ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

TÍTULO DE LA INVESTIGACIÓN

Desarrollo de un modelo predictivo para la fijación de precios de renta de vehículos en una de las más grandes empresas de renta de vehículos en Ecuador.

Profesor

Mario Salvador González

Autor

Gibson Steven Andagoya

2023

RESUMEN

El presente trabajo está enfocado en el estudio las dinámicas del precio de vehículos en una empresa de renta de autos o dynamic prycing, el mismo que se ve influenciado por diversos factores cuya definición puede dar como resultado un desempeño positivo como negativo dentro de los ingresos netos de la empresa.

La investigación plantea determinar cuáles son estas variables que comprometen el precio, y en contraposición brinda buscar una solución a través de la definición de un algoritmo de machine learning que permita predecir las tarifas de los vehículos acorde a la influencia de las variables, entre las cuales están la demanda, la temporalidad, y la cantidad de reservas que posee la empresa.

Los datos fueron obtenidos directamente de la empresa objeto de estudio, a través del procesamiento de los datos se determinó que la investigación se basara en dos modelos de vehículos, compactos y económicos, el modelo de regresión lineal en random forest es la mejor opción para el modelo de precios.

A través del Random Forest se determinó el número de estimadores o arboles a utilizar dentro del modelo, también utilizamos el feature importance para determinar las variables más influyentes dentro del modelo, se encontró que la cantidad de contratos abiertos es aquella que tiene más peso al momento de crear nuestro algoritmo, finalmente se determinó a través del R2 y MSE que el modelo con los datos actuales de la empresa no se ajusta a las predicciones debido a una mala definición de precios iniciales por parte de la empresa la cual debe ser corregida a través del web scraping.

ABSTRACT

The present work is focused on the study of the dynamics of the price of vehicles in a car rental or dynamic pricing company, which is influenced by several factors whose definition can result in a positive or negative performance within the net income of the company.

The research proposes to determine which are these variables that compromise the price, and on the other hand, it offers a solution through the definition of a machine learning algorithm that allows to predict the rates of the vehicles according to the influence of the variables, among which are the demand, the seasonality, and the amount of reservations that the company has.

The data were obtained directly from the company under study, through data processing it was determined that the research will be based on two vehicle models, compact and economic, the linear regression model in random forest is the best option for the pricing model.

Through the Random Forest we determined the number of estimators or trees to be used within the model, we also used the feature importance to determine the most influential variables within the model, it was found that the number of open contracts is the one that has more weight when creating our algorithm, finally it was determined through the R2 and MSE that the model with the current data of the company does not fit the predictions due to a poor definition of initial prices by the company which must be corrected through web scraping.

ÍNDICE DEL CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

1. INTRODUCCIÓN

Dentro del sector turistico, la llegada del internet planteo una nueva forma en la cual se desenvolvería el giro de negocio, tradicionalmente los sistemas de reservas se gestionaban directo de la compañía a tarifas establecidas por la misma, sin embargo, hoy por hoy con la globalización y la digitalización el cliente tiene la capacidad de poder comparar precios y características desde la palma de su mano (Montiel, 2018)

Este incremento del poder de negociación del cliente significó para las empresas turísticas un gran ajuste del precio de sus servicios, aun mas en las compañías de renta de autos donde el servicio ofertado es prácticamente idéntico, el incremento del poder del cliente significo una disminución y volatilidad de precios en las tarifas de renta de vehículos, esta volatilidad a su vez es decisiva al momento de la toma de decisión del cliente según su valor percibido (Kotler & Keller, 2016).

Es pues así que en esta investigación buscamos determinar cuáles son las variables decisivas que afectan directamente el precio en busca de incrementar el poder de negociación de la empresa en contraposición de los sitios web y brokers de gestión de reservas en alquiler de coches, nuestro objetivo al determinar estas variables es tener la facultad de generar un algoritmo que utilice el machine learning como arista principal al momento de la determinación del precio.

