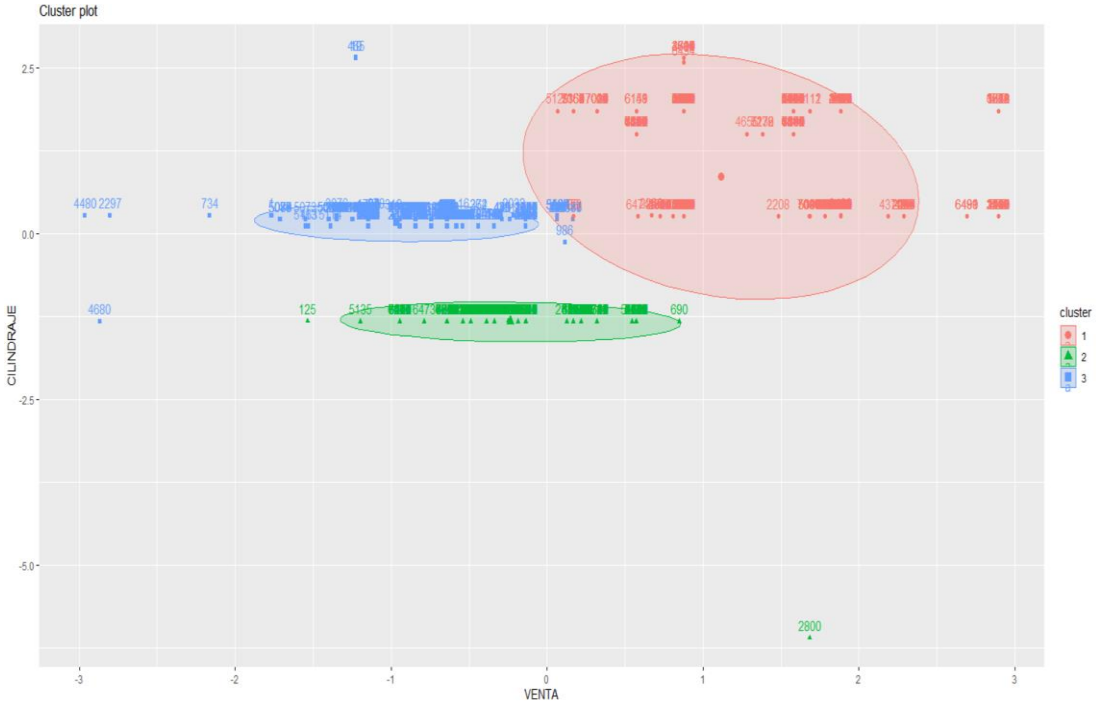
El gráfico muestra el resultado más específico ya que incluye detalles como el centroide de cada uno de los clusters y su conexión con cada uno de los puntos que pertenecen a los grupos, se mantienen los colores de cada uno. El algoritmo trata de agrupar los datos de acuerdo a su cercanía con los centroides de cada uno de los segmentos identificados, por lo que el cluster el color verde tiene un dato que está más distante que los otros, el cual se ha identificado como un dato atípico es decir que su valor tiene gran diferencia en comparación con los otros datos, sin embargo el algoritmo lo ha incluido en el cluster verde.

```
PTrainIC <- fviz_cluster(K3, data = SKtraindata,
                ellipse.type = "norm",repel = FALSE)
```

PTrainIC= es el vector que va a representar el tercer resultado gráfico de los clusters
Fviz_cluster= función que ayudará a visualizar los resultados de los segmentos que se obtuvieron.
k3= Es el objeto que almacena los datos de los 3 clusters creados.
SKtraindata= son los datos con el que se está desarrollando el gráfico de clusters, en este caso los datos de entrenamiento.
ellipse.type=norm: es un función que dibuja los elipses basadas en una distribución norma, por lo que los elipses se dibujan tomando como referencia la varianza y la covarianza de los puntos de cada cluster, lo que muestra como los puntos se encuentran distribuidos en el espacio.
repel=FALSE= es una función para evitar que las etiquetas se sobrepongan y que sean legibles en el gráfico.

Figura 16

Resultado del modelo de clustering de manera optimizada.



El gráfico muestra los elipses de cada cluster que son hechos de acuerdo a la varianza y covarianza de cada uno de los grupos, lo que se puede observar es que el cluster rojo tiene mucha dispersión en las variables que se están estudiando, en este caso el cilindraje y las ventas, por lo que sus datos están dispersos. El cluster verde tiene una elipse más larga y estrecha lo que quiere decir que los datos están más agrupados, hay concentración en valores más bajos por lo que representa un grupo de productos con características más específicas y con valores en ventas más bajos. El cluster azul tiene una dispersión moderada lo que quiere decir que existe más variación entre las ventas y el cilindraje, este cluster tiene una dispersión horizontal en Ventas y estrecha en cilindraje por lo hay más variabilidad en ventas pero no varía mucho en las características técnicas es decir el cilindraje.

Posteriormente se va a trabajar con los datos de prueba

Kt3 <- kmeans(SKtestdata, centers = 3, nstart = 25)

Kt3: Es el objeto que va a almacenar el resultado de los 3 clusters creados con los datos de prueba, es decir aquellos datos que no han sido vistos por el modelo con anterioridad.
kmeans: Algoritmo aplicado
SKtestdata: vector representativo del escalamiento de los datos de prueba.
Centers=3: Especificación de dividir los datos en 3 clusters.
Nstart=25: Número de veces que se ejecutará el algoritmo de kmeans, esto para asegurar resultados más confiables.

Peugeot_data_cleaned$class <- NA
Peugeot_data_cleaned$class: son los datos limpios con las variables que se han escogido, en específico la variable de clasificación que muestra los 3 cluster clasificados con anterioridad.
NA= se eliminan los NA, es decir los datos que están faltantes

Peugeot_data_cleaned$class[trainIndex] <- as.factor(K3$cluster)

Peugeot_data_cleaned$class: son los datos limpios con las variables que se han escogido, en específico la variable de clasificación que muestra los 3 cluster clasificados con anterioridad.
TrainIndex= son los datos a usarse.
As.factor= es una función que convierte el número de cluster en factor, que es un dato categórico en R.
(K3$cluster)= es el vector que representa el número de clusters calculado
Peugeot_data_cleaned$class[-trainIndex] <- as.factor(Kt3$cluster)
Peugeot_data_cleaned$class: son los datos limpios con las variables que se han escogido, en específico la variable de clasificación que muestra los 3 cluster clasificados con anterioridad.
-TrainIndex= son los datos que van a usarse para evaluar el modelo, en esta caso existe un signo de - ya que se van a utilizar los datos restantes del conjunto de datos, es decir si ya se ha utilizado los de entrenamiento ahora es turno de utilizar los de prueba.
As.factor= es una función que convierte el número de cluster en factor, que es un dato categórico en R.
(K3$cluster)= es el vector que representa el número de clusters calculado

Peugeot2$clasificacion <- Peugeot_data_cleaned$class

En este caso se pasa la clasificación ya obtenida a la base de datos original

Peugeot2$clasificacion= es el vector nuevo que va a contener todas las variables incluyendo la nueva columna de la clasificación obtenida.

Peugeot_data_cleaned$class= es la base de datos que contiene los nuevos resultados de clasificación

```
G1=ggplot(Peugeot2,aes(x=Peugeot2$clasificacion,
                       y=Peugeot2$VENTA))+
  geom_boxplot(fill="yellow",colour="black")+
  theme(axis.title = element_blank())+
  theme(legend.position = "none")+
  labs("Precio con respecto al segmento")
```

G1: es el vector que va a almacenar el gráfico que se realizó con la finalidad de hacer una comparación de las ventas según la clasificación dada por Kmeans.

Peugeot2$clasificacion: es el objeto que contiene las clasificaciones que se obtuvieron de la base de datos principal, en este caso está puesto en el eje de las X en el plano cartesiano.

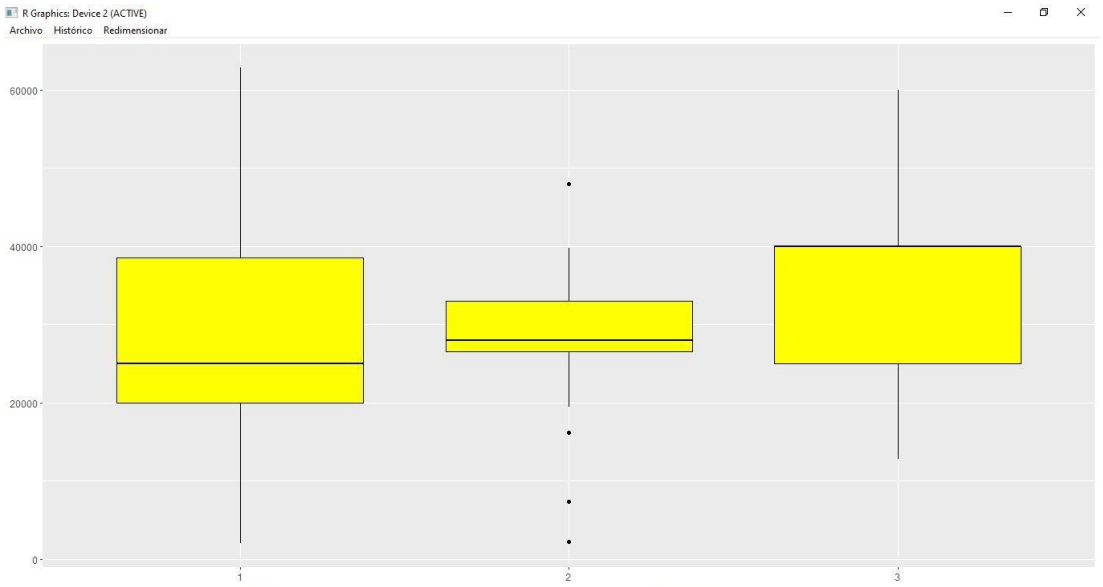
Peugeot2$VENTA: es el objeto que contiene las ventas de los segmentos que fueron clasificados, en este caso está ubicado en el eje de las Y dentro del plano cartesiano.

Geom_boxplot, theme y labs: son funciones que permiten al usuario la configuración del gráfico para mejorar la visualización del mismo.

Se realiza un gráfico que muestre la relación que existe entre las variables que fueron estudiadas, en este caso Precio en relación a la clasificación ya obtenida.

Figura 17

Gráfico de relación entre precio y segmento de cliente.



En el gráfico se observan las 3 clasificaciones hechas por Kmeans, las cuales representan 3 segmentos de mercados diferentes con características distintas. El segmento 1 representa el segmento de mercado medio, con ventas entre $20.000 y $39.000. El segmento 2 refleja una caja más pequeña, lo que sugiere que las ventas tienen menor dispersión y están en un rango entre $25.000 y $40.000, debido a esto la clasificación representa un mercado de escala comercial baja. Finalmente, el segmento 3 es el que mayor espacio abarca dentro del plano cartesiano por lo que muestra un rango de ventas entre $20.000 y $60.000 con una mediana cercana a los $40.000, por lo que es el segmento que tiene mayor ventas y que tiene mayor significancia de las tres clasificaciones.

Posteriormente se realizó otro gráfico boxplot con la finalidad de hacer una mayor distinción entre los segmentos clasificados y que se pueden ver con mayor facilidad sus atributos.

```
G2= ggplot(Peugeot2, aes(x = Peugeot2$clasificacion, y = Peugeot2$VENTA, fill =
Peugeot2$clasificacion)) +
geom_boxplot(outlier.shape = 21, outlier.size = 3, outlier.color = "red", notch =
TRUE) +
```

stat_summary(fun = mean, geom = "point", shape = 4, size = 4, color = "blue") +

geom_jitter(aes(color = Peugeot2$clasificacion), width = 0.2, alpha = 0.5, size = 2) +

scale_fill_brewer(palette = "Set3") +

scale_color_brewer(palette = "Dark2") +

labs(title = "Distribución de Precios por Segmento de Cliente",

subtitle = "Visualización detallada de la relación entre precios y segmentos",

x = "Segmentos de Cliente",

y = "Precios",

fill = "Segmentos",

color = "Segmentos") +

theme_minimal(base_size = 15) +

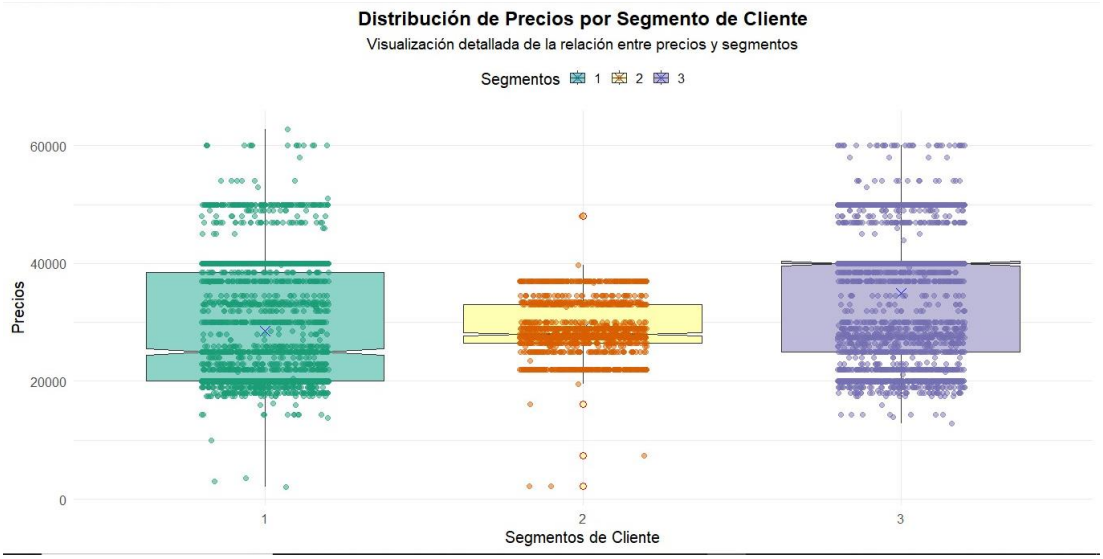
theme(legend.position = "top",

plot.title = element_text(hjust = 0.5, face = "bold", size = 18),

plot.subtitle = element_text(hjust = 0.5, size = 14))

Figura 18

Gráfico de distribución de precio por segmento optimizado.



En este gráfico se observan con mayor detalle los puntos pertenecientes a las observaciones que se usaron para el desarrollo del modelo, lo que permite visualizar de mejor manera la densidad y la dispersión de los datos. El segmento 1 (verde) presenta una mediana que está cerca a los $25.000, tiene una gran variabilidad en los

precios y tiene un rango intercuartílico amplio que va aproximadamente desde los $20.000 hasta los $39.000 y varios valores atípicos por encima y por debajo de la caja, por lo que tiene una distribución de datos amplia. El segmento de mercado 2 (amarillo) tiene una mediana aproximada de $29.000 y se muestra que es la más baja de los tres segmentos. Este grupo tiene una dispersión mucho menor con la mayoría de los datos agrupados entre los $25.000 y $35.000, y existe menor cantidad de datos atípicos para este segmento, en conclusión tiene una menor variabilidad de los datos y los valores centrales más bajos. Finalmente, el clúster 3 (morado) tiene una mediana alrededor de los $40.000 un poco similar al segmento 1, su dispersión es más amplia debido a que tiene un rango intercuartílico desde los $25.000 hasta los $40.000 aproximadamente y con presencia de pocos datos atípicos. Este segmento de mercado tiene los precios con mayor concentración de datos.

- **Forecasting por tipo de clasificación**

A continuación se procederá con el forecasting, para poder ver la tendencia y el comportamiento en ventas que está teniendo la empresa Peugeot en los últimos años para luego poder predecir las ventas hasta el 2026, para el forecasting se tienen en cuenta los 3 segmentos de clientes clasificados por K-means previamente.
Primero, se llaman a las librerías a usar en este algoritmo específicamente, las cuales son las siguientes:

```
Library(dplyr)
Library(lubridate)
Library(randomForest)
Library(ggplot2)
Library(forecast)
```

Luego, se valida la columna de fecha de compra del auto de la base de datos en un formato tipo "fecha" el cual R studio pueda comprender.

```
Peugeot2 <- Peugeot2 %>%mutate(FECHA = as.Date(FECHA))
```

Peugeot2: Base de datos nueva, la cual contiene la clasificación de la base en 3 segmentos.

Peugeot 2 %>%mutate: El operador pipe (%>%) es el encargado de pasar el valor de la izquierda en este caso Peugeot 2 como entrada en el primer argumento de la

función de la derecha en este caso mutate, la función mutate nos permite modificar las columnas existentes en un conjunto de datos, en este caso la columna fecha.

FECHA=as.Date(Fecha): Convierte los valores de la columna fecha de la base de datos y los transforma al formato Date, el cual es el formato estándar para fechas en R.

A continuación, se extrae el mes de la fecha, debido a que nuestra base original tiene las ventas por día, se suman los datos que están en el rango de días para obtener una base de datos más fácil de manejar con ventas mensuales.

Peugeot2 <- Peugeot2 %>%mutate(Mes = floor_date(FECHA, "month"))

Peugeot2: Base de datos nueva, la cual contiene la clasificación de la base en 3 segmentos.

Peugeot2 %>%mutate: El operador pipe (%>%) es el encargado de pasar el valor de la izquierda en este caso Peugeot 2 como entrada en el primer argumento de la función de la derecha en este caso mutate, la función mutate nos permite en este caso crear una nueva columna llamada "mes" calculada a partir de la columna fecha.

Mes = floor_date(FECHA, "month"): En este caso se usó la función floor_date del paquete de R Lubridate esta función sirve para redondear las fechas de cierta manera y mandar todos los días de un mes específico al día 1 del mes correspondiente, las nuevas fechas se guardarán bajo la columna "mes" en el dataset "Peugeot2"

Después, se suman las ventas por clasificación y por mes:

ventas_mensuales<-

Peugeot2%>%group_by(Mes,clasificacion)%>%summarise(Ventas = sum(VENTA, na.rm = TRUE)) %>%ungroup()

Ventas_mensuales: Nuevo dataset que contiene la sumatoria de ventas por mes y por segmentos.

Peugeot2%>%: El operador pipe (%>%) es el encargado de pasar el valor de la izquierda en este caso Peugeot 2 como entrada en el primer argumento de la función de la derecha, este operador se usa más de una vez en este código.

Group_by(Mes,clasificacion): Esta función es la encargada de agrupar los datos según el mes y la clasificación, es decir clasificará todos los datos que correspondan al mismo mes y todos los que correspondan a la misma clasificación.

Summarise(Ventas = sum(VENTA, na.rm = TRUE)): La función summarise se usa para resumir cada grupo, en esta función se trata de resumir la suma de las ventas para cada grupo omitiendo los valores NA, con esto se crea una nueva columna bajo el título "Ventas" que contiene el valor total de las mentas de cada mes dependiendo de su categoría o clasificación.
ungroup(): Esta función ayuda a que los datos vuelvan a un dataframe normal, los desagrupa.