Este algoritmo a su vez debe utilizar las variables que determinan el precio, que al igual que en las aerolíneas se encuentra la temporalidad y la oferta y demanda del servicio (Tomas & Vega., 2020).

Esto con la finalidad de poder incrementar no solo el poder de negociación de la empresa, sino de optimizar mejor su flota de autos, así como incrementar sus ganancias y ofrecer un mejor servicio al cliente al establecer tarifas dinámicas.

Finalmente, el estudio determinara cuales son las variables y a su vez cual es el mejor modelo que nos permita tener tarifas dinámicas y que puedan ser predichas acorde a nuestras variables independientes, para esto se plantea el uso del random forest como modelo de regresión lineal.

2. REVISIÓN DE LITERATURA

El trabajo realizado busca determinar un enfoque claro sobre la determinación de precios, en este caso la industria es el servicio de renta de vehículos, los estudios planteados como revisión bibliográfica provienen de la experiencia del autor en su trabajo como agente de una compañía de renta de autos, así como información histórica de la demanda de vehículos y la fijación de precios de estos.

La determinación del precio es una de las variables fundamentales dentro de cualquier giro de negocio, el precio usualmente es uno de los principales símbolos de atractivo o desdén de un producto o servicio, la determinación de este puede afectar como beneficiar a una industria, para el establecimiento de precios existen diversos métodos que se han utilizado partiendo de modelos y bases económicas.

Tradicionalmente un modelo de establecimiento de precios era el planteado por Alfred Marshal, en el cual se explica la variación de la demanda de un producto o servicio, esta dependía netamente del aumento o disminución del precio, a través de la elasticidad de la demanda lo que se busca es determinar un precio exacto donde las ganancias se maximicen, (Marshall, 1931).
Este modelo nos permite definir el precio en base a la oferta y la elasticidad de la demanda[1], la misma que se ve influenciada por varios factores, en un estudio realizado sobre la elasticidad de demanda en el precio de repuestos automotrices, se planteó que la elasticidad están influenciada por varios factores característicos de la demanda, desde variables macro a micro económicas, la sensibilidad al precio del consumidor dentro de este estudio determino un precio optimo donde se incrementaban los precios de los repuestos automotrices en un

[1] Sensibilidad al incremento o disminución del precio, la elasticidad de la demanda esta influenciada por varias características de los compradores, como su ingreso, su gasto, expectativa del producto entre otros (Garzon, Lozada, & Monroy., 2018)

7,9% sin embargo esto no significaba un consumo menor o mayor de los repuestos automotrices (Garzon, Lozada, & Monroy., 2018).

Esto significaría que muchas veces la sensibilidad depende del tipo de productos o servicios objeto de estudio, usualmente en los commodities, la sensibilidad al precio suele ser mayor, también en bienes de consumo primario o de primera necesidad, ya que afectan directamente al consumidor, sin embargo, para el enfoque de este estudio, la oferta de servicio tiene un enfoque más suntuario y de ocio, el mismo que se ve influenciado por variables diferentes tanto macro como microeconómicamente.

Por otro lado, otra estrategia de determinación de precios está enfocada en el proceso constructivo de una hoja de costos, tradicionalmente las empresas de manufactura realizan un estudio de costos al cual al precio bruto se le añade un porcentaje de la ganancia esperada, esta determinación de precios es típica de empresas de producción de bienes (Collin, 2015).

Collins plantea la contabilidad de costos como un medio per se dé la fijación de un precio, el libro tiene su enfoque en la determinación del precio óptimo de un producto o servicio, así como las implicaciones de los costos fijos y variables, el objetivo principal es buscar un punto óptimo de producción que permita ajustarse a la demanda tradicional de Marshal y que al mismo tiempo permita generar un porcentaje de utilidad neta de ingresos.