El siguiente paso es seleccionar una clasificación específica y ver las ventas mensuales de esa clasificación.

clasificacion_filtro <- 2

Clasificacion_filtro <- 2: Este código indica la clasificación que se quiere filtrar en la base de datos (En este caso el segmento de clientes 2).

ventas_clasificacion <-
ventas_mensuales%>%filter(clasificacion==clasificacion_filtro)

Ventas_clasificacion: Nombre del nuevo dataset que contiene la sumatoria de las ventas de cada una de los segmentos.
Ventas_mensuales%>%filter: Este código usa el operador pipe, el cual es el encargado de pasar el valor de la izquierda en este caso Ventas mensuales como entrada en el primer argumento de la función de la derecha, en este caso filter.
Filter(clasificacion==clasificacion_filtro): Filter es una función del paquete dplyr la cual es usada para como su nombre lo indica filtrar filas de un conjunto de datos, en este caso la condición es que se seleccione únicamente las filas donde el valor de clasificación sea el mismo al de clasificación filtro, para así poder ver únicamente la sumatoria de ventas en el segmento de clientes requerido.

A continuación se debe verificar que haya datos disponibles para la clasificación especificada, para evitar errores y cálculos innecesarios.
if (nrow(ventas_clasificacion) == 0) {stop("No hay datos disponibles para la clasificación especificada.")}
if (nrow(ventas_clasificacion) == 0): Condición que verifica si hay alguna fila del dataset que esté vacía o que sea = 0.

{stop("No hay datos disponibles para la clasificación especificada.")}: Si la condición anterior se cumple, el programa mostrará un mensaje de error el cual dirá "No hay datos disponibles para la clasificación especificada."

Con todo lo necesario para que la base de datos esté lista para trabajar, se podrá continuar con la creación de la serie temporal.

ts_ventas <- ts(ventas_clasificacion$Ventas, start = c(year(min(ventas_clasificacion$Mes)), month(min(ventas_clasificacion$Mes))), frequency = 12)

ts_ventas: Nuevo objeto llamado ts_ventas el cual contiene una serie temporal con los valores de las ventas mensuales.

ts(ventas_clasificacion$Ventas: la columna ventas del dataset ventas_clasificación son los datos que se van a usar para la creación de la serie temporal.

start = c(year(min(ventas_clasificacion$Mes)): El argumento start ayuda a definir el valor inicial de la serie temporal, la función year(min……) nos ayuda a extraer el año de la fecha más antigua que tiene el dataset ventas_clasificación específicamente en la columna mes.

month(min(ventas_clasificacion$Mes))): La función month(min……) nos ayuda a extraer el mes de la fecha más antigua que tiene el dataset ventas_clasificación específicamente en la columna mes.

frequency = 12): Esta función define la frecuencia de la serie de tiempo, en este caso es mensual.

Luego de crear la serie de tiempo se procede a graficarla para poder entender de mejor manera el comportamiento de las ventas de Peugeot desde el 2018 hasta el 2023.

autoplot(ts_ventas) + labs(title = "Serie Temporal de Ventas", x = "Tiempo", y = "Ventas") + theme_minimal()

autoplot(ts_ventas): Autoplot es una función que está dentro del paquete ggplot2 la cual es encargada de generar gráficos basados en el objeto de datos de la serie de tiempo, se usó el objeto que contiene la serie temporal original (ts_ventas).

labs(title = "Serie Temporal de Ventas": La función labs es la encargada de modificar el título del gráfico, el título a agregar es: "Serie Temporal de Ventas".

x = "Tiempo", y = "Ventas"): La misma función labs es la encargada de agregarle un título tanto al eje x (tiempo) como al eje y (ventas).

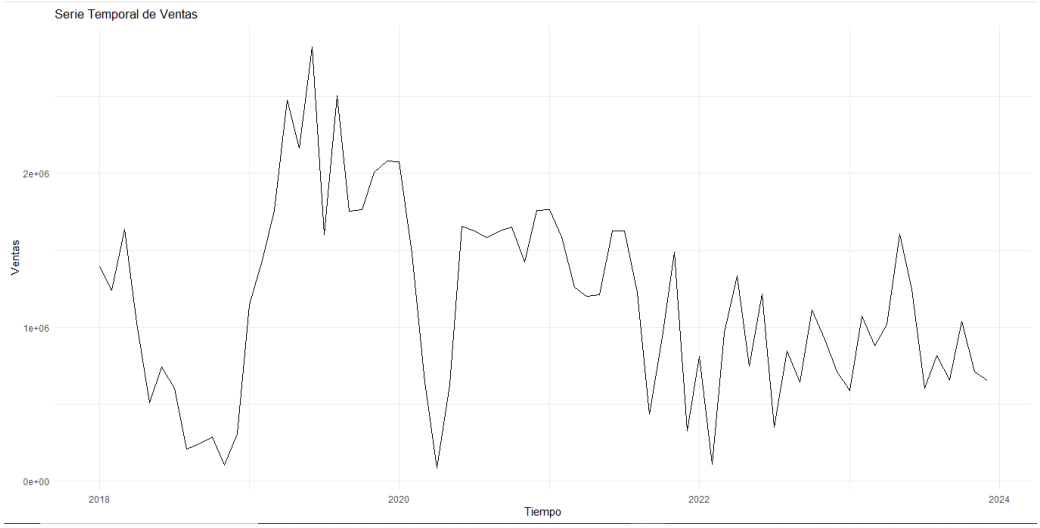
theme_minimal(): A lo anterior se le agrega un tema minimalista, lo cual le aporta un diseño más limpio y fácil de entender.

A continuación, se presentan las gráficas haciendo uso del filtro aplicado anteriormente, se pudieron visualizar las gráficas del segmento de clientes 1, 2 y 3

Gráfica segmento 1: Se puede visualizar que en el segmento 1, las ventas en dólares al inicio empezaron con una decaída, para luego tener su mayor pico entre el 2019 e inicios del 2020, para ya después decaer en el resto del 2020 seguramente como resultado de la pandemia, se pudo notar también una recuperación en el 2021 y una tendencia mantenida con picos pequeños hasta el 2023.

Figura 19

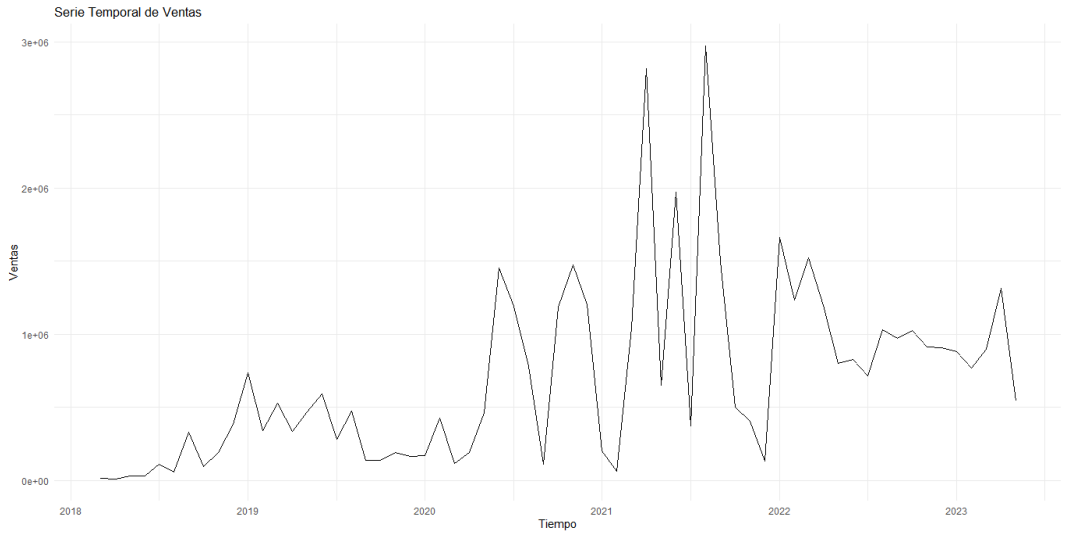
Comportamiento de precios del segmento 1 desde el 2018 hasta el 2023.



Gráfica segmento 2: En el segmento 2 en cambió se puede observar que desde el 2018 hasta inicios del 2020 las ventas en dólares se mantuvieron bajas, pero a partir de finales del 2020 hasta el 2022 hubo un pico en las ventas bastante alto, para luego volver a decaer a finales del 2022 hasta el 2023, pero a pesar de eso no alcanzaron números tan bajos como las del 2018-2019, también se pudo notar que el segmento 2 no alcanza ventas tan altas y por tiempos más prolongados como el segmento 1 y 3, lo que nos puede indicar que se trata del segmento de clientes de gama baja.

Figura 20

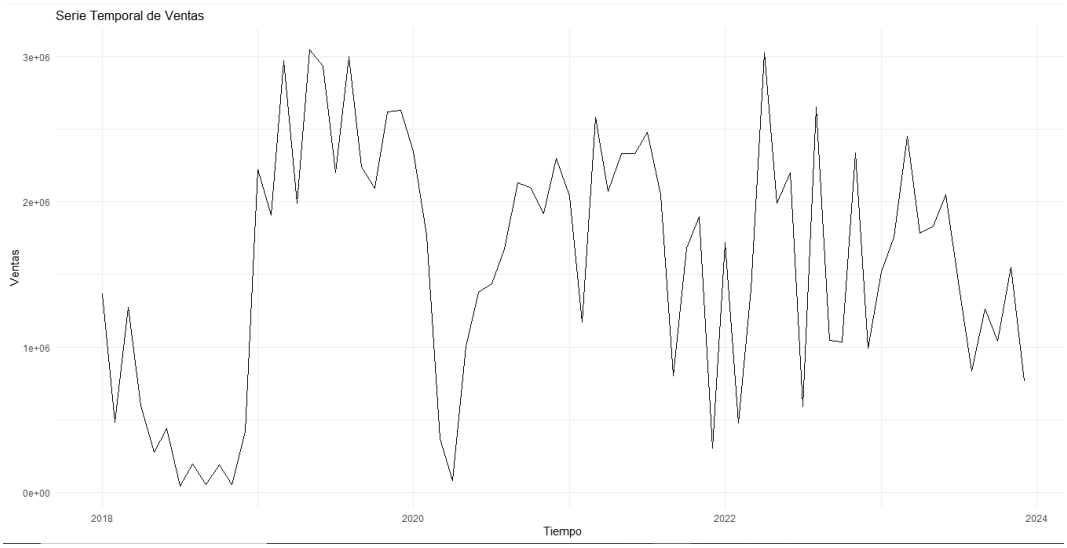
Comportamiento de precios del segmento 2 desde el 2018 hasta el 2023.



Gráfica segmento 3: Se puede notar que en el segmento de clientes 3 se tiene un alto nivel en las ventas en dólares, con una tendencia baja solo en el 2018, ya para el 2019 las ventas alcanzan un pico bastante alto, esta tendencia de precios altos se mantiene no de una manera completamente uniforme, pero sí se mantuvo con números altos hasta el 2023, pese a una baja significativa que hubo en el 2020, con esto se podría interpretar que se trata de una gama alta de clientes.

Figura 21

Comportamiento de precios del segmento 3 desde el 2018 hasta el 2023.



Después de ver las gráficas se procedió a ajustar el modelo ARIMA automáticamente, esto se hace porque el modelo ARIMA permite pronosticar de mejor manera los valores basándose en datos históricos, además de ser muy útil al momento de trabajar con tendencias, estacionalidad y ruidos blancos.

modelo_arima <- auto.arima(ts_ventas)

modelo_arima: objeto donde se almacenará el modelo ajustado ARIMA.

auto.arima(ts_ventas): Auto.arima es una función del paquete forecast el cual logra ajustar el modelo ARIMA a la serie de tiempo, la serie de tiempo la toma del objeto ts_ventas.

Luego, se hizo el resumen del modelo ARIMA

summary(modelo_arima)

Summary: es una función que va a permitir la visualización de los principales indicadores y coeficientes de evaluación del modelo

modelo_arima: es el objeto donde se almacenaron los datos del modelo ya ajustado por ARIMA.

Resultados:

Segmento 1:

ARIMA (1,1,0):

p=1: Es el orden de autoregresión, esto quiere decir que el modelo utiliza 1 periodo anterior para predecir.

d=1: Se refiere a la diferenciación, en este se usó 1 diferencia, esto se usa para hacer a la serie de tiempo estacionaria ya que ARIMA asume la estacionariedad de la misma.

q=0: Esta parte es el orden de media móvil del modelo, como tenemos q=0 quiere decir que el modelo no usó errores de previsión retardados.

Segmento 2:

ARIMA (0,1,1):

p=0: Es el orden de autoregresión, esto quiere decir que el modelo no utiliza en este segmento valores pasados para predecir.

d=1: Se refiere a la diferenciación, en este se usó 1 diferencia, esto se usa para hacer a la serie de tiempo estacionaria ya que ARIMA asume la estacionariedad de la misma.

q=1: Esta parte es el orden de media móvil del modelo, como tenemos q=1 quiere decir que el modelo usó el error de 1 periodo anterior.

Segmento 3:

ARIMA (3,0,1):

p=3: Es el orden de autoregresión, esto quiere decir que el modelo utiliza 3 periodos anteriores para predecir.

d=0: Se refiere a la diferenciación, en no se usaron diferencias lo que quiere decir que la serie probablemente era estacionaria, esto es necesario ya que ARIMA asume la estacionariedad de la misma.

q=1: Esta parte es el orden de media móvil del modelo, como tenemos q=1 quiere decir que el modelo usó el error de 1 periodo anterior.

Training set error measures:

Training set

Los indicadores presentados sirven para medir el error de modelo en el set de entrenamiento, por lo tanto al revisar el RMSE (Error Cuadrático Medio) se puede saber la precisión que tiene el modelo desarrollado, su función es la de medir cuánto se desvía el valor que se predijo tomando en cuenta los valores reales presentados en el data set.

Segmento 1:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-12851.2	465101.6	360717.4	-36.60098	63.54935	0.5229929	-0.001630353

ME (Mean Error): El error medio es el promedio de los errores del segmento 1 en este caso, desafortunadamente no es tan preciso o recomendable de usar debido a que toma en cuenta las sobreestimaciones (errores positivos) y subestimaciones (errores negativos), el modelo puede cometer errores grandes pero con signos opuestos y esta

métrica de evaluación de errores los anularía. En este segmento se tiene un ME de -12851.2 lo que quiere decir que en el modelo se tiene tendencia a subestimar los valores.

RMSE (Root Mean Square Error): La raíz del error cuadrático medio arregla la cancelación de errores por lo tanto es más recomendable de usar para medir el error del modelo, la ventaja es que se mide en las mismas unidades que la variable a predecir. En este caso se quiere predecir las ventas de autos Peugeot en dólares, se tiene un error cuadrático medio de 465101.6 lo cual indica que el modelo no es tan preciso en el segmento 1.

MAE (Mean Absolute Error): El error absoluto medio usa el valor absoluto de los errores lo cual soluciona el problema del error absoluto. En este caso el error absoluto medio es 360717.4 el cual también indica que el modelo no tiene tanta precisión pero sí mejoró en comparación a la raíz del error cuadrático medio.

MPE (Mean Percentage Error): El error porcentual medio mide el promedio de errores pero a manera de porcentaje, en el segmento 1 existe un -36.60098% de error entre los valores reales y los predichos, es decir que existe una tendencia a subestimar los valores reales.

MAPE (Mean Absolute Percentage Error): El error porcentual absoluto medio mide el promedio de errores a manera de porcentaje pero toma el valor absoluto de los extremos de los datos y no toma en cuenta los ceros. En este segmento el error porcentual absoluto medio es de 63.54935 lo cual en porcentaje es bastante alto ya que el modelo en este segmento tiene más del 50% de error.

MASE (Mean Absolute Scaled Error): Este indicador compara el error absoluto medio del modelo con otro modelo base simple, como en este caso el valor es de 0.5229929, es menor a 1 quiere decir que el modelo es competitivo.

ACF1: En este caso tenemos un valor cercano a 0 (-0.001630353) lo que indica que los errores del modelo se correlacionan bien con los valores anteriores, esto significa que se ha capturado bien los patrones.

Segmento 2:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
77513.35	554377.2	360658.4	-49.39362	93.16265	0.5358038	-0.0115087

ME (Mean Error): El error medio es el promedio de los errores del segmento 2 en este caso, desafortunadamente no es tan preciso o recomendable de usar debido a que toma en cuenta las sobreestimaciones (errores positivos) y subestimaciones (errores negativos), el modelo puede cometer errores grandes pero con signos opuestos y esta métrica de evaluación de errores los anularía. En este segmento se tiene un ME de 77513.35 lo que quiere decir que en el modelo se tiene tendencia a sobrestimar los valores.

RMSE (Root Mean Square Error): La raíz del error cuadrático medio arregla la cancelación de errores por lo tanto es más recomendable de usar para medir el error del modelo, la ventaja es que se mide en las mismas unidades que la variable a predecir. En este caso se quiere predecir las ventas de autos Peugeot en dólares, se tiene un error cuadrático medio de 554377.2 que es mayor al segmento 1 lo cual sugiere que el modelo sigue sin ser preciso ya que el error es alto considerando en promedio los valores de este segmento.

MAE (Mean Absolute Error): El error absoluto medio usa el valor absoluto de los errores lo cual soluciona el problema del error absoluto. En este caso el error absoluto medio es 360658.4 el cual también indica que el modelo no tiene tanta precisión pero sí mejoró en comparación a la raíz del error cuadrático medio.

MPE (Mean Percentage Error): El error porcentual medio mide el promedio de errores pero a manera de porcentaje, en el segmento 2 existe un -49.39362% de error entre los valores reales y los predichos, es decir que existe una tendencia a subestimar los valores reales.

MAPE (Mean Absolute Percentage Error): El error porcentual absoluto medio mide el promedio de errores a manera de porcentaje pero toma el valor absoluto de los extremos de los datos y no toma en cuenta los ceros. En este segmento el error porcentual absoluto medio es de 93.16265 lo cual en porcentaje es bastante alto ya que el modelo en este segmento tiene casi un 100% de error según este indicador.

MASE (Mean Absolute Scaled Error): Este indicador compara el error absoluto medio del modelo con otro modelo base simple, como en este caso el valor es de 0.5358038, es decir, es menor a 1 lo cual indica que el modelo es competitivo.

ACF1: En este caso tenemos un valor cercano a 0 (-0.0115087) lo que indica que los errores del modelo se correlacionan bien con los valores anteriores, esto significa que se ha capturado bien los patrones.

Segmento 3:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
7949.425	656752.3	528446.1	-94.47555	118.8982	0.4888368	0.04601699

ME (Mean Error): El error medio es el promedio de los errores del segmento 3 en este caso, desafortunadamente no es tan preciso o recomendable de usar debido a que toma en cuenta las sobreestimaciones (errores positivos) y subestimaciones (errores negativos), el modelo puede cometer errores grandes pero con signos opuestos y esta métrica de evaluación de errores los anularía. En este segmento se tiene un ME de 7949.425 lo que quiere decir que en el modelo se tiene tendencia a sobrestimar los valores.

RMSE (Root Mean Square Error): La raíz del error cuadrático medio arregla la cancelación de errores por lo tanto es más recomendable de usar para medir el error del modelo, la ventaja es que se mide en las mismas unidades que la variable a predecir. En este caso se quiere predecir las ventas de autos Peugeot en dólares, se tiene un error cuadrático medio de 656752.3 que es mayor al segmento 2 lo cual sugiere que el modelo sigue sin ser preciso ya que el error es alto considerando en promedio los valores de este segmento.

MAE (Mean Absolute Error): El error absoluto medio usa el valor absoluto de los errores lo cual soluciona el problema del error absoluto. En este caso el error absoluto medio es 528446.1 el cual también indica que el modelo no tiene tanta precisión pero sí mejoró en comparación a la raíz del error cuadrático medio.

MPE (Mean Percentage Error): El error porcentual medio mide el promedio de errores pero a manera de porcentaje, en el segmento 3 existe un -94.47555% de error entre los

valores reales y los predichos, es decir que existe una tendencia bastante alta a subestimar los valores reales.

MAPE (Mean Absolute Percentage Error): El error porcentual absoluto medio mide el promedio de errores a manera de porcentaje pero toma el valor absoluto de los extremos de los datos y no toma en cuenta los ceros. En este segmento el error porcentual absoluto medio es de 118.8982 lo cual en porcentaje es bastante alto de error.

MASE (Mean Absolute Scaled Error): Este indicador compara el error absoluto medio del modelo con otro modelo base simple, como en este caso el valor es de 0.4888368, es decir, es menor a 1 lo cual indica que el modelo es competitivo.

ACF1: En este caso tenemos un valor cercano a 0 (0.04601699) lo que indica que los errores del modelo se correlacionan bien con los valores anteriores, esto significa que se ha capturado bien los patrones.

El siguiente paso es la verificación de los residuos del modelo, por lo que va a desarrollar el siguiente comando

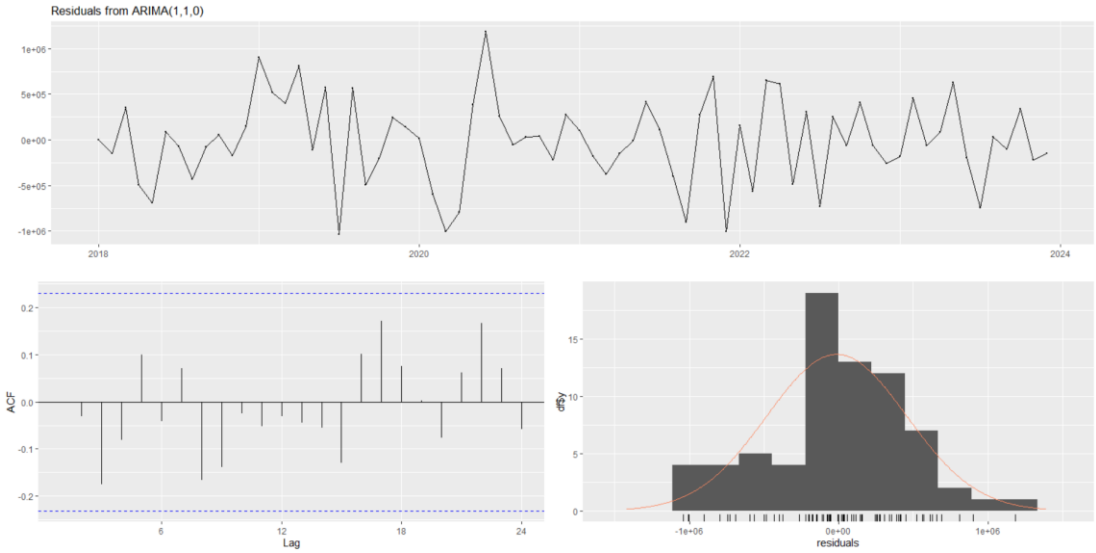
checkresiduals(modelo_arima)

Checkresiduals: es una función cuyo objetivo es visualizar en forma de tablas los residuos del modelo desarrollado.

Modelo_arima: es el objeto donde se almacenó los datos del modelo ya ajustado por ARIMA.

Figura 22

Residuos representados a lo largo del tiempo.



En esta gráfica se representa la diferencia que existe entre los valores reales y los valores que se predijeron mostrados en el tiempo que se está evaluando, en este caso son los años entre 2018 y 2024. También este gráfico es muy útil para verificar la existencia de ruido blanco, es decir si existen patrones temporales ya que un modelo bien desarrollado es necesario que no tenga tendencias o estacionalidades.

Gráfica 2: Función de autocorrelación de los residuos

Esta gráfica mide la correlación existente entre los residuos y el número de periodos que se han estudiado en el pasado. El objetivo es saber si los residuos obtenidos son independientes del tiempo evaluado, si el modelo está bien desarrollado las líneas deben permanecer dentro del espacio que forma las líneas azules, ya que así se indica que no existe una correlación significativa, caso contrario si las líneas salen de los líneas azules el modelo podría mejorarse.

Gráfico 3: Histograma de los residuos con curva de densidad

El histograma de los residuos sirve para ver si siguen una distribución normal, si la forma del histograma y la curva de densidad se alinean bien a la campana de Gauss los residuos tienen una tendencia a la normalidad. El gráfico no muestra asimetría o colar largas a los costados por lo que no se afecta la calidad del modelo desarrollado.

El siguiente paso es el desarrollo del pronóstico con ARIMA

*horizonte <- (2026 - year(max(ventas_clasificacion$Mes))) * 12*

Horizonte: es la variable que contiene el número total de meses que se van a predecir, entre el último dato disponible en el data set y el año de pronóstico.

2026-year(max(ventas_clasificación$Mes))*12: es el comando que va a desarrollar el pronóstico al año 2026 tomando como referencia los datos de la clasificación de las ventas.

forecast_arima <- forecast(modelo_arima, h = horizonte)

forecast_arima: es el vector que va a contener el resultado del pronóstico con ARIMA

forecast: es la función de pronóstico

Modelo_arima: es el objeto donde se almacenó los datos del modelo ya ajustado por ARIMA.

h=horizonte: es la variable que contiene el periodo de tiempo que se va a pronosticar.

Gráfico del pronóstico usando ARIMA

autoplot(forecast_arima) + labs(title = paste("Pronóstico ARIMA para Clasificación", clasificacion_filtro), x = "Mes", y = "Ventas") + theme_minimal()

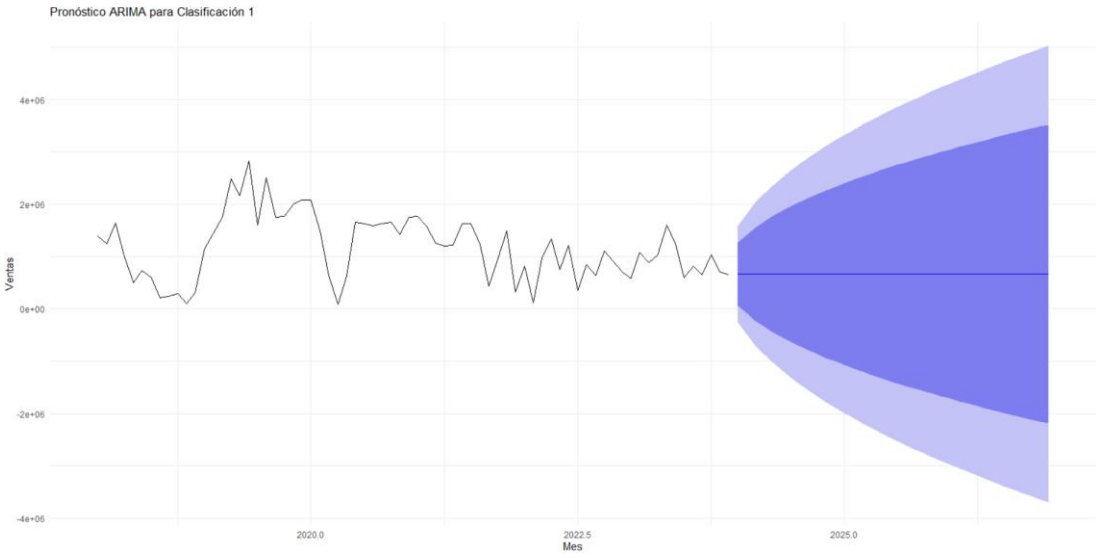
autoplot: en una función de ggplot2 que va a permitir la visualización de los resultados del pronóstico usando ARIMA.

forecast_arima: es el vector que va a contener el resultado del pronóstico con ARIMA

labs: función que permite la configuración del título y las variables “x” y “y” en el gráfica.

Figura 23

Gráfica del pronóstico con ARIMA.



Resultados:

En la gráfica se puede ver que al utilizarse los datos de los años anteriores 2018-2024, al hacer la parte del pronóstico el modelo ARIMA no tuvo un resultado favorable ya que a partir del año 2025 el modelo no respondió dando como resultado una línea recta, esto se debe a lo que los datos usados para la predicción tienen un alta variabilidad o ruido, por lo que este modelo queda sin utilidad.

Modelo ETS (Error, Trend, Seasonality)

modelo_ets <- ets(ts_ventas)

Modelo_ets: objeto donde se almacenará el modelo ajustado ETS.

ets(ts_ventas): ETS es una función del paquete forecast el cual logra ajustar el modelo ETS a la serie de tiempo, la serie de tiempo la toma del objeto ts_ventas.

summary(modelo_ets)

Summary: es una función que va a permitir la visualización de los principales indicadores y coeficientes de evaluación del modelo

Modelo_ets: objeto donde se almacenará el modelo ajustado ETS.

Resultados:

Training set error measures:

Training set

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
18883.73	616811.3	458634.1	-79.60357	109.8924	0.5164403	-0.03366493

Luego se va a desarrollar el forecast con el modelo ETS

forecast_ets <- forecast(modelo_ets, h = horizonte)

forecast_est: es el vector que va a contener el resultado del pronóstico con ETS

forecast: es la función de pronóstico

Modelo_ets: es el objeto donde se almacenó los datos del modelo ya ajustado por ETS.

h=horizonte: es la variable que contiene el periodo de tiempo que se va a pronosticar.

Gráfico del pronóstico usando ETS

autoplot(forecast_ets) + labs(title = paste("Pronóstico ETS para Clasificación", clasificacion_filtro), x = "Mes", y = "Ventas") +theme_minimal()

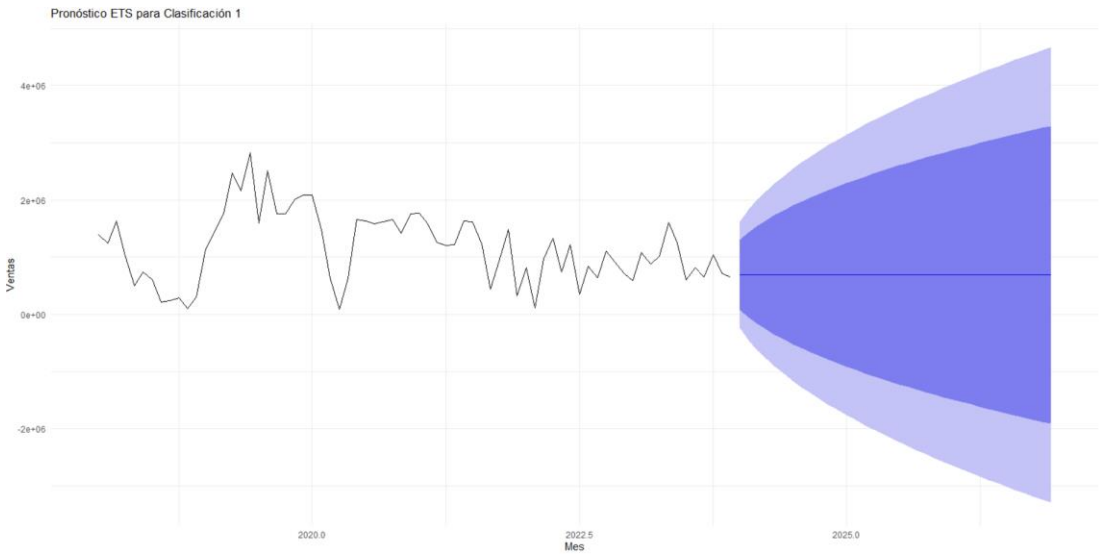
autoplot: en una función de ggplot2 que va a permitir la visualización de los resultados del pronóstico usando ETS.

forecast_ets: es el vector que va a contener el resultado del pronóstico con ETS

labs: función que permite la configuración del título y las variables “x” y “y” en el gráfico.

Figura 24

Gráfica del pronóstico con ETS.



Resultados:

En la gráfica se puede ver que al utilizarse los datos de los años anteriores 2018-2024, al hacer la parte del pronóstico el modelo ETS no tuvo un resultado favorable ya que a partir del año 2025 el modelo no respondió dando como resultado una línea recta, tal como ya había pasado anteriormente con el modelo ARIMA.

Descomposición de la serie temporal

descomposicion <- decompose(ts_ventas)

Descomposición: variable que va a almacenar la descomposición de la serie de tiempo

Descompose: función que permite descomponer una serie de tiempo en su estacionalidad, tendencia y ruido.

ts_ventas: es el objeto que contiene la serie de tiempo

autoplot(descomposicion) + labs(title = "Descomposición de la Serie Temporal")

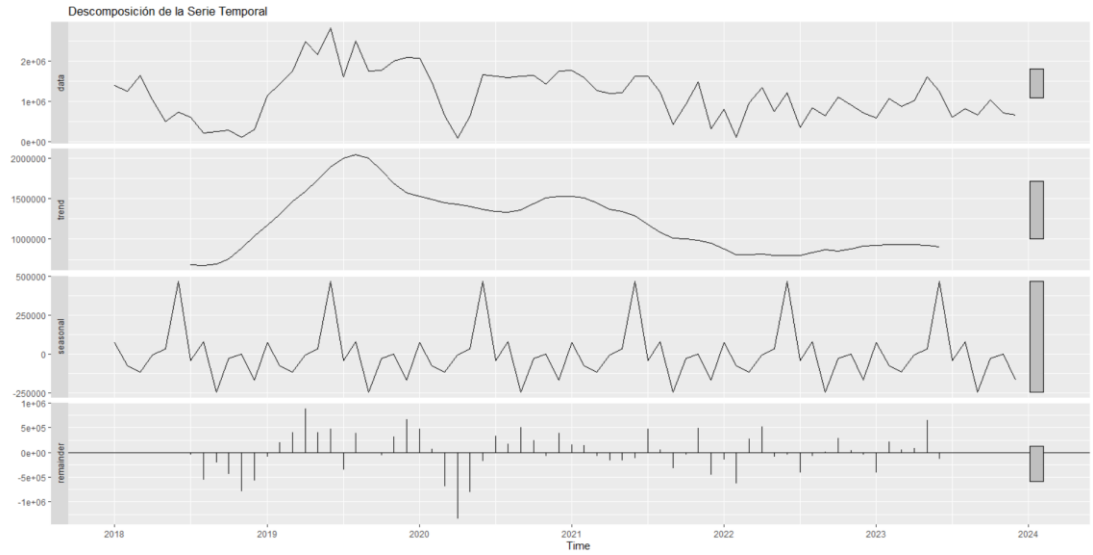
autoplot: en una función de ggplot2 que va a permitir la visualización de los resultados de la descomposición

Descomposición: variable que va a almacenar la descomposición de la serie de tiempo

labs: función que permite la configuración del título del gráfico.

Figura 25

Tendencia, estacionalidad y ruido de la serie temporal del segmento 1.



Resultados:

En la gráfica se muestran la descomposición de la serie de tiempo en su tendencia, estacionalidad, ruido. El primer gráfico muestra la serie temporal original.

El siguiente gráfico es el de tendencia que describe los datos a largo plazo, ayuda a identificar patrones como crecimientos o caídas en los datos que se están estudiando.

El tercer gráfico es la estacionalidad, los mismos que muestran los ciclos que se repiten a lo largo del tiempo, que están asociadas a algunos meses del año específicos, estaciones o eventos que son recurrentes.

El cuarto gráfico corresponde a los residuales o ruidos, esta evalúa que tan bien la tendencia y estacionalidad pueden explicar la serie estudiada.

Ajuste manual del modelo ARIMA

modelo_manual <- arima(ts_ventas, order = c(1, 1, 1), seasonal = c(1, 1, 1))

Modelo_manual: objeto que va a almacenar el modelo manual

Arima: es la función que se usará para pronosticar con ARIMA

ts_ventas: es el objeto que contiene la serie de tiempo

order=c(1,1,1): este es el componente no estacional del modelo ARIMA. Cada número representa el factor autoregresivo, differencing y medias móviles.

seasonal=c(1,1,1): este esl componente estacional del modelo ARIMA, que toma en cuenta el autoregresivo estacional, differencing estacional y las medias móviles estacionales.

summary(modelo_manual)

Summary: es una función que va a permitir la visualización de los principales indicadores y coeficientes de evaluación del modelo

Modelo_manual: objeto que va a almacenar el modelo manual

Luego se va a desarrollar el forecast con el modelo manual

forecast_manual <- forecast(modelo_manual, h = horizonte)

forecast_manual: es el vector que va a contener el resultado del pronóstico con el modelo manual

forecast: es la función de pronóstico

Modelo_manual: es el objeto donde se almacenó los datos del modelo ya ajustado por ETS.

h=horizonte: es la variable que contiene el periodo de tiempo que se va a pronosticar.

Gráfico del pronóstico usando el modelo Manual

autoplot(forecast_manual) + labs(title = paste("Pronóstico ARIMA Ajustado Manualmente para Clasificación", clasificacion_filtro), x = "Mes", y = "Ventas") + theme_minimal()

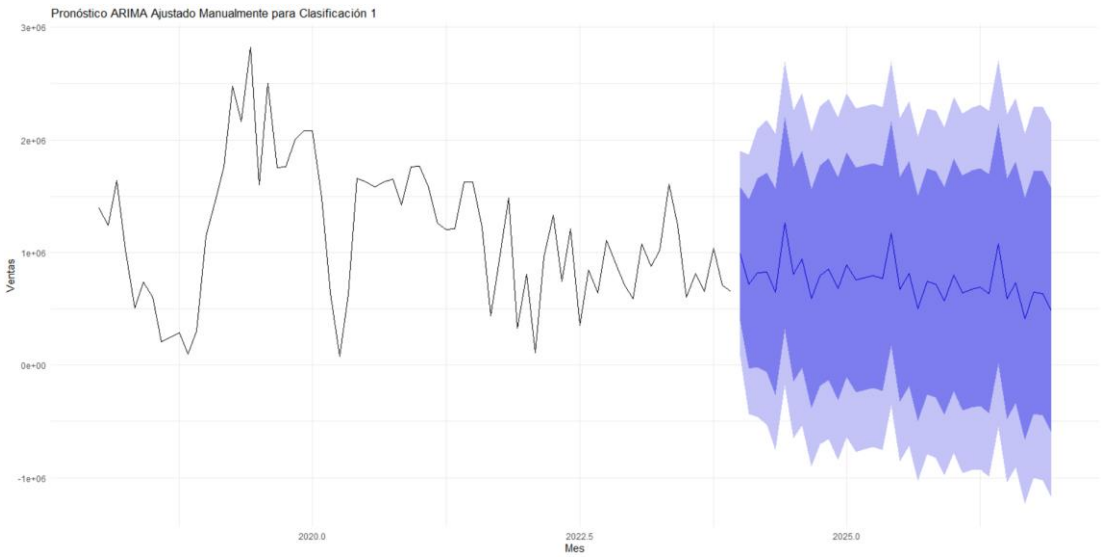
autoplot: en una función de ggplot2 que va a permitir la visualización de los resultados del pronóstico usando el modelo Manual.

forecast_manual: es el vector que va a contener el resultado del pronóstico con el modelo Manual

labs: función que permite la configuración del título y las variables “x” y “y” en el gráfico.