La contabilidad de costos es un enfoque óptimo para el control de los recursos, la administración y la toma de decisiones, sin embargo, Collins no determina un enfoque exacto en la contabilidad de costos enfocados a los servicios, ya que el costo de los servicios no se establece netamente por el número de recursos utilizados, el precio de este es más subjetivo y está atado a varias aristas distintas a los de la producción de un bien.

Bajo ambos preceptos encontramos diversas maneras de fijar el precio de un bien, sin embargo, no para la determinación del precio del servicio, nos apoyaremos de estas teorías y adicionaremos una ultima la cual tiene que ver con el valor percibido por el mismo, esta teoría es fundamental debido a que el servicio muchas veces establece su precio a través del valor percibido por el cliente.

Según Kotler, en el capítulo 5 sobre como establecer relaciones con los clientes basadas en la fidelidad y el largo plazo, el valor percibido es esencia todos aquellos beneficios que el cliente evalúa pre contra de un bien o servicio, dentro de esta evaluación no solo se inmiscuye el precio, sino que también forman parte los beneficios psicológicos, sociológicos el costo de los mismos y al relación entre las diferentes ofertas de mercado ergo permite tomar una decisión que maximice su satisfacción (Kotler & Keller, 2016).

Para un mejor entendimiento de esto, muchas veces la percepción del valor del cliente no está en el precio sino en los beneficios per se que el cliente puede sacar de un producto donde considere que su decisión maximiza su satisfacción, un ejemplo claro es la tecnología de Apple, cuyo enfoque está en la calidad de sus productos, pero no solo en eso sino también en satisfacer necesidades específicas de sus clientes, sean estas, diseño, componentes, y acoplarse plenamente a sus necesidades.

El precio dentro de estos productos es secundario, el consumidor considera que el precio es acorde al bien que está adquiriendo, de igual manera en el negocio de renta de autos el valor percibido por el cliente muchas veces no está en el precio, el precio es simplemente un atractivo primario al momento de gestionar una reserva, el servicio per se es aquel que incrementa el valor de la renta de vehículos o la disminuye.

Dado que estos tres planteamientos son subyacentes podemos utilizar varios de estos para definir las determinantes del precio, las mismas que pueden ser la demanda de autos, el valor percibido por el cliente y el costo del servicio.

Por otro lado la percepción del valor del cliente muchas veces se ve distorsionada, según Poundstone, muchas consumidores tienen una percepción alterada del precio, esto puede jugar a favor o en contra de nuestros objetivos empresariales, si bien Kotler nos plantea que el cliente es en su máximo esplendor un comprador nato con habilidades informativas gracias a la Red Poundstone nos plantea que muchos de estos se guían en base a estrategias ficticias o artificiales que pueden incrementar la perspectiva del valor de un bien o servicio.

En la parte de su libro “Priceless” Anchor for Dummies (Poundstone, 1995), la perspectiva del valor del cliente se ve distorsionada por precios “Ancla” los cuales son básicamente el establecimiento de precios flat o fijos bastante altos en relación con su costo real y cualquier disminución de los mismos, lo que general que aunque el costo posterior sea inferior al flat Price el precio real sigue aún muy ínfimo y la ganancia aún muy alta, pero la perspectiva del consumidor se torna positiva al encontrar un excelente precio (Poundstone, 1995).

Esto en la renta de autos es conocido como la tabla RACK o tabla de precios fijos, que es muy superior a los precios establecidos en los brókeres de renta de autos, es por tal motivo que los clientes optan por realizar reservas en línea donde los precios comparativos son extremadamente bajos en comparación con la tabla rack.

Sin embargo, la problemática gira entorno a la determinación de este precio bajo en comparación con el precio ancla, si bien el precio bajo genera un atractivo a los clientes, el objeto es encontrar cuan bajo debe ser ese precio para ser competitivo y al mismo tiempo ser el precio más alto para ofertar al cliente.