Figura 26

Gráfica del pronóstico usando el modelo ARIMA ajustado manualmente.



Resultados:

En la gráfica se puede ver que al utilizarse los datos de los años anteriores 2018-2024, al hacer la parte del pronóstico el modelo Manual tuvo un resultado favorable ya que a partir del año 2025 el modelo si respondió dando como resultado un pronóstico más factible de la serie de tiempo para los años que se tenían que pronosticar.

Comparación de los modelos

Precisión del modelo ARIMA

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-12851.2	465101.6	360717.4	-36.60098	63.54935	0.5229929	-0.001630353

Precisión del modelo ETS

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-3518.323	469135.3	469135.3	-38.48198	65.75669	0.5321095	0.003832902

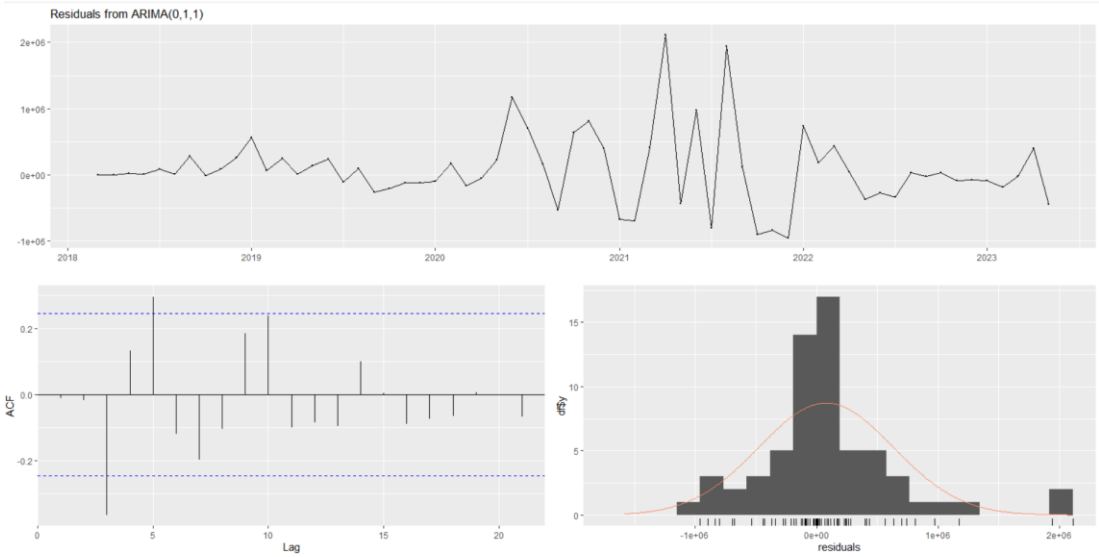
Precisión del modelo Manual

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-43785.52	374757	275907.4	-29.24484	46.23954	0.7119178	0.00076578

Pronóstico de la Clasificación 2

Figura 27

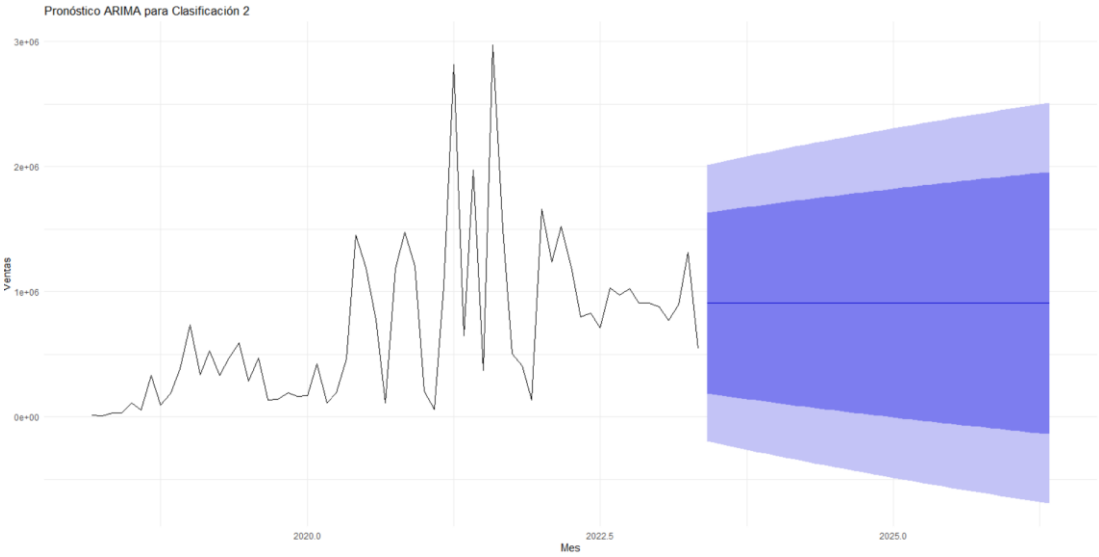
Residuos modelo ARIMA del segmento 2.



Pronóstico de la Clasificación 2 con ARIMA

Figura 28

Pronóstico con ARIMA del segmento 2.



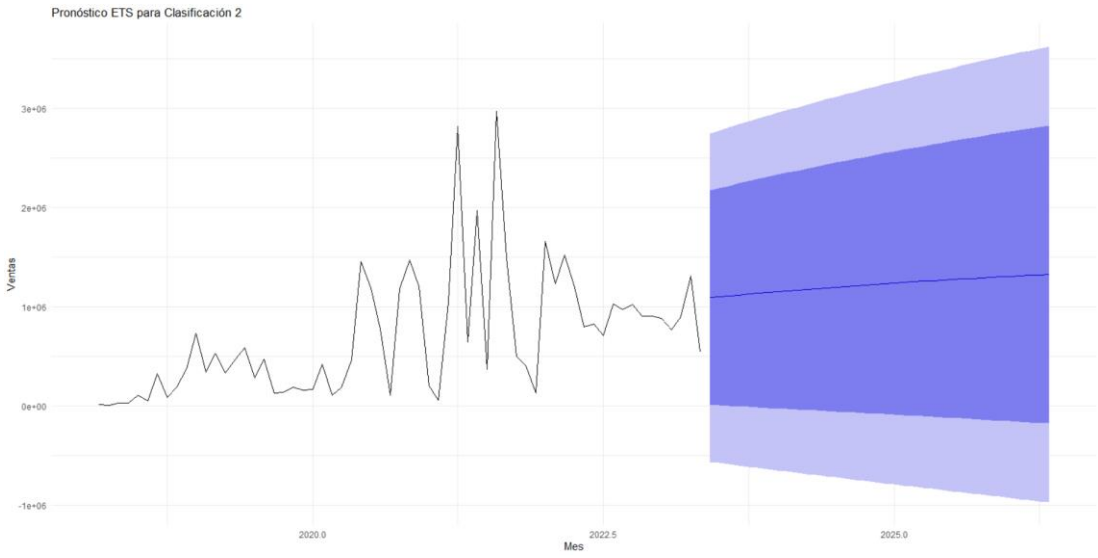
En la gráfica se puede ver que al utilizarse los datos de los años anteriores 2018-2024 para la clasificación 2 al hacer la parte del pronóstico el modelo ARIMA no tuvo un resultado favorable ya que a partir del año 2025 el modelo no respondió dando como

resultado una línea recta, esto se debe a lo que los datos usados para la predicción tienen un alta variabilidad o ruido, por lo que este modelo queda sin utilidad.

Pronóstico de la Clasificación 2 con ETS

Figura 29

Pronóstico con ETS del segmento 2.

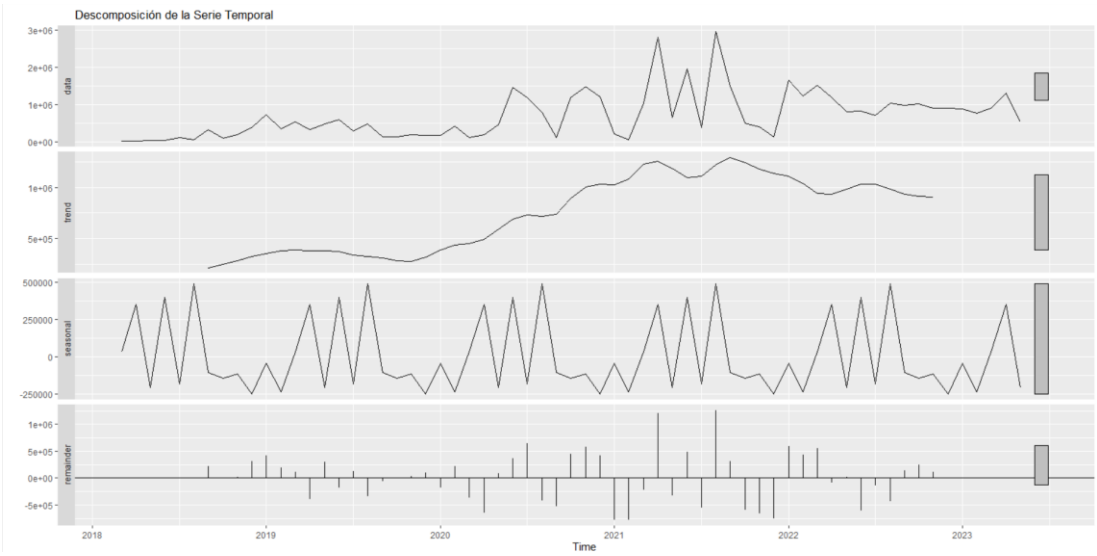


En la gráfica se puede ver que al utilizarse los datos de los años anteriores 2018-2024 para la clasificación 2, al hacer la parte del pronóstico el modelo ETS no tuvo un resultado favorable ya que a partir del año 2025 el modelo no respondió dando como resultado una línea recta, tal como ya había pasado anteriormente con el modelo ARIMA.

Descomposición de la serie temporal

Figura 30

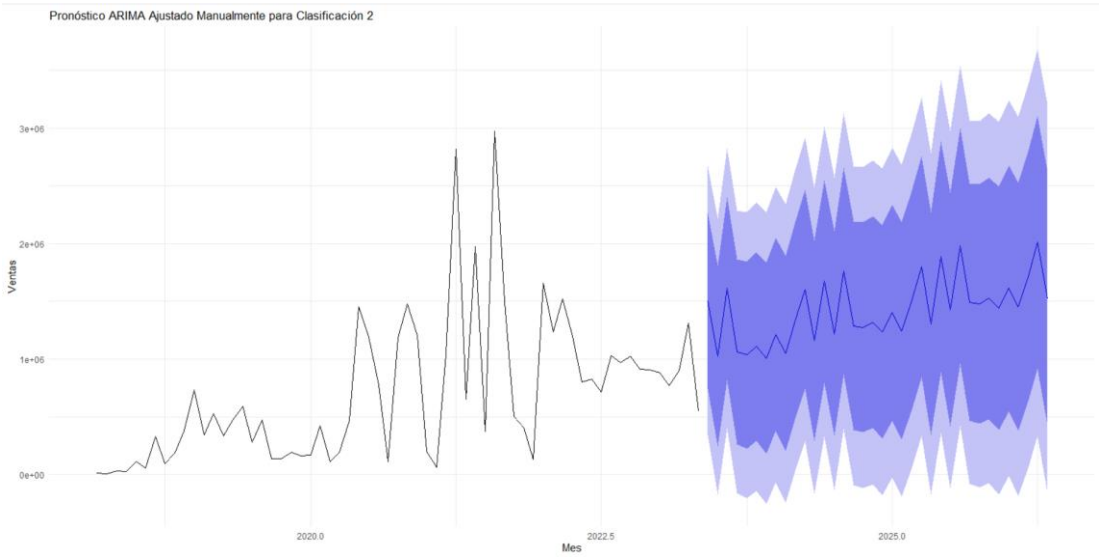
Tendencia, estacionalidad y ruido de la serie temporal del segmento 2.



Pronóstico Manual de la clasificación 2

Figura 31

Pronóstico de ARIMA ajustado manualmente del segmento 2.



Comparación de los modelos de la clasificación 2

Precisión del modelo ARIMA

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
77513.35	554377.2	360658.4	-49.39362	93.16265	0.5358038	-0.0115087

Precisión del modelo ETS

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-16241.27	539439.2	372762.9	-79.0357	115.2754	0.5537866	0.0489522

Precisión del modelo Manual

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-50213.52	478212.3	310565.4	-67.0765	83.46983	0.6949952	-0.01415488

Pronóstico de la Clasificación 3

checkresiduals(modelo_arima)

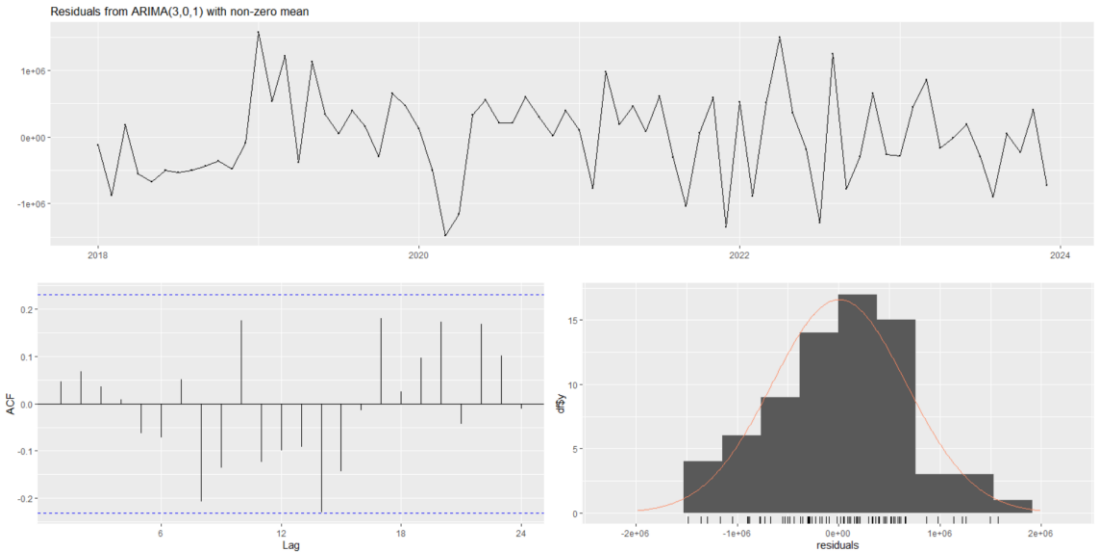
Checkresiduals: es una función cuyo objetivo es visualizar en forma de tablas los residuos del modelo desarrollado.

Modelo_arima: es el objeto donde se almacenó los datos del modelo ya ajustado por ARIMA.

Resultados:

Figura 32

Residuos a lo largo del tiempo del segmento 3.



En esta gráfica se representa la diferencia que existe entre los valores reales y los valores que se predijeron mostrados en el tiempo que se está evaluando.

Gráfica 2: Función de autocorrelación de los residuos

Esta gráfica mide la correlación existente entre los residuos y el número de periodos que se han estudiado en el pasado. En este caso los picos siguen dentro las líneas azules, por lo que el modelo si tiene buenos resultados.

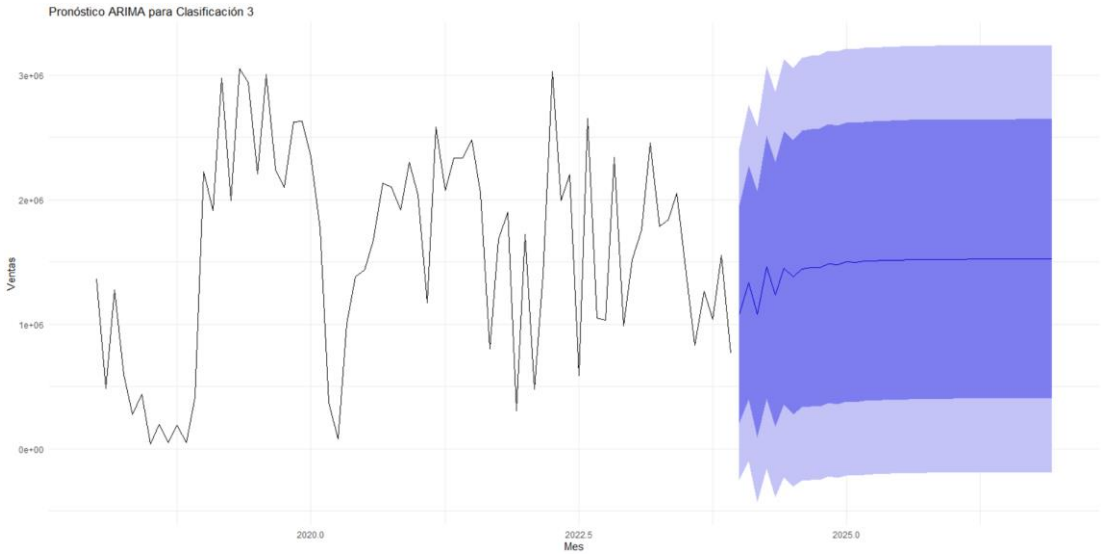
Gráfico 3: Histograma de los residuos con curva de densidad

El histograma de los residuos sirve para ver si siguen una distribución normal, en esta gráfica se puede observar que hay myaor concentración de los residuos se encuentran cerca del valor 0, lo cual es un indicador positivo ya que quiere decir que los errores están cercanos a la media lo que se espera de un buen modelo. La asimetría del gráfico muestra que hay un sesgo hacia la derecha lo que puede representar que existe un problema con la normalidad de los residuos.

Pronóstico de la clasificación 3 con ARIMA

Figura 33

Pronóstico con el modelo ARIMA del segmento 3.