Neagle, plantea el establecimiento del precio de competencia como Dynamic Prices, estos son básicamente precios que se ajustan en tiempo real, acorde a varias circunstancias, sean estas la demanda, el cliente, la ubicación geográfica, el tiempo entre otros (Neagle & Muller., 2018).
Neagle plantea un ejemplo muy interesante que tiene que ver con Álamo (nuestra competencia), Álamo plantea los Dynamic Prices debido a la segmentación de mercado, la compañía a pesar de ser una de las más pequeñas en ese entonces en USA, utiliza la estrategia de Dynamic Prices para ofrecer los mejores precios a sus clientes, siendo aun mejores que sus competidores Avis y Hertz, la estrategia se basa en ofertar diferentes precios acorde al segmento de mercado, el segmento variaba acorde a la temporada, por ejemplo cuando la temporada apuntaba hacia el renting corporativo, los precios automáticamente subían, dado que este segmento tenía una mayor capacidad adquisitiva, por otro lado cuando el renting era enfocado en la parte turística, el precio automáticamente baja dado que el segmento era más familiar y ahorrativo, esto permitió a Álamo brindar precios más bajos dado que compensaban con precios altos en otros segmentos.

Los Dynamic Prices de Neagle y Müller, son en si una composición conjunta de varios de los métodos de fijación de precio, sin embargo, este se acopla totalmente a nuestro giro de negocio, en el encontramos la influencia y sensibilidad de los precios de Marshall (Marshall, 1931), también el valor percibido de Kotler (Kotler & Keller, 2016) y el constructivo de Collins.
Una aproximación de estudios que sirven como base a esta investigación son aquellos enfocados al precio de boletos de aerolíneas, las mismas que siguen un patrón de comportamiento similar al de la renta de autos, si bien el fin es el mismo (movilización) el medio es diferente, a través de estos estudios determinaremos la importancia de la fijación de un precio dinámico que incremente la rentabilidad de la empresa y a su vez sea el precio más competitivo.

En su estudio sobre las principales barreras de decisión para la compra de boletos de aerolínea digital en México, Montiel Solís, determina que un 74% de

las personas que compra boletos por internet están de acuerdo debido a que pueden encontrar mayor información a través de la web y de igual manera encuentran mejores precios (Montiel, 2018), las aerolíneas al igual que la renta de autos utilizan los Dynamic Prices en sus sitios web los cuales ofertan precios acorde a diferentes factores como la demanda, el tipo de asiento, la anticipación de compra entre otros, Montiel plantea en su hipótesis 4 que el precio es un factor decisivo al momento de optar hacer la reservas a través de la web.

Esto tiene especial relevancia dado que el manejo de reservas hoy por hoy se realiza a través de los brókeres de internet, tradicionalmente ir directamente con el proveedor del servicio tenía costes significativamente más bajos, pero con el anchor Price y el Dynamic Price, hoy por hoy se pueden encontrar mejores ofertas a través de metabuscadores en internet cuyas tarifas no son fijas como las de la empresa directa.

Tomas & Vega, 2019 plantean por otro lado que parte de las estrategias de fijación de precio dentro de las aerolíneas tiene que ver con un estudio de segmentos de temporalidad anual, básicamente los precios se fijan acorde al segmento del cliente, al igual que Álamo, el ejecutivo y el turista poseen diferentes precios, esto de la mano con el tiempo de reserva, la cantidad de equipaje y la disponibilidad de asientos (Tomas & Vega., 2020).

En retrospectiva en el negocio de renta de autos la fijación de precios es similar, la cantidad de equipaje se resume a la capacidad de carga del vehículo, sea en pasajeros o a su vez en maletas, entre mayor es esta, el precio incrementa, también acorde a la disponibilidad de los autos el precio aumento o disminuye, finalmente la temporada influye también en el precio de los autos.
Una vez ejemplificada la problemática subyacente y las características que definen el precio, procederemos a describir las metodologías que usan algunos autores para poder realizar una predicción de precios para la implementación de un Dynamic Pricing.