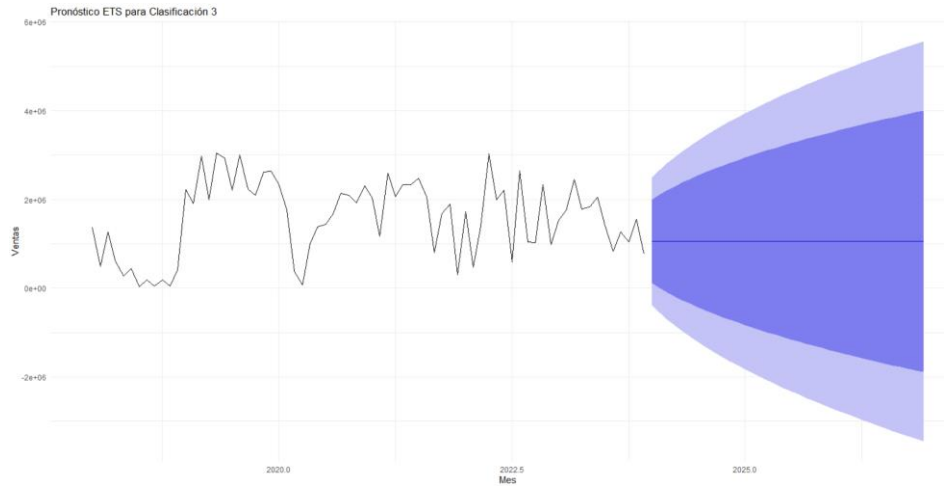


En la gráfica se puede ver que al utilizarse los datos de los años anteriores 2018-2024 para la clasificación 3 al hacer la parte del pronóstico el modelo ARIMA no tuvo un resultado favorable ya que a partir del año 2025 el modelo no respondió dando como resultado una línea recta, esto se debe a lo que los datos usados para la predicción tienen un alta variabilidad o ruido, por lo que este modelo queda sin utilidad.

Pronóstico de la clasificación 3 con ETS

Figura 34

Pronóstico con el modelo ETS del segmento 3.

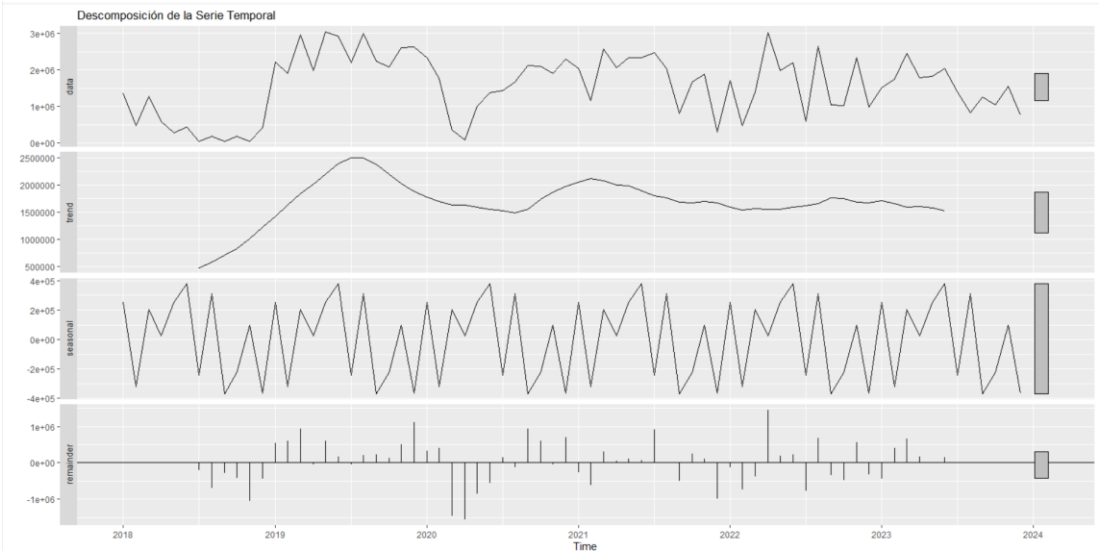


En la gráfica se puede ver que al utilizarse los datos de los años anteriores 2018-2024 para la clasificación 3, al hacer la parte del pronóstico el modelo ETS no tuvo un resultado favorable ya que a partir del año 2025 el modelo no respondió dando como resultado una línea recta, tal como ya había pasado anteriormente con el modelo ARIMA.

Descomposición de la serie temporal

Figura 35

Tendencia, estacionalidad y ruido de la serie temporal del segmento 3.



Resultados:

En la gráfica se muestran la descomposición de la serie de tiempo en su tendencia, estacionalidad, ruido. El primer gráfico muestra la serie temporal original de la clasificación 3.

El siguiente gráfico es el de tendencia que describe los datos a largo plazo, ayuda a identificar patrones como crecimientos o caídas en los datos que se están estudiando.

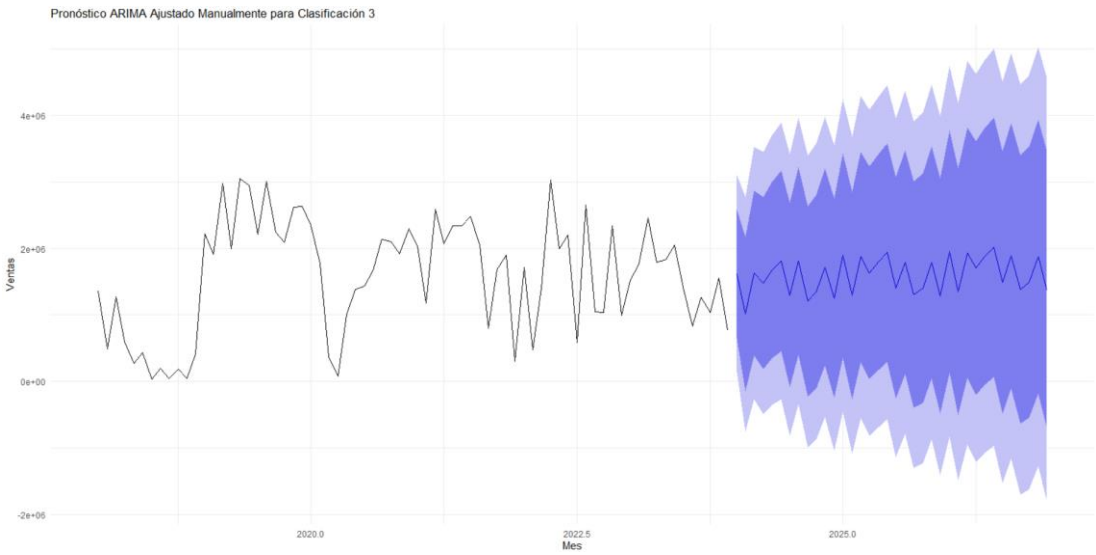
El tercer gráfico es la estacionalidad, los mismos que muestran los ciclos que se repiten a lo largo del tiempo, que están asociadas a algunos meses del año específicos, estaciones o eventos que son recurrentes.

El cuarto gráfico corresponde a los residuales o ruidos, esta evalúa que tan bien la tendencia y estacionalidad pueden explicar la serie estudiada.

Pronóstico Manual de la clasificación 3

Figura 36

Pronóstico con el modelo ARIMA ajustado manualmente del segmento 3.



Resultados:

En la gráfica se puede ver que al utilizarse los datos de los años anteriores 2018-2024 para la clasificación 3, al hacer la parte del pronóstico el modelo Manual tuvo un resultado favorable ya que a partir del año 2025 el modelo si respondió dando como resultado un pronóstico más factible de la serie de tiempo para los años que se tenían que pronosticar. Las sombras azules representan los intervalos de confianza, el de color más oscuro tiene un intervalo de confianza más estrecho, es decir de menor confianza, la capa más clara representa una confianza más amplia, lo que indica que el pronóstico de las ventas se encuentran dentro de esa región del gráfico. Por lo tanto, el modelo pronostica que las ventas continuarán con fluctuaciones pero en un rango predecible dentro de las sombras de los intervalos de confianza.

Comparación de los modelos de la clasificación 3

Precisión del modelo ARIMA

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
7949.425	656752.3	528446.1	-94.47555	118.8982	0.4888368	0.04601699

Precisión del modelo ETS

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
14214.09	720301	553442.5	-66.4967	95.85323	0.5119597	-0.04929636

Precisión del modelo Manual

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-99428.7	624440.2	411504.5	-34.9802	48.61469	0.6419472	-0.1361095

Resultados:

De acuerdo al desarrollo de los modelos K-means y Forecasting se obtuvieron los siguientes resultados:

Para garantizar un buen resultado de la clasificación se utilizaron tres métodos diferentes para determinar el número óptimo de clústeres. El primer método usado fue el Silhoutte, el cual evalúa la cohesión y separación de los clústeres. El segundo es el WSS (Within-cluster sum of squares), en español Suma de cuadrados intraclústeres, el mismo mide la compacidad de los clústeres, es decir que tan compacto son los

clústeres. Finalmente, el método de Gap Statistic que compara la dispersión que existe dentro de cada clúster con lo que se espera por una distribución aleatoria. El resultado de estos métodos fue que el número óptimo de clusters es 3.

La interpretación de las 3 clasificaciones es que cada clúster representa un segmento de comportamiento de ventas y que gracias a las diferencias existentes entre cada uno de los grupos se puede visualizar los 3 clusters bien definidos, es decir que los clientes de Peugeot están bien delimitados por sus características, aunque haya presencia de datos atípicos que pertenecen a cada clúster dependiendo de su proximidad al centroide.

Para el modelo de Forecasting se utilizó el modelo ARIMA para obtener las proyecciones de las ventas en dólares mensuales para los años 2025 y 2026 y se comparó con otras metodologías.

Para el desarrollo del Forescasting se evaluaron 3 modelos, el ARIMA, ETS (Error-Trend-seasonal) y el ARIMA con ajuste manual, cabe recalcar que los dos primeros modelos son modelos de machine learning en el que se usa la Inteligencia Artificial para generar los resultados, mientras que el modelo de ARIMA con ajuste manual es un proceso netamente estadístico, este detalle es importante de considerarse para la comparación de los resultados que se van a explicar a continuación.

Para la validación de los 3 modelos se tomaron en cuenta los indicadores de desempeño como son el RMSE y el MAE, el RMSE (Root Mean Squared Error) busca mostrar la precisión del ajuste de las observaciones en comparación con los datos históricos, mientras que el MAE (Mean Absolute Error) es el error absoluto promedio que ayuda a la detección de errores más amplios relacionados con la proyección.

Este resultado tiene como particularidad que los dos modelos que se hicieron mediante Machine Learning no tuvieron proyecciones acertadas, mientras que el modelo estadístico de ARIMA Manual es el que mejor se ajustó a los datos que fueron evaluados, dando a entender que dependiente de los datos que se tienen y que se van a estudiar muchas veces la inteligencia artificial no es viable y que modelos estadísticos son más precisos para el análisis que posteriormente va a ayudar a la implementación de estrategias de ventas que está diferenciadas para cada uno de los segmentos que se

obtuvieron, además de planificar recursos y presupuestos de acuerdo a las proyecciones obtenidas.

Por otro lado, los pronósticos hechos por segmento muestran comportamientos diferentes dependiendo del mercado al que pertenecen. Además, este segmento de mercado 1 representativo de la clase media presencia fluctuaciones con picos más marcados y que tienen tendencia a la baja. Esto se presenta de la siguiente manera: lo más bajo del pronóstico es 500.000 y lo más alto 1.200.000 en ventas, sin embargo los indicadores de confianza del 80% muestran ventas entre 1.500.000 y 2.100.000, mientras que el intervalo de confianza del 95% muestra un punto más en 1.800.000 y uno más alto en 2.600.000

El segmento de mercado 2, que corresponde a los consumidores de economía baja, presenta un valor de ventas inferior que oscilan entre 1.000.000 y los 1.500.000 con algunos picos en 2.000.000, de acuerdo a los intervalos de confianza estudiadas los niveles de ventas pueden variar 2.000.000 y 3.000.000 en sus puntos más bajos y altos de fluctuación respectivamente, esto con un intervalo de confianza del 80%, sin embargo si se toma como referencia la confianza del 95% cambian los puntos y varían entre 2.300.000 y 3.100.000 como puntos inferiores y superiores. El crecimiento de este grupo es una buena oportunidad para la innovación de productos de gama económica y generar oportunidades de volumen para que de esta forma los ingresos se impulsen no por la compra de productos más caros sino por la cantidad de autos económicos que se vendan.

El segmento 3 correspondiente a la categoría de autos de gama alta muestra un valor de ventas esperado de aproximadamente entre los 1.000.000 y los 2.000.000 de ventas mensuales, con intervalos de confianza entre los 80% y 95%, que indican que el nivel más alto de las ventas puede ser hasta de 3.000.000 con una confianza de 80% y entre 4.000.000 y 5.000.000 con una confianza del 95%. El segmento 3 muestra un comportamiento no tan variante como los otros dos segmentos antes analizados, por lo que se espera que las ventas se mantengan con la misma incidencia que los años posteriores y no tenga un cambio drástico. Esto se debe a que los consumidores de este grupo en específico tienen preferencia a las características de los automóviles de gama alta.

Gracias a los resultados obtenidos por medio del uso y ejecución de los algoritmos de K-means y Forecasting, se obtuvieron la clasificación de mercados de la marca Peugeot en Ecuador, a su vez que se hizo una proyección de las ventas de los mismos por segmento, lo que facilita la preparación de diferentes estrategias dinámicas y favorables en los distintos departamentos de Peugeot, ya sea marketing, ventas, importaciones, con la finalidad de cumplir con la demanda de los vehículos en el país.

Después de analizar el comportamiento de las proyecciones hechas a partir del modelo de ARIMA ajustado manualmente se proponen las siguientes estrategias para el cumplimiento de los objetivos de la marca, con enfoques diferentes para cada segmento de mercado con sus respectivos consumidores:

El segmento de clase baja tiene un enfoque en la accesibilidad y el valor, lo que se busca son precios accesibles, financiamientos flexibles que se ajusten a sus presupuestos y facilidades de pago. En marketing una buena estrategia son las campañas donde se promueva financiamientos accesibles, ofrecer cuotas cómodas por un periodo de tiempo. En ventas, se puede hablar de Test drives en comunidades y ferias que se hagan en las diferentes ciudades del país, así como incentivos o descuentos para las compras en grupos, de esta forma se puede aumentar las ganancias no por los precios de los automóviles sino por los volúmenes de compra que se pueden llegar a obtener.

El segmento de clase media tiene un enfoque en la relación calidad-precio, debido a que el cliente busca un equilibrio entre el diseño, la estética, la tecnología y el precio del carro de su exigencia. En marketing, se debe destacar la tecnología, seguridad y diseño mediante plataformas como las redes sociales, también funcionaría las alianzas con bancos para poder ofrecer tasas de interés atractivas dependiendo del perfil del cliente. En ventas, se puede incorporar planes de financiamiento personalizados o promociones con bonos por el retorno de autos usados siempre y cuando sean de la misma marca Peugeot. En operaciones se puede agilizar la entrega de autos ya que un problema frecuente es el tiempo que el consumidor debe esperar para que se le entregue el carro.

El segmento de clase alta tiene un enfoque especial, ya que este tipo de consumidor tiene preferencias por la exclusividad y la personalización, se exige lujo,

tecnología, estética y una experiencia premium. En marketing, se puede destacar los eventos exclusivos y lanzamientos VIP, ofrecer experiencias de prueba exclusivas y programas de personalización de los vehículos ya sea en los detalles como los colores y accesorios internos. En ventas es importante la atención personalizada y asesoramiento especializado que brinde una experiencia de compra única, entregas a domicilio y acceso a los últimos modelos y los más exclusivos. Asimismo, en operaciones se debe incluir un stock limitado para los modelos exclusivos ya que lo que se busca es que la experiencia de compra que tiene el cliente promueva la percepción de exclusividad.

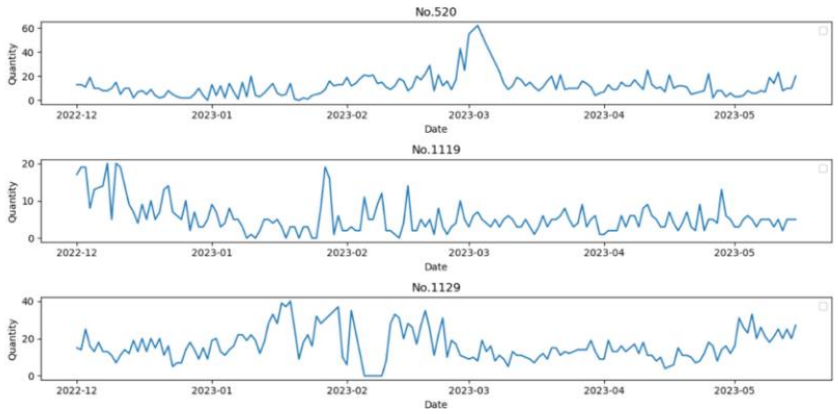
Discusión

Adecuándose al contexto de este trabajo de titulación, es indispensable garantizar la veracidad y exactitud de los resultados y metodologías usadas, es por esto que se comparará los resultados presentados con los resultados del artículo titulado "Research on E-commerce retail demand forecasting based on SARIMA model and K-means clustering algorithm", el cual trata de la integración del modelo SARIMA con el algoritmo de agrupamiento K-means en la industria del e-commerce para alinear los niveles de inventario con la expectativa de ventas, esto con la finalidad de reducir costos.

En el artículo científico se recolectaron datos de productos con sus respectivos comerciantes y almacén, se tienen en total 1996 series de tiempo y los productos que se comercializan varían entre alimentos, bebidas, muebles, materiales de construcción, instrumentos musicales y juguetes. Se aplicó data cleaning para tener una base de datos más manejable y libre de errores significativos, luego con un total de 35 comerciantes, 1212 tipos de productos y 54 almacenes se empezó con el análisis. Se tienen datos desde diciembre del 2022 hasta mayo del 2023 y se quieren predecir los siguientes 15 días. Para empezar, hicieron un análisis donde seleccionaron un grupo específico (520, 1119, 1129) para ver el comportamiento de los datos gráficamente.

Figura 37

Comportamiento de los datos a través del tiempo.

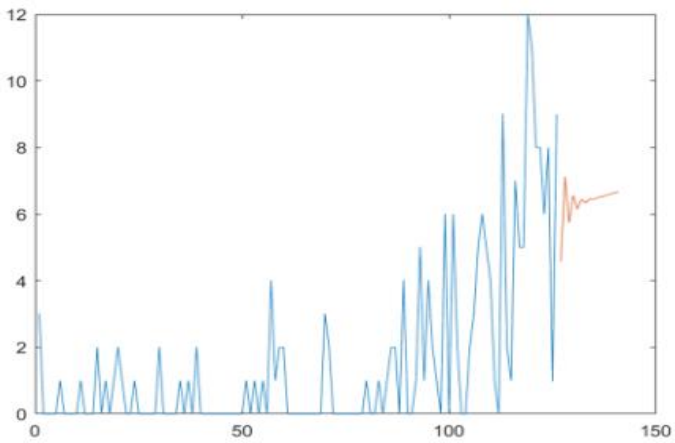


Nota: Comportamiento del grupo (520, 1119, 1129) a través del tiempo. Fuente: Academic Journal of Science and Technology.

Debido a la complejidad de los datos se decide evaluar 3 diferentes modelos para predecir y así poder escoger el óptimo, primero se aplicó el método auto-arima el cual evalúa y escoge los parámetros óptimos para ARIMA Y SARIMA, luego, escogiendo series de tiempo específicas se evaluaron los modelos empezando por ARIMA, para el gráfico de este modelo se tomó como ejemplo el grupo (19, 448, 30) los resultados de esta representación se muestran a continuación:

Figura 38

Evaluación del comportamiento de los datos con ARIMA.



Nota: Evaluación de la serie de tiempo (19,448,30) con el modelo ARIMA. Fuente: Academic Journal of Science and Technology.

Esta gráfica muestra que el modelo está teniendo problemas para ajustarse a los picos altos y que este no es tan efectivo para estos datos.

Continuando con la regresión lineal se aplicó una fórmula para poder predecir la demanda de un mes:

$$\hat{y}_i = a + bx_i \tag{21}$$

$$a = \frac{1}{n}\sum_{i=1}^{n} y_i - b\frac{1}{n}\sum_{i=1}^{n} x_i \tag{22}$$

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \tag{23}$$

Donde: a es el intercepto cuando $x_i = 0$, b es la pendiente que indica cuánto cambia la demanda por cada cambio en el tiempo, y_i es la cantidad demandada y x_i es la fecha. Después de calcular los parámetros avanzaron con el valor de la demanda en el periodo t+n con la siguiente fórmula:

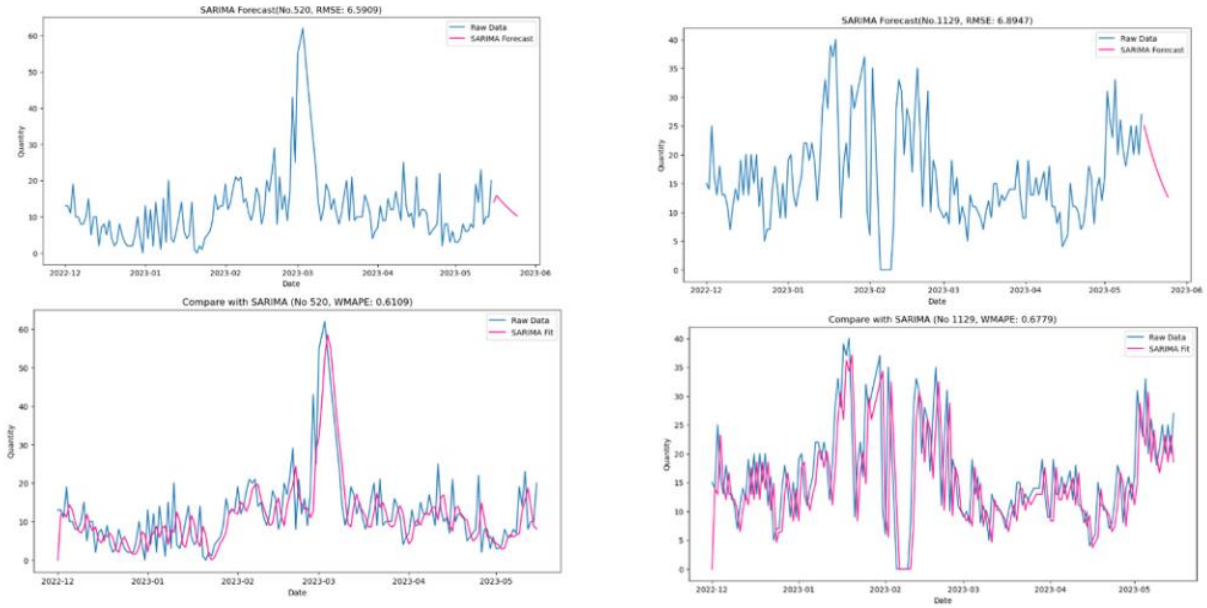
$$\hat{y}_{t+n} = a + bx_{t+n} \tag{24}$$

Ya con estos resultados de la predicción llegaron a la conclusión que este modelo era muy sensible a la volatilidad de los datos, por lo que se decide intentar con otro modelo.

Finalizando con SARIMA, el cuál es una versión mejorada del modelo ARIMA, para este modelo se escogieron 2 series de tiempo de las 1996 para analizar el comportamiento del mismo. El resultado para la serie de tiempo 520 y 1129 fue la siguiente:

Figura 39

Evaluación del comportamiento de los datos con el modelo SARIMA.



Nota: Evaluación de las series de tiempo (520, 1129) con el modelo SARIMA. Fuente: Academic Journal of Science and Technology.

Con este resultado se pudo demostrar que el modelo SARIMA fue el que mejor se ajustó a los datos, más que el modelo ARIMA, como se puede ver en el gráfico este modelo logró que coincidan mejor las predicciones del modelo con los datos reales pasados.

Luego, gracias al error porcentual absoluto medio ponderado (WMAPE) y el error cuadrático medio (RMSE) se llegó a la conclusión que el modelo con menos error era

el SARIMA, con este modelo se pudo observar gracias a los picos y bajadas que existen temporadas altas y bajas en ciertos productos.

Tabla 2

Evaluación de los errores de cada modelo.

Table 1. Evaluation index value of each model

Model	1-wmape	RMSE
LR	1-0.8119	10.8236
ARIMA	1-0.5944	4.4595
SARIMA	1-0.6278	6.7428

Nota: Evaluación de los errores donde se muestra que el que mejor se ajusta es el modelo SARIMA. Fuente: Academic Journal of Science and Technology.

Por otro lado, en nuestra investigación se recolectaron datos de ventas de autos Peugeot desde el 2018 hasta el 2023, asimismo se aplicó data cleaning y se adecua la base de datos para que queden las ventas de manera mensual, se lograron predecir los 2 años siguientes de ventas evaluando los datos bajo 3 modelos (ARIMA, ARIMA ajustado manualmente y ETS), al igual que en el artículo existía complejidad en los datos y después de analizar los errores en cada uno de los modelos se definió que el más óptimo para predecir en este caso era el modelo ARIMA ajustado manualmente, conclusión a la que se llegó tomando en cuenta indicadores de error como el RMSE y el MAPE, gracias a esto pudimos observar con ayuda de los gráficos en que rango de precios se ubica cada segmento y ver si los precios cambiarán drásticamente o se mantendrán en una rango constante en los próximos 2 años.

En el artículo seleccionado para su análisis se usa el modelo de K-means con la finalidad de mejorar la precisión de la predicción de la demanda que se haga con el SARIMA. Después de aplicar el Data Cleaning se desarrolló la segmentación de los datos de las series temporales que se tenían previamente, se obtuvo la clasificación de los productos, comerciantes y almacenes de acuerdo a sus características, esto facilitó la simplificación de los datos que anteriormente se había mencionado eran un complejos de tratar debido a su naturaleza, además de que gracias a la segmentación

hecha por K-means se aumentó el rendimiento de los resultados, esto quiere decir que los resultados se obtuvieron con un error menor al momento de hacer los clústeres. Este agrupamiento ayudó a analizar las tendencias y predicciones para cada uno de los grupos en lugar de hacer el análisis de manera individual, lo que aumentaría la complejidad del tratamiento de los datos para el desarrollo de los modelos.

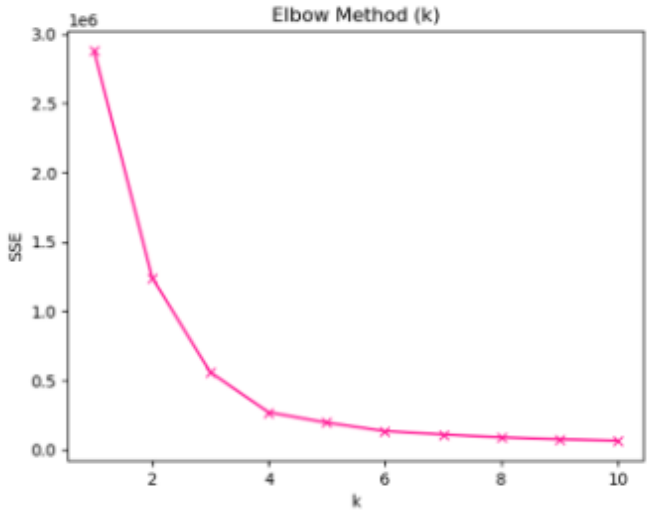
Este artículo usó el método de codo para determinar el número óptimo de clusters y el indicador usado para este método fue el de la suma de errores cuadrados el cual usa la siguiente fórmula:

$$SSE=\sum_{i=1}^{k} \sum_{p\epsilon C_i} |p - m_i|^2 \quad (25)$$

Donde C_i es el cluster i, p es el punto de muestra del cluster i, y m_i es el centroide del respectivo cluster y SSE se refiere al error de las muestras en un cluster lo cual representa la calidad del agrupamiento, con esto se generó un gráfico que contiene el rango de errores y valores de k, en el número del eje de las x donde se muestre un quiebre o codo es el número óptimo de clusters ya que a partir de ahí el error no disminuirá significativamente así se aumente el número de k, tal y como se muestra en el siguiente gráfico:

Figura 40

Método de codo con K-means.



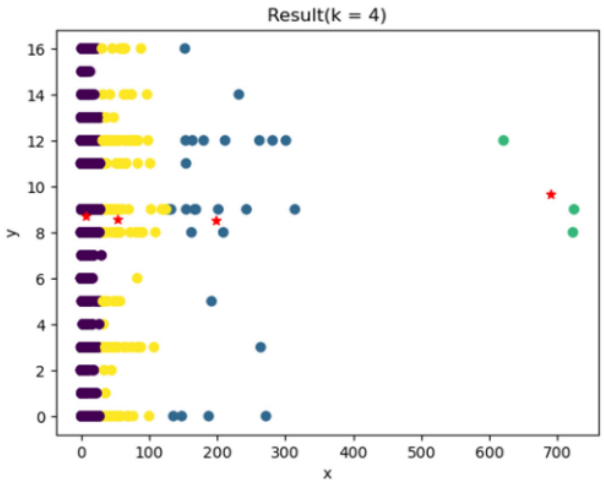
Nota: Se muestra mediante el gráfico que el número óptimo de clusters es 4. Fuente: Academic Journal of Science and Technology.

Mediante el método del codo se obtuvo que el número de clústeres es 4, esto quiere decir que son 4 segmentos que se van a utilizar en la segmentación. Los datos de los grupos incluyen productos, comerciantes y almacenes.

Luego se muestra la gráfica donde ya están los datos divididos en 4 segmentos:

Figura 41

Resultados de la segmentación.



Nota: Comportamiento de los 4 segmentos diferentes. Fuente: Academic Journal of Science and Technology.

Asimismo esta gráfica se explica de manera más explícita en una tabla que muestra como quedaron divididos los datos para un mejor manejo y entendimiento.

Tabla 3

Tabla de resultados de la clasificación de cada segmento.

Table 2. Table of classification centers

Category	Demand Quantity	Merchant Classification	Inventory Classification	Merchant Scale	Warehouse Category	Warehouse Region
1	6.4919	8.690	8.6906	0.3552	0.3414	2.6796
2	199.004	8.5	8.5	0.3846	0.8077	2.1154
3	689.331	9.666	9.6667	1	1	2
4	53.9763	8.5478	1.1783	0.4268	0.6943	2.2675

Fuente: Academic Journal of Science and Technology.

La categoría 1 muestra que existe una demanda baja, este segmento es el de electrodomésticos y materiales de construcción, además la escala de los comerciantes es de media a alta y en cuanto a los almacenes estos se ubican mayormente en regiones pequeñas o almacenes locales en su mayoría en el sur de China.

El segmento 2 muestra una demanda media e incluye productos relacionados con materiales de construcción. Tiene una escala comercial de media a grande con almacenes principalmente en el centro de China.

La categoría 3 muestra una demanda alta con productos de cuidado personal y de belleza, lo que quiere decir que este nicho de mercado es el que mayor significancia tiene. Tiene una escala comercial mediana y los almacenes se encuentran ubicados en el este de China.

Finalmente, la categoría 4 muestra similitud de comportamiento con el segmento 1, debido a que tiene una demanda baja pero tiene una escala comercial superior al segmento 1 y sus almacenes se encuentran distribuidos en diferentes partes de China.

Haciendo un contraste entre el artículo seleccionado y el presente trabajo de titulación ambos tienen similitudes que tienen gran significancia en el desarrollo de los modelos. Por un lado, los resultados de K-means obtenidos en esta investigación dio como resultados 3 segmentos de mercados claramente diferenciados en el contexto de ventas de los carros Peugeot, a pesar de que en ambos casos los datos que se utilizaron para las investigaciones tuvieron algún tipo de complejidad se pudo obtener los segmentos de mercados que den resultados acertados para la identificación y clasificación de las demandas de los productos que se incluyen en el modelo con la finalidad de mejorar las predicciones sobre las expectativas de ventas.

En ambas investigaciones se integraron dos modelos de Machine Learning que se complementaron de manera que se obtengan resultados óptimos para el pronóstico de ventas en contextos de ventas diferentes, demostrando así que para un correcto tratamiento de datos y evaluación de los mismos por medio de modelos de Business Intelligence se pueden obtener resultados veraces que posteriormente servirán de base para la toma de decisiones acertadas en los múltiples contextos empresariales que se pueden presentar.

Conclusión

La presente investigación muestra la importancia del uso del Machine Learning en el mundo automotriz. El uso de las múltiples herramientas que ofrece el Business Intelligence ayuda a obtener resultados más confiables, eficientes e informados, debido a que el desarrollo de los modelos de manera correcta permite minimizar los errores para de esta manera se tomen mejores decisiones con base a datos que muestren la realidad de los problemas empresariales que se estudian, además de que estas herramientas son muy versátiles al ser compatibles con varios ámbitos de diferentes industrias, en este caso en el contexto empresarial y económico.

Por esta razón, de acuerdo a este trabajo de investigación, comenzando con una base teórica, se logró explicar de forma sencilla y sintetizada las definiciones de los distintos modelos usados, como lo son kmeans y forecast y los distintos elementos dentro de estos, además de explicar la importancia de la aplicación del Machine Learning. En este caso el objetivo principal fue segmentar mercados mediante el algoritmo de K-means para su posterior pronóstico de precios de ventas de los dos años siguientes, lo que se buscó era entender el comportamiento de los segmentos de mercados correspondientes a diferentes tipos de consumidores, midiendolos por los precios de ventas mensuales, además de saber clasificar a los clientes por las características que prefieren en un carro de la marca Peugeot. Lo que el algoritmo K-means ayuda es a clasificar datos por características que permiten que los grupos estén bien definidos para así poder desarrollar estrategias que estén bien orientadas a los grupos segmentados, debido a que no todos los consumidores tienen el mismo poder adquisitivo ni buscan las mismas características en carros, por lo cual se vuelve vital agrupar los clientes de forma de forma adecuada ya que un buen enfoque en las ventas por segmento ayuda a incrementar los ingresos de la empresa.

Los resultados que se obtuvieron de la clasificación fueron 3 segmentos de mercado con sus diferentes características. El segmento 1 corresponde a los consumidores de clase media debido a que sus ingresos son los que tienen cantidades que se encuentran en la mitad de los niveles de ventas, este grupo de clientes es que el que busca características más exigidas pero que no cuentan con el presupuesto o la disposición

de pagar más dinero por un carro de alta gama pero que si busca tener especificaciones no tan básicas como las de un carro de gama básica.

El segmento 2 representa el segmento de clase baja, este tiene el menor nivel de ventas por lo que las características de los automóviles son más básicas y no tiene avance tecnológico superior. Estos clientes están en búsqueda de automóviles que se consideren accesibles pero que cuenten con las características propias de Peugeot.

El segmento 3 es el correspondiente a la gama alta, son clientes cuyas exigencias son más avanzadas tanto en especificaciones como en tecnología, estos clientes son los que muestran un poder adquisitivo superior y que tienen la capacidad y disposición de pagar por un carro que les brinde comodidades, seguridad y tecnología.

Referencias

Allen, W. H. (Mayo de 2006). AUTOMÓVILES PEUGEOT CUMPLE 110 AÑOS. (I. S. Echeverría, Ed.) *Revista de Ingeniería Mecánica*, 9(2), 66-67. Obtenido de https://www.redalyc.org/articulo.oa?id=225117943011

Álvarez-Indacochea, E., Figueroa-Soledispa, L., & Peñafiel-Loor, A. (2020). Estrategias de marketing y su incidencia en la satisfacción del cliente en tiempos de COVID-19. *FIPCAEC*, 5(1), 48-61. Recuperado de https://www.fipcaec.com/index.php/fipcaec/article/view/281/493

Asociación de Empresas Automotrices del Ecuador (AEADE). (2024). Boletín de ventas de marzo 2024. Recuperado de: https://www.aeade.net/wp-content/uploads/2024/03/BOLETIN-VENTAS_PRENSA_MARZO-2024.pdf

Asociación de Empresas Automotrices del Ecuador (AEADE). (2024). *Boletín de ventas de marzo 2024* [Gráfico]. Recuperado de https://www.aeade.net/wp-content/uploads/2024/03/BOLETIN-VENTAS_PRENSA_MARZO-2024.pdf

Areneda, P. (27 de Abril de 2021). Tidyverse para Data Análisis. *RPUBS*. Obtenido de https://rpubs.com/paraneda/tidyverse

Armijos, S. (14 de Enero de 2021). Se proyecta recuperación de ventas para vehículos nuevos y usados. *Vistazo*. Obtenido de https://www.vistazo.com/enfoque/se-proyecta-recuperacion-de-ventas-para-vehiculos-nuevos-y-usados-IDVI218716

Asociación Latinoamericana de Distribuidores de Automotores (ALADDA), A. L. (19 de Julio de 2024). Mercado Automotor Regional: Mercado regional al cierre de junio de 2024. *AEADE*, 1-6.

Berrendero, J. (s.f.). Introducción a CARET. *RPUBS*. Obtenido de https://rpubs.com/joser/caret

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (3 de Noviembre de 2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of statistical Software, 61*(6), 1-36. doi:https://doi.org/10.18637/jss.v061.i06

Codificandobits. (s.f.). Los sets de entrenamiento, validación y prueba. *Codificandobits*. Obtenido de https://codificandobits.com/blog/sets-entrenamiento-validacion-y-prueba/

CRAN.R PROJECT. (s.f.). Extract and Visualize the Results of Multivariate Data Analyses. *CRAN.R PROJECT*. Obtenido de https://cran.r-project.org/web/packages/factoextra/readme/README.html

Datacamp. (25 de Abril de 2024). ¿Qué es el machine learning? *Datacamp*. Obtenido de https://www.datacamp.com/es/blog/what-is-machine-learning

DataCamp. (29 de Julio de 2024). ¿Qué es la Distancia Manhattan? *DataCamp*. Obtenido de https://www.datacamp.com/es/tutorial/manhattan-distance

DataScientest. (13 de Diciembre de 2022). Machine learning: definición, funcionamientos y usos. *DataScientest*. Obtenido de https://datascientest.com/es/machine-learning-definicion-funcionamiento-usos

Exponentis. (15 de julio del 2019). *Ejemplo de clustering con k-means en Python* [Gráfico]. Obtenido de https://exponentis.es/ejemplo-de-clustering-con-k-means-en-python

Fernandez, C. (06 de Marzo de 2024). La OCDE explica el contenido de su definición actualizada de los sistemas de la inteligencia artificial. *Diario LaLey*. Obtenido de: https://diariolaley.laleynext.es/dll/2024/03/07/la-ocde-explica-el-contenido-de-su-definicion-actualizada-de-los-sistemas-de-inteligencia-artificial

Ferrell, O. C., & Hartline, M. D. (2012). Estrategia de marketing (5ª ed., M.E. Treviño Rosales & M. P. Carril Villarreal, Trads.). *Cengage Learning* Editores. ISBN: 978-607-481-824-6.