El Dynamic Pricing nos permitirá maximizar nuestras ganancias a través de una correcta determinación de las variables del precio, la predicción no es algo nuevo en el ámbito de los negocios, tradicionalmente está relacionada con la predicción de precios en la bolsa lo cual permite generar un redito de la compra y venta de las mismas, para ejemplificar como la predicción puede ayudar a maximizar el rendimiento Broncano, 2022 nos plantea que la variabilidad de los precios en este caso de las acciones de Apple dependen de varios factores, sin embargo la predicción de las acciones pueden ser descritas a través de varios modelos de machine Learning, En su estudio sobre el machine learning para predecir valores de acciones, plantea que el modelo más eficaz para la predicción de las acciones es el de regresión lineal o la técnica de Extreme Gradient Boosting, esto se basó realizando una comparativa de los resultados de otros modelos como el ARIMA, el objetivo de esta investigación era determinar el mejor método predictivo en cuanto a precios de acciones (Broncano, 2022).

Este estudio nos brinda una pauta principal de que el método de regresión lineal es probablemente el mejor método al momento de realizar una predicción del precio, por otro lado aunque no es nuestro giro de negocio, Medina (2020), plantea en su estudio sobre predicción de precios de viviendas de igual manera que la mejor metodología para la predicción de precios es en sí la regresión lineal, Medina plantea un método de aprendizaje supervisado con un 80% de datos que serán usados para entrenar el modelo y el 20% serán usados para prueba (Medina, 2020).

Sin embargo, los datos de prueba y entrenamiento los veremos posteriormente en el apartado de metodología.

3. IDENTIFICACIÓN DEL OBJETO DE ESTUDIO

El objeto de estudio de esta investigación se centra en el desarrollo, evaluación y aplicación de un modelo de regresión lineal para predecir los precios de renta de automóviles en la industria de alquiler de vehículos.

El estudio será enfocado a determinar las variables que afectan directamente a las tarifas de renta de autos, al igual que las aerolíneas el precio está definido por variables ajenas al control de esta, esto con el objetivo de establecer precios dinámicos que puedan elevar la rentabilidad de la empresa.

Los datos serán obtenidos directamente de la empresa de renta de autos, con el objetivo de poder predecir las tarifas acordes a las variables planteadas, esta descripción de precios permitirá anticiparse a una subida o bajada de precios acorde al comportamiento del mercado, asegurando la competitividad tanto en precios como en la rentabilidad de la empresa.

4. PLANTEAMIENTO DEL PROBLEMA

La industria de la renta de autos ha experimentado un cambio significativo en las últimas décadas, con la adopción generalizada de precios dinámicos (Neagle & Muller., 2018) como una estrategia para maximizar los ingresos y optimizar la utilización de flotas de vehículos. Esta práctica implica la variación de las tarifas de alquiler en tiempo real en función de una serie de factores, como la demanda, la temporada, la ubicación y otros datos relevantes. Aunque los precios dinámicos los cuales ofertan beneficios tanto para las empresas como para los clientes.

La problemática gira entorno a la maximización del uso de recursos a través de la asignación de precios de estos los cuales deben ser establecidos de manera proactiva mas no reactiva, el problema yace en que desde la aparición de la web y sobre todo de las plataformas de brókeres en línea, el precio del servicio de renta de autos comenzó a ser variable y no a estar establecido directamente por el proveedor del servicio, sino más bien por factores ajenos al control de la compañía.

En ese sentido los Stakeholders como clientes y brókeres comenzaron a tener mayores cantidades de poder de negociación en cuanto a las fuerzas de Porter (Porter, 1980), esto hace a su vez exista una gran barrera de ingreso para pequeñas empresas al mundo del renting internacional, debido a que las tarifas son establecidas a través de la oferta y la demanda, muchas de estas se encuentran en precios extremadamente bajos, lo cual hace que la empresa requiera de cierta cantidad de vehículos para poder sostenerse, a estas tarifas tan bajas una empresa con una flota pequeña se vuelve inviable, por lo que la cantidad de autos es imprescindible para poder subsistir, similar a las economías de escala, entre más autos, es más fácil desglosar los gastos pudiendo proporcionar el servicio a costos más asequibles (Castro, 2009).

Bajo este precepto dado que es una empresa de renting internacional, puede asumir estos costos, así como proporcionar el servicio, también se potencia a

través de la venta de upsales y la cantidad de autos es un gran referente que nos brinda ventaja competitiva. Sin embargo lo que se busca es maximizar estas tarifas, el problema también gira en torno a que las tarifas se establecen a través de factores como el precio de la competencia (Montiel, 2018), a precios más económicos, se abarca una mayor cantidad de clientes, entre las grandes compañías de renting existe una lucha eterna entre definir qué empresa es más proactiva definiendo el precio de sus vehículos a través del tiempo, bajando los mismos en temporadas malas y encareciéndolos en periodos de bonanza

5. OBJETIVO GENERAL

Determinar las variables que afectan a las tarifas de la renta de autos, para poder generar un algoritmo de regresión lineal con machine learning, que pueda generar precios dinámicos que maximicen la rentabilidad, a través de la obtención de la información directa de la compañía, en el periodo 2023 - 2024

6. OBJETIVOS ESPECÍFICOS

- Contextualizar los fundamentos de la regresión lineal y el Machine Learning.
- Contextualizar los fundamentos del Dynamic Pricing.
- Determinar el mejor algoritmo de machine learning para la predicción de precios
- Diagnosticar las variables influyentes en el precio de las tarifas de renta de autos.
- Determinar las factibilidad de elaboración de un modelo de machine learning que pueda predecir precios de tarifas de renta de autos con tarifas dinámicas.

7. JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA

Previamente se hizo un acercamiento hacia la importancia del precio sobre la determinación del ingreso dentro de la compañía, si bien existen varios factores macro y microeconómicos que afectaran de manera positiva o negativa el desempeño de la empresa, en el caso de las rentadoras de coches, el precio es una variable decisiva para el cliente al momento de elegir una opción u otra.

Por tal motivo el siguiente proyecto tendrá como finalidad crear un modelo predictivo que permita definir el precio más competitivo y que de igual manera incremente la rentabilidad de la empresa.

El análisis de igual manera pretende demostrar que variables son las más influyentes dentro de la definición del precio, el anticipar las tarifas permitirá generar un mayor número de reservas, a un mejor precio, por otro lado, la determinación de variables influyentes dentro de la organización permitirá tener una mejor estrategia operativa.

Parte importante del proyecto también es dar visibilidad a la analítica como parte esencial del negocio, a pesar de ser una organización considerada como PYME la misma aun toma decisiones en base a la experiencia mas no en base al análisis de resultados, esto hace que muchas de las decisiones sean acertadas y por otro lado otras tengan serias repercusiones en el desempeño de las actividades, ejemplo claro es la gran inversión que se hacen en la renovación de la flota de vehículos sin un estudio de rentabilidad de los mismos, que a largo plazo hace que varios autos se encuentren parqueados y exista escasez en otra categoría de vehículos.

7.1. Recolección De Datos.

7.1.1. Base de datos.

La base de datos está conformada por observaciones recopiladas desde el 10 de enero del 2020 hasta la actualidad, la misma consta de los valores diarios de las variables independientes y el precio diario que es la variable dependiente. Se obtiene esta información de dos herramientas fundamentales dentro de la empresa, la primera es TSD, la cual nos arroja información sobre el estado de la flota, de esta herramienta obtenemos toda la información diaria de la misma, y en cuanto a la información financiera, optamos por usar la herramienta “latinium”, la cual es el sistema da facturación aprobada por el SRI, en esta podemos obtener el informe real de pagos diarios efectuados y acreditados a la cuenta de la empresa.