Fiallos, A. (Noviembre de 2016). *Evaluación de algoritmos de datamining para clustering* [Gráfico]. Obtenido de https://www.researchgate.net/figure/Clusters-obtenidos-con-algoritmo-K-Means_fig3_310951203

Gavara, J. (24 de Mayo de 2022). Explorando la Ciencia de Datos. *Themegraphy*. Obtenido de https://joseignaciogavara.com/graficos-en-r-lattice/

GeeksforGeeks. (05 de Agosto de 2024). Euclidean distance. *GeeksforGeeks*. Obtenido de https://www.geeksforgeeks.org/euclidean-distance/

González, J. (15 de Enero de 2021). Los pedidos de carros Peugeot aumentaron 15% y mejora su posición en América Latina. *La República*. Obtenido de https://www.larepublica.co/empresas/los-pedidos-de-carros-de-peugeot-aumentaron-15-y-mejoran-su-posicion-en-america-latina-3111254

Gonzalez, L. (02 de Noviembre de 2020). Aprendizaje no supervisado. *AprendeIA*. Obtenido de https://aprendeia.com/aprendizaje-no-supervisado-machine-learning/

González, P. (08 de Mayo de 2024). *Las ventas de carros en Ecuador se desplomaron en abril por el alza del IVA* [Gráfica]. Obtenido de https://www.primicias.ec/noticias/economia/ventas-carros-iva-chevrolet/

Hernández, M., & Muñoz, M. (2004). Diseño de una metodología para la planeación y programación de la producción de café tostado y molido en la planta de Colcafé Bogotá. Pontificia Universidad Javeriana, 11-130.

Hierro, A. (Mayo de 2006). Automóviles Peugeot cumple 110 años. *Revista de Ingenieria Mecánica*, 9(2), 66-67. Obtenido de https://www.redalyc.org/articulo.oa?id=225117943011

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). *OTexts*. Recuperado de: https://otexts.com/fpp2/

Hyndmanm, R. (19 de Junio de 2024). Forecast. *RDocumantation*. Obtenido de https://www.rdocumentation.org/packages/forecast/versions/8.23.0

IBM. (s.f.). ¿Qué es el aprendizaje supervisado? *IBM*. Obtenido de https://www.ibm.com/mx-es/topics/supervised-learning

INESDI. (02 de Noviembre de 2022). Qué es el aprendizaje no supervisado. *INESDI TECHSCHOOL*. Obtenido de https://www.inesdi.com/blog/que-es-aprendizaje-no-supervisado/

La República. (24 de Junio de 2018). Peugeot Ecuador presentó su nueva gama de vehículos a diesel. *La República*. Obtenido de https://www.larepublica.ec/blog/2018/06/24/peugeot-ecuador-presento-su-nueva-gama-de-vehiculos-a-dieselpeugeot-ecuador-presento-su-nueva-gama-de-vehiculos-a-dieseleva-gama-de-vehiculos-a-diesel-en-el-autoshow-de-guayaquil/

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting: Methods and applications (3rd ed.). Wiley.

Martí, A. (3 de Abril de 2020). El curioso origen de Peugeot. *Xataka*. Obtenido de https://www.xataka.com/historia-tecnologica/curioso-origen-peugeot-salto-tecnologico-que-se-consagro-garajes-cuando-empezo-cocinas#:~:text=Peugeot%20empez%C3%B3%20siendo%20un%20peque%C3%B1o,Industrial%20evolucion%C3%B3%20por%20otro%20camino

Mauricio, J. A. (2007). Análisis de series temporales . *Universidad complutense de Madrid* , 1-295.

Morgan, S. (16 de Enero de 2021). La fusión de FCA y Groupe PSA ha sido completada. Obtenido de Stellantis: https://media.stellantisnorthamerica.com/newsrelease.do?id=22462&esp=si&mid=

Naula, P. (5 de febrero de 2024). Venta de vehículos nuevos con leve caída en 2023. *Expreso*. Obtenido de: https://elmercurio.com.ec/2024/02/05/vehiculos-ventas-disminucion-ecuador-202/

Núria, E. (s.f.). Machine learning. *Bismart*. Obtenido de https://blog.bismart.com/machine-learning-que-es-como-funciona-para-que-sirve

Peña, S. (2017). Análisis de datos. Fondo editorial Areandino, Fundación Universitaria del Área Andina.

Peugeot. (11 de Abril de 2023). Historia del símbolo del león de Peugeot. *Peugeot Uruguay*, 1-5. Obtenido de

https://www.peugeot.com.uy/marca/nosotros/peugeot-magazine/historia-del-simbolo-del-leon-de-peugeot.html

Peugeot (2023). SUV PEUGEOT COMPACTOS, GRANDES Y FAMILIARES. *Peugeot Ecuador*. Recuperado de https://www.peugeot.com.ec/gama/categoria/suv.html#:~:text=Los%20SUV%2C%20o%20Sport%20Utility,m%C3%A1s%20utilizados%20por%20las%20familias.

Ponce, J., Torres, A., Quezada, F., & Pedreño, O. (Marzo de 2014). Inteligencia artificial. *Iniciativa latinoamericana de libros de texto abiertos*, 10-225. doi:http://dx.doi.org/10.13140/2.1.3720.0960

Redacción Empresas. (26 de Febrero de 2024). Tres de cada 10 ecuatorianos usaron IA en sus tareas diarias de trabajo. *Primicias*. Obtenido de https://www.primicias.ec/nota/primicias-empresas/eco-naranja/el-36-de-ecuatorianos-uso-la-ia-en-sus-tareas-diarias-dentro-del-trabajo/#:~:text=A%20pesar%20de%20que%20solo,los%20pa%C3%ADses%20de%20la%20regi%C3%B3n.

Ríos, G., & Hurtado, C. (2008). Series de tiempo. *Universidad de Chile*, 1-52.

Rouhiainen, L. (noviembre de 2018). Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro. *Editorial Planeta*, 1-22.

Salvador Maceira, M. (2019). Machine Learning aplicado al trading.

SEON. (2024). Modelos supervisados machine learning. *SEON*. Obtenido de https://seon.io/es/recursos/glosario/modelos-supervisados-machine-learning/

Statista. (12 de Noviembre de 2024). Ranking de los modelos de automóviles con mayores ventas en América Latina en 2024. Obtenido de Statista Research Department: https://es.statista.com/estadisticas/1087895/cuota-mercado-america-latina-marcas-automoviles/#:~:text=El%20Volkswagen%20Polo%20fue%20el,Chevrolet%20Onix%20completaron%20el%20podio.

Statista. (12 de Noviembre de 2024). Ranking de los modelos de automóviles con mayores ventas en América Latina en 2024 [Gráfico]. *Statista*. Recuperado de https://es.statista.com/estadisticas/1087895/cuota-mercado-america-latina-marcas-automoviles/#:~:text=El%20Volkswagen%20Polo%20fue%20el,Chevrolet%20Onix%20completaron%20el%20podio.

Stellantis Media - *La fusión de FCA y Groupe PSA ha sido completada.* (2021, 16 enero). https://media.stellantisnorthamerica.com. https://media.stellantisnorthamerica.com/newsrelease.do?id=22462&esp=si&mid=

Tapia, E. (09 de Diciembre de 2023). *Carros europeos entrarán a Ecuador con arancel cero desde 2024* [Gráfico]. Obtenido de https://www.primicias.ec/noticias/economia/carros-europeos-arancel-cero-acuerdo-comercial-ue/

Tapia, E. (19 de Febrero de 2024). *Por el alza del IVA, el ICE también aumentará y los carros serán más caros en 2024* [Gráfico]. Obtenido de https://www.primicias.ec/noticias/economia/iva-incremento-precios-carros-ice-impuestos/

Tejada, D. (28 de Mayo de 2023). K-means clustering. *RPubs*. Obtenido de https://rpubs.com/Dariel1102/1046632

Tirado, L. (17 de Julio de 2024). Ley de la inteligencia artificial en Ecuador: Un nuevo marco regulatorio para 2024. *GlobalSuite Solutions*. Obtenido de https://www.globalsuitesolutions.com/es/ley-inteligencia-artificial-ecuador/

Villavicencio, J. (11 de octubre de 2022). Introducción a series de tiempo. *Slideshare*, 1-33. Recuperado de https://es.slideshare.net/slideshow/manualintroseriestiempopdf/253508535

Vivero, M. (14 de Octubre de 2024). Regulación de la inteligencia artificial en Ecuador. *Lexis Blog*. Obtenido de https://www.lexis.com.ec/blog/legaltech/no-todo-aporta-regulacion-de-la-inteligencia-artificial-en-ecuador

Yamin, M., Mahandari, C. P., & Mumtaz, M. M. (2023). *Simulation and analysis of hatchback car driving comfort and handling performance.* Journal of Engineering Science and Technology (JEST), 2(3), 89-94. https://doi.org/10.56741/jnest.v2i03.399

Yu, L., Zhang, Y., Jian, G., Gutman, I., 2017. Classification for Microarray Data Based on K-Means Clustering Combined with Modified Single-to-Noise-Ratio Based on Graph Energy. J. Comput. Theor. Nanosci. 14, 598–606. https://doi.org/10.1166/jctn.2017.6248

Zhao, Y. (27 de Abril de 2024). Research on E-Commerce Retail Demand Forecasting Based on SARIMA Model and K-means Clustering Algorithm. *Academic journal of science and technology*, 10(3), 226-231. doi:https://doi.org/10.54097/45acxz19

Thompson, I. (2005, agosto). La segmentación del mercado. *MarketingPower*. Recuperado de: http://www.marketingpower.com/mg-dictionary.php

Autocasión. (2016, agosto 5). ¿Qué es un coche urbano? *Autocasión.* Recuperado de: https://www.autocasion.com/actualidad/reportajes/que-es-un-coche-urbano

América Economía. (28 de Enero de 2019). Peugeot creció en 2018 más de 200% en ventas en Ecuador. *America Economía*. Obtenido de https://www.americaeconomia.com/negocios-industrias/peugeot-crecio-en-2018-mas-de-200-en-ventas-en-ecuador

Alarcón, M. (2023, agosto 18). ¿Cuáles son los tipos de vehículos comerciales? *Freeway Insurance*. Recuperado de https://freewayseguros.com/blog/vehiculos-comerciales/cuales-son-los-tipos-de-vehiculos-comerciales/

Aguilera, N. (2024, agosto 30). Ventas de vehículos chinos aumentan en Ecuador. *Voz de América*. Recuperado de https://www.vozdeamerica.com/a/ventas-de-vehiculos-chinos-aumentan-en-ecuador/7765634.html

Cámara de la Industria Automotriz del Ecuador. (2024, octubre). Cifra de ventas en el sector automotor: Septiembre 2024. CINADE.

Bacorelle, J. (2024, 8 febrero). Las ventas de coches Peugeot en el mundo crecen un 6 % hasta los 1,2 millones. *ABC Motor*. Recuperado de:

https://www.abc.es/motor/economia/ventas-coches-peugeot-mundo-crecen-millones-20240208110328-nt.html

ANEXOS

Anexo 1

Base de datos

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
FRANCIA	AUTOMOVIL	STATION WAGON	13847.22	1600	DIESEL	2018-10-22
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	26990.00	1560	DIESEL	2018-11-23
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-11-12
FRANCIA	JEEP	JEEP	48990.00	1598	GASOLINA	2018-11-13
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-11-12
FRANCIA	JEEP	JEEP	28874.90	1200	GASOLINA	2018-10-18
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-10-18
FRANCIA	JEEP	JEEP	31990.00	1587	GASOLINA	2018-10-17
FRANCIA	JEEP	JEEP	31990.00	1587	GASOLINA	2018-10-18
FRANCIA	OMNIBUS	MINIBUS	19196.00	2198	DIESEL	2018-09-06
FRANCIA	JEEP	JEEP	31990.00	1587	GASOLINA	2018-10-17
FRANCIA	JEEP	JEEP	31990.00	1587	GASOLINA	2018-10-18
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-10-17
FRANCIA	AUTOMOVIL	CONVERTIBLE	7384.30	1600	GASOLINA	2018-09-11
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-10-16
FRANCIA	CAMIONETA	FURGONETA	26060.20	1600	DIESEL	2018-09-25
FRANCIA	JEEP	JEEP	48990.00	1598	GASOLINA	2018-10-05
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-10-05
FRANCIA	OMNIBUS	MINIBUS	19196.00	2198	DIESEL	2018-09-06
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-09-28
ESPA&#65533;A	CAMIONETA	REPARTO	24990.00	1560	DIESEL	2018-09-06

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESPA&#65533;A	CAMIONETA	REPARTO	24990.00	1560	DIESEL	2018-09-06
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	26990.00	1560	DIESEL	2018-09-17
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-09-19
FRANCIA	JEEP	JEEP	31990.00	1587	GASOLINA	2018-09-04
FRANCIA	JEEP	JEEP	48990.00	1598	GASOLINA	2018-09-12
ESPA&#65533;A	CAMIONETA	REPARTO	24990.00	1560	DIESEL	2018-09-07
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-08-29
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-08-29
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-08-31
FRANCIA	JEEP	JEEP	48990.00	1598	GASOLINA	2018-10-05
FRANCIA	JEEP	JEEP	48990.00	1598	GASOLINA	2018-10-05
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2018-08-24
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-08-23
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-08-23
FRANCIA	JEEP	JEEP	48990.00	1598	GASOLINA	2018-08-22
ESPA&#65533;A	CAMIONETA	FURGONETA	27990.00	1560	DIESEL	2018-08-15
ESPA&#65533;A	CAMIONETA	FURGONETA	27990.00	1560	DIESEL	2018-08-15
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-08-13
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-08-09
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-08-09
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-08-09

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-07-24
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2018-07-31
FRANCIA	JEEP	JEEP	31990.00	1587	GASOLINA	2018-07-25
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2018-03-28
FRANCIA	JEEP	JEEP	31990.00	1587	GASOLINA	2018-03-28
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-11-12
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2018-07-25
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-07-24
FRANCIA	JEEP	JEEP	31990.00	1587	GASOLINA	2018-07-09
ESPA&#65533;A	CAMIONETA	FURGONETA	27990.00	1560	DIESEL	2018-07-20
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-07-20
FRANCIA	JEEP	JEEP	48990.00	1598	GASOLINA	2018-07-20
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-07-18
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2018-07-17
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-07-17
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-07-17
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-07-17
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-07-17
ESPA&#65533;A	CAMIONETA	FURGONETA	27990.00	1560	DIESEL	2018-07-13
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-07-09
ESPA&#65533;A	CAMIONETA	FURGONETA	27990.00	1560	DIESEL	2018-07-10

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-06-26
ESPA&#65533;A	CAMIONETA	REPARTO	24990.00	1560	DIESEL	2018-06-27
ESPA&#65533;A	CAMIONETA	REPARTO	24990.00	1560	DIESEL	2018-06-27
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-06-21
ESPA&#65533;A	CAMIONETA	FURGONETA	27990.00	1560	DIESEL	2018-06-25
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-06-22
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2018-07-17
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-06-21
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-06-20
ESPA&#65533;A	CAMIONETA	FURGONETA	25990.00	1560	DIESEL	2018-06-20
FRANCIA	JEEP	JEEP	48990.00	1598	GASOLINA	2018-06-20
ESPA&#65533;A	CAMIONETA	FURGONETA	27990.00	1560	DIESEL	2018-06-19
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-06-18
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2018-06-18
ESPA&#65533;A	CAMIONETA	FURGONETA	27990.00	1560	DIESEL	2018-06-15
ESPA&#65533;A	CAMIONETA	REPARTO	24990.00	1560	DIESEL	2018-06-14
FRANCIA	JEEP	JEEP	48990.00	1598	GASOLINA	2018-06-14
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-06-13
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-06-07
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	24990.00	1598	GASOLINA	2018-06-07
ARGENTINA	AUTOMOVIL	SEDAN	39325.00	1598	GASOLINA	2018-01-30

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-12-05
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-12-05
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-11-30
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2019-12-06
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1560	DIESEL	2019-12-06
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-11-30
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-10-28
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-12-05
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-12-02
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-12-13
ESPA&#65533;A	AUTOMOVIL	SEDAN	22990.00	1587	GASOLINA	2019-12-06
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-12-06
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-12-06
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-12-06
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-12-06
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-12-06
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-12-04
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-11-20
FRANCIA	AUTOMOVIL	STATION WAGON	19990.00	1560	DIESEL	2019-08-28
FRANCIA	AUTOMOVIL	STATION WAGON	25990.00	1560	DIESEL	2019-08-29
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-11-30
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-12-13

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-11-28
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-11-28
ESPA&#65533;A	AUTOMOVIL	SEDAN	22990.00	1587	GASOLINA	2019-11-21
FRANCIA	JEEP	JEEP	49990.00	1997	DIESEL	2019-11-18
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-11-20
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-11-28
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-11-18
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-11-27
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-11-26
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-11-27
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-11-26
FRANCIA	JEEP	JEEP	26490.00	1199	GASOLINA	2019-11-26
FRANCIA	JEEP	JEEP	26490.00	1199	GASOLINA	2019-11-28
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-11-28
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-11-27
ESPA&#65533;A	AUTOMOVIL	SEDAN	22990.00	1587	GASOLINA	2019-11-27
FRANCIA	AUTOMOVIL	STATION WAGON	19990.00	1560	DIESEL	2019-11-21
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-11-27
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-11-27
FRANCIA	CAMIONETA	FURGONETA	20492.72	1600	DIESEL	2019-11-18
FRANCIA	AUTOMOVIL	STATION WAGON	19990.00	1560	DIESEL	2019-11-18
FRANCIA	AUTOMOVIL	STATION WAGON	19990.00	1560	DIESEL	2019-12-17

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-10-31
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-10-31
FRANCIA	JEEP	JEEP	49990.00	1997	DIESEL	2019-10-30
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-10-29
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-10-16
FRANCIA	JEEP	JEEP	26490.00	1199	GASOLINA	2019-10-21
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-10-21
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-10-30
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-10-25
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-10-15
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-10-29
FRANCIA	AUTOMOVIL	STATION WAGON	19990.00	1560	DIESEL	2019-10-29
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-10-29
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-10-23
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-10-26
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-10-29
ESPA&#65533;A	CAMIONETA	REPARTO	24990.00	1560	DIESEL	2019-10-28
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-09-30
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-10-28
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-10-24
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-10-28

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESPA&#65533;A	CAMIONETA	REPARTO	22990.00	1560	DIESEL	2019-10-01
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-09-30
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-09-27
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-09-30
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-30
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-09-30
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-09-30
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-28
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-09-28
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-27
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-30
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-27
FRANCIA	AUTOMOVIL	STATION WAGON	25990.00	1560	DIESEL	2019-09-30
ESPA&#65533;A	AUTOMOVIL	SEDAN	22990.00	1587	GASOLINA	2019-10-01
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1560	DIESEL	2019-07-30
ESPA&#65533;A	CAMIONETA	REPARTO	24990.00	1560	DIESEL	2019-07-30
FRANCIA	AUTOMOVIL	STATION WAGON	25990.00	1560	DIESEL	2019-09-30
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-09-30
FRANCIA	JEEP	JEEP	49990.00	1997	DIESEL	2019-10-15
FRANCIA	JEEP	JEEP	26490.00	1199	GASOLINA	2019-09-30
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-09-18
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-09-20

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-13
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-08-30
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1560	DIESEL	2019-09-12
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-09-11
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1560	DIESEL	2019-09-10
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-11
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-08-22
FRANCIA	JEEP	JEEP	49990.00	1997	DIESEL	2019-09-11
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-09-10
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-09-06
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-10
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-09-10
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-10
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-08-31
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-09-06
FRANCIA	JEEP	JEEP	26490.00	1199	GASOLINA	2019-09-04
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-09-06
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-08-30
FRANCIA	CAMIONETA	FURGONETA	46990.00	1997	DIESEL	2019-08-31
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-09-03
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-06
FRANCIA	CAMIONETA	FURGONETA	46990.00	1997	DIESEL	2019-09-06

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-07-26
FRANCIA	AUTOMOVIL	STATION WAGON	25990.00	1560	DIESEL	2019-07-26
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-07-18
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-07-26
FRANCIA	JEEP	JEEP	26490.00	1199	GASOLINA	2019-07-12
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-07-24
ESPA&#xFFFD;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2019-07-24
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-07-25
FRANCIA	JEEP	JEEP	27990.00	1199	GASOLINA	2019-07-16
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-07-05
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-07-25
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	22990.00	1587	GASOLINA	2019-07-16
FRANCIA	JEEP	JEEP	26490.00	1199	GASOLINA	2019-07-24
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-07-17
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1560	DIESEL	2019-07-10
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-07-18
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-07-24
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-07-11
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-07-22
FRANCIA	JEEP	JEEP	28990.00	1199	GASOLINA	2019-07-16
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-07-10

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
FRANCIA	JEEP	JEEP	29990.00	1587	GASOLINA	2019-02-22
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1560	DIESEL	2019-02-22
FRANCIA	JEEP	JEEP	39990.00	1997	DIESEL	2019-02-28
FRANCIA	JEEP	JEEP	29990.00	1587	GASOLINA	2019-02-20
FRANCIA	JEEP	JEEP	53990.00	1598	GASOLINA	2019-02-18
FRANCIA	JEEP	JEEP	53990.00	1598	GASOLINA	2019-02-21
FRANCIA	JEEP	JEEP	59990.00	1598	GASOLINA	2019-02-21
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-02-21
FRANCIA	JEEP	JEEP	59990.00	1598	GASOLINA	2019-02-18
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-02-18
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-02-18
FRANCIA	JEEP	JEEP	59990.00	1598	GASOLINA	2019-02-18
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-02-15
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-02-18
ESPA&#xFFFD;A	AUTOMOVIL	SEDAN	19990.00	1587	GASOLINA	2019-02-15
FRANCIA	JEEP	JEEP	59990.00	1598	GASOLINA	2019-07-29
FRANCIA	JEEP	JEEP	29990.00	1587	GASOLINA	2019-02-18
FRANCIA	JEEP	JEEP	27990.00	1199	GASOLINA	2019-02-12
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-02-18
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-02-13
FRANCIA	JEEP	JEEP	49990.00	1598	GASOLINA	2019-02-08

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-08-22
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-08-29
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-14
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-11
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-08-25
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-17
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-09
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-29
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-17
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-26
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-29
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-10-05
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-10-07
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-16
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-10-05
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-10-29
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-11-06
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-09-21
FRANCIA	AUTOMOVIL	STATION WAGON	21990.00	1560	DIESEL	2020-11-24
FRANCIA	AUTOMOVIL	STATION WAGON	19990.00	1560	DIESEL	2020-07-28
FRANCIA	AUTOMOVIL	STATION WAGON	19990.00	1560	DIESEL	2020-08-18

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2019-12-24
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-07
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-12
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-12
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-13
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-13
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-14
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-19
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-01-31
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-27
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-27
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-03-03
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-03-05
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-02-29
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-05-13
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-05-29
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-06-05
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-06-05
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-06-12
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-07-24
ESPA&#65533;A	CAMIONETA	FURGONETA	29990.00	1560	DIESEL	2020-07-08

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-04
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-01-21
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-01-20
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-13
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-01-28
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-07
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-17
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-14
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-14
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2019-09-27
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-01-30
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-18
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-11
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-04
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-18
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-21
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-13
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-29
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-21
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-03-05
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-02-28
FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2020-03-06

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-29
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-29
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-30
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-29
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-29
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-30
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-30
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-30
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-30
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-30
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-12-06
ESPA&#65533;A	JEEP	JEEP	32990	1199	GASOLINA	2021-11-29
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-17
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-07
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-17
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-14
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-18
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-18
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-18
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-14
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-18
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-17

3,767 to 3,788 of 7,604 entries, 7 total columns

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-14
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-11
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-28
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-10
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-15
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-08
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-21
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-18
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-16
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-10
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-31
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-22
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-05-31
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-25
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-23
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-28
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-29
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-28
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-25
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-24
ESPA&#65533;A	JEEP	JEEP	27490	1199	GASOLINA	2021-06-29

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-18
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-23
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-24
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-24
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-24
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-30
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-24
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-31
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-24
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-09-05
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-09-05
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-23
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-24
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-09-16
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-09-05
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-09-27
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-08-26
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-09-30
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-09-30
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-09-26
ESLOVAQUIA	AUTOMOVIL	STATION WAGON	21990	1199	GASOLINA	2022-10-21

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-04-12
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-04-20
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-04-18
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-06-22
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-07-07
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-07-19
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-08-05
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-08-05
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-08-11
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-08-26
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-08-26
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-09-02
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-08-29
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-11-09
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-10-25
ESPA&#65533;A	AUTOMOVIL	SEDAN	23990	1587	GASOLINA	2022-12-21
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990	1560	DIESEL	2022-01-28
ESPA&#65533;A	AUTOMOVIL	SEDAN	19990	1560	DIESEL	2022-06-30
FRANCIA	CAMIONETA	FURGONETA	46990	1997	DIESEL	2022-01-27
FRANCIA	CAMIONETA	FURGONETA	46990	1997	DIESEL	2022-01-26
FRANCIA	CAMIONETA	REPARTO	36990	1997	DIESEL	2021-11-30

	PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
9	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-09-25
0	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-09-28
1	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-09-22
2	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-09-29
3	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-09-29
4	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-05
5	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-10
6	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-23
7	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-11
8	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-23
9	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-23
0	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-24
1	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-25
2	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-21
3	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-22
4	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-24
5	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-27
6	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-30
7	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-26
8	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-21
9	FRANCIA	JEEP	JEEP	39990.00	1598	GASOLINA	2023-10-31

PAIS	CLASE	SUBCLASE	VENTA	CILINDRAJE	COMBUSTIBLE	FECHA
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-03-02
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-02-24
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-02-28
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-02-24
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-02-24
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-02-28
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-02-28
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-03-10
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-03-17
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-03-17
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-04-20
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-03-30
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-04-05
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-03-29
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-04-20
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-04-20
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-05-16
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-05-12
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-05-29
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-06-09
ESPA&#65533;A	JEEP	JEEP	33490	1199	GASOLINA	2023-06-09

Anexo 2

Script completo en R studio

```
Peugeot2 <- read.csv2("../Downloads/Compilado ventas peugeot2.csv")

Peugeot_data_cleaned <- Peugeot2[,c(4,5)]

#Llamamos a las librerias a usar
library(class)
library(caret)
library(tidyverse)
library(cluster)
library(factoextra)
library(NbClust)

#Plantamos semilla
set.seed(123)

#Dividimos los datos en entrenamiento y prueba
trainIndex <- createDataPartition(Peugeot_data_cleaned$VENTA,
                                  p=.7, list = FALSE)
train_data <- Peugeot_data_cleaned[trainIndex,]
test_data <- Peugeot_data_cleaned[-trainIndex]

#Escalamos los datos
SKtraindata <- scale(train_data)
SKtestdata <- scale(test_data)

################TRABAJANDO CON LOS DATOS DE ENTRENAMIENTO######

fviz_nbclust(SKtraindata, kmeans, method = "silhouette")
fviz_nbclust(SKtraindata, kmeans, method = "wss")
fviz_nbclust(SKtraindata, kmeans, method = "gap_stat")

#Calculamos los clústers

K3 <- kmeans(SKtraindata, centers = 3, nstart = 25)
```

```
#Plot
PTrainOC <- fviz_cluster(K3, data = SKtraindata,repel = FALSE)
PTrainOC
PTainMC <- fviz_cluster(K3, data = SKtraindata,
                        ellipse.type = "euclid",
                        repel = FALSE, star.plot= TRUE)
PTainMC
PTrainIC <- fviz_cluster(K3, data = SKtraindata,
                         ellipse.type = "norm",repel = FALSE)
PTrainIC

################TRABAJANDO CON LOS DATOS DE PRUEBA######

#Calculamos los clústers

Kt3 <- kmeans(SKtestdata, centers = 3, nstart = 25)

#agregamos las clasificaciones al conjunto de datos
Peugeot_data_cleaned$class <- NA
Peugeot_data_cleaned$class[trainIndex] <- as.factor(K3$cluster)
Peugeot_data_cleaned$class[-trainIndex] <- as.factor(Kt3$cluster)

#Pasamos las clasificaciones a la base original
Peugeot2$clasificacion <- Peugeot_data_cleaned$class

#Boxplot

Peugeot2$clasificacion=factor(Peugeot2$clasificacion)

G1=ggplot(Peugeot2,aes(x=Peugeot2$clasificacion,
                       y=Peugeot2$VENTA))+
  geom_boxplot(fill="yellow",colour="black")+
  theme(axis.title = element_blank())+
  theme(legend.position = "none")+
  labs("Precio con respecto al segmento")
```

```r
G2= ggplot(Peugeot2, aes(x = Peugeot2$clasificacion, y = Peugeot2$VENTA, fill = Peugeot2$clasificacion)) +
  geom_boxplot(outlier.shape = 21, outlier.size = 3, outlier.color = "red", notch = TRUE) +
  stat_summary(fun = mean, geom = "point", shape = 4, size = 4, color = "blue") +
  geom_jitter(aes(color = Peugeot2$clasificacion), width = 0.2, alpha = 0.5, size = 2) +
  scale_fill_brewer(palette = "Set3") +
  scale_color_brewer(palette = "Dark2") +
  labs(title = "Distribución de Precios por Segmento de Cliente",
       subtitle = "Visualización detallada de la relación entre precios y segmentos",
       x = "Segmentos de Cliente",
       y = "Precios",
       fill = "Segmentos",
       color = "Segmentos") +
  theme_minimal(base_size = 15) +
  theme(legend.position = "top",
        plot.title = element_text(hjust = 0.5, face = "bold", size = 18),
        plot.subtitle = element_text(hjust = 0.5, size = 14))

G2

##########FORCASTING POR TIPO DE CLASIFICACIÓN##########

# Cargar librerías necesarias
library(dplyr)
library(lubridate)
library(randomForest)
library(ggplot2)
library(forecast)

# Convertir la fecha a formato Date
Peugeot2 <- Peugeot2 %>%
  mutate(FECHA = as.Date(FECHA)) # Convertir la variable 'FECHA' a tipo fecha

# Crear la columna de Mes (agregando agrupación por meses)
Peugeot2 <- Peugeot2 %>%
  mutate(Mes = floor_date(FECHA, "month")) # Extraer el mes de la fecha
```

```r
# Sumar ventas por clasificación y por mes
ventas_mensuales <- Peugeot2 %>%
  group_by(Mes, clasificacion) %>%
  summarise(Ventas = sum(VENTA, na.rm = TRUE)) %>%
  ungroup()

# Seleccionar una clasificación específica (puedes iterar sobre otras clasificaciones si lo necesitas)
clasificacion_filtro <- 3 # Cambia a 1, 2 o el número deseado
ventas_clasificacion <- ventas_mensuales %>%
  filter(clasificacion == clasificacion_filtro)

# Verificar que haya datos disponibles
if (nrow(ventas_clasificacion) == 0) {
  stop("No hay datos disponibles para la clasificación especificada.")
}

# Crear la serie temporal
ts_ventas <- ts(ventas_clasificacion$Ventas,
                start = c(year(min(ventas_clasificacion$Mes)),
                          month(min(ventas_clasificacion$Mes))),
                frequency = 12)

# Graficar la serie temporal original
autoplot(ts_ventas) +
  labs(title = "Serie Temporal de Ventas", x = "Tiempo", y = "Ventas") +
  theme_minimal()

# Ajustar el mejor modelo ARIMA automáticamente
modelo_arima <- auto.arima(ts_ventas)

# Resumen del modelo ARIMA
summary(modelo_arima)

# Verificar residuos del modelo
checkresiduals(modelo_arima)
```

```r
# Pronóstico con ARIMA
horizonte <- (2026 - year(max(ventas_clasificacion$Mes))) * 12
forecast_arima <- forecast(modelo_arima, h = horizonte)

# Graficar el pronóstico ARIMA
autoplot(forecast_arima) +
  labs(title = paste("Pronóstico ARIMA para Clasificación", clasificacion_filtro),
       x = "Mes", y = "Ventas") +
  theme_minimal()

# Alternativa: Modelo ETS (Error, Trend, Seasonality)
modelo_ets <- ets(ts_ventas)
summary(modelo_ets)

# Pronóstico con ETS
forecast_ets <- forecast(modelo_ets, h = horizonte)

# Graficar el pronóstico ETS
autoplot(forecast_ets) +
  labs(title = paste("Pronóstico ETS para Clasificación", clasificacion_filtro),
       x = "Mes", y = "Ventas") +
  theme_minimal()

# Descomposición de la serie temporal
descomposicion <- decompose(ts_ventas)
autoplot(descomposicion) +
  labs(title = "Descomposición de la Serie Temporal")

# Ajuste manual del modelo ARIMA
modelo_manual <- arima(ts_ventas, order = c(1, 1, 1), seasonal = c(1, 1, 1))
summary(modelo_manual)

# Pronóstico manual
forecast_manual <- forecast(modelo_manual, h = horizonte)

# Graficar pronóstico manual
```

```r
summary(modelo_ets)

# Pronóstico con ETS
forecast_ets <- forecast(modelo_ets, h = horizonte)

# Graficar el pronóstico ETS
autoplot(forecast_ets) +
  labs(title = paste("Pronóstico ETS para Clasificación", clasificacion_filtro),
       x = "Mes", y = "Ventas") +
  theme_minimal()

# Descomposición de la serie temporal
descomposicion <- decompose(ts_ventas)
autoplot(descomposicion) +
  labs(title = "Descomposición de la Serie Temporal")

# Ajuste manual del modelo ARIMA
modelo_manual <- arima(ts_ventas, order = c(1, 1, 1), seasonal = c(1, 1, 1))
summary(modelo_manual)

# Pronóstico manual
forecast_manual <- forecast(modelo_manual, h = horizonte)

# Graficar pronóstico manual
autoplot(forecast_manual) +
  labs(title = paste("Pronóstico ARIMA Ajustado Manualmente para Clasificación", clasificacion_filtro),
       x = "Mes", y = "Ventas") +
  theme_minimal()

# Comparar modelos
accuracy(modelo_arima)
accuracy(modelo_ets)
accuracy(modelo_manual)
```

DECLARACIÓN Y AUTORIZACIÓN

Yo, **Jiménez Polanco, Angie Stephanie**, con C.C: # 0951140136 y **Lindao Barros, Dennise Nicole**, con C.C: # 0932304678 autor/a del trabajo de titulación: **Análisis de datos para la evaluación del rendimiento y optimización de las ventas de carros de marca Peugeot en el Ecuador,** previo a la obtención del título de **Licenciada en Negocios Internacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, **7** de **febrero** de **2025**

Nombre: **Jiménez Polanco, Angie Stephanie**

C.C: 0951140136

Nombre: **Lindao Barros, Dennise Nicole**

C.C: 0932304678

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TEMA Y SUBTEMA:	Análisis de datos para la evaluación del rendimiento y optimización de las ventas de carros de marca Peugeot en el Ecuador		
AUTOR(ES)	Jiménez Polanco, Angie Stephanie Lindao Barros, Dennise Nicole		
REVISOR(ES)/TUTOR(ES)	Carrera Buri, Félix Miguel		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	**Facultad de Economía y Empresa**		
CARRERA:	**Negocios Internacionales**		
TITULO OBTENIDO:	**Licenciada en Negocios Internacionales**		
FECHA DE PUBLICACIÓN:	**7/2/2025**	**No. DE PÁGINAS:**	**120 p.**
ÁREAS TEMÁTICAS:	**Análisis de datos, Negocios, Marketing y ventas.**		
PALABRAS CLAVES/ KEYWORDS:	Optimización, Análisis de datos, Machine Learning, Forecasting, K-means, Segmentación de mercados, Peugeot, Rendimiento de ventas.		

RESUMEN/ABSTRACT (150-250 palabras):
Este trabajo de investigación se basa en analizar el rendimiento y optimización de las ventas de los vehículos Peugeot en Ecuador. Después de un análisis exhaustivo pudimos notar que actualmente aquí en Ecuador no se le da tanta importancia a la inteligencia artificial en las empresas, es por esto que se ve la necesidad de incrementar técnicas de machine learning para analizar el rendimiento de las ventas de los autos Peugeot, y con esta información poder desarrollar recomendaciones para su optimización. Para poder realizar de manera exitosa este análisis, primero se revisaron conceptos claves para entender más a profundidad los modelos a usar, luego se aplicaron dichos conceptos en la práctica por medio de RStudio donde se realizó primero un modelo de k-means donde se lograron segmentar los clientes en 3 grupos y después en base a esta información se logró predecir las ventas en dólares de los dos años siguientes para cada segmento y de esta manera ver cual era el grupo de clientes que generaba más ingresos para la empresa y cuál era el segmento que menos compraba autos de esta marca. Al final como resultados obtuvimos que el segmento que más autos Peugeot compraba era el segmento 3 de clase alta, pero decidimos dar recomendaciones distintas para incrementar las ventas de cada uno de estos grupos distintos de clientes, enfocándonos un poco más en el segmento con más ventas y en mantener la calidad y esencia de la marca.

ADJUNTO PDF:	☒ SI	☐ NO
CONTACTO CON AUTOR/ES:	**Teléfono:** +593-996820555 +593-990776864	E-mail: **angiejimenezpolanco@gmail.com** **dnlindao12@gmail.com**
CONTACTO CON LA INSTITUCIÓN (C00RDINADOR DEL PROCESO UTE)::	**Nombre:** Freire Quintero César Enrique	
	Teléfono: +593-990090702	
	E-mail: cesar.freire@cu.ucsg.edu.ec	

SECCIÓN PARA USO DE BIBLIOTECA	
Nº. DE REGISTRO (en base a datos):	
Nº. DE CLASIFICACIÓN:	
DIRECCIÓN URL (tesis en la web):	