7.1.2. TSD Rental.

TSD es una herramienta enfocada como una solución a las empresas de movilización, específicamente a aquellas que se dedican a la renta de vehículos, esta herramienta gestiona desde el inicio hasta el final del proceso de la renta, es decir, desde cuando el cliente realiza la reservación e ingresa toda su información personal hasta cuando el cliente ya entrega el vehículo una vez acabado su periodo de renta.

7.1.3. Tipos de datos.

Los tipos de datos que nos ofrece TSD, son datos estructurados y nos ofrece información del consumidor a nivel global el primer paso para la renta del vehículo viene a través de la reservación, TSD se encarga de gestionar y recibir las reservaciones de clientes de todo el mundo a través de códigos de reserva dentro de esta información se encuentran datos demográficos del cliente como su edad el país su género sus números de teléfono y correos electrónicos y también recopila información financiera de tarjetas de crédito, ya que en este negocio es un requisito fundamental poseer una tarjeta de crédito para realizar un hola que queda como garantía para la renta del vehículo.

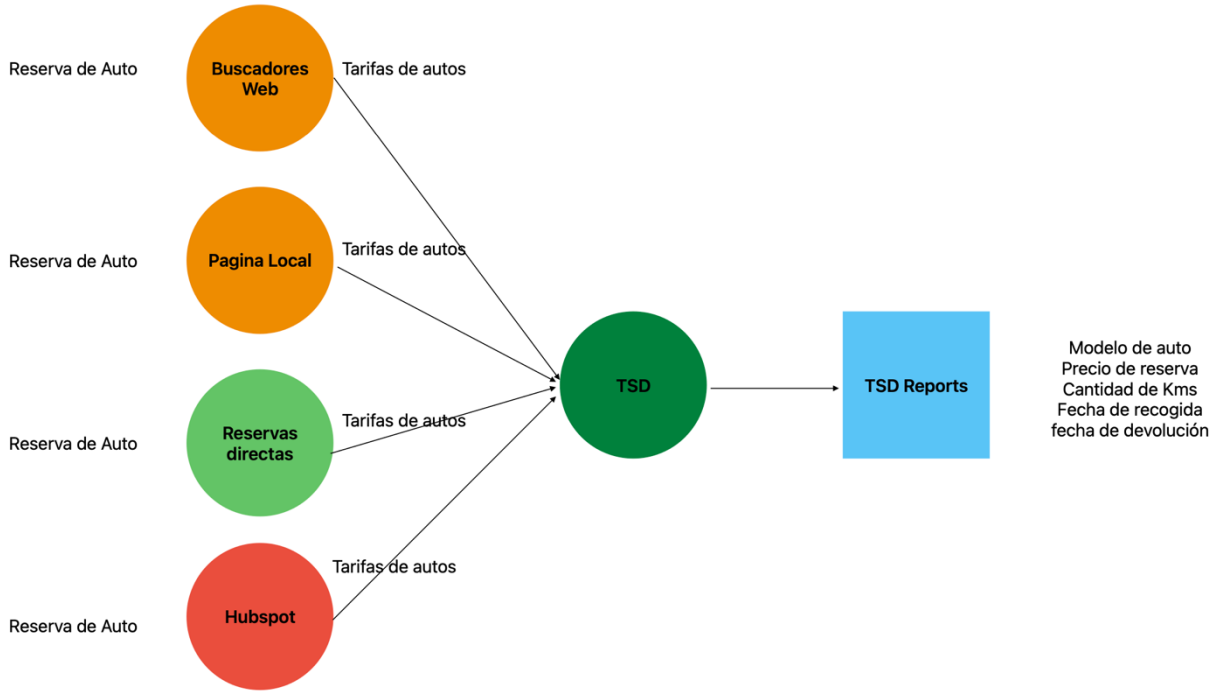
También recopila información sobre la demanda de los servicios que ofrecemos el precio de la renta de los vehículos varía dependiendo del tipo de vehículo existen: vehículos económicos pequeños hasta los más costosos en tipo SUV.
Finalmente recoge información sobre el desempeño de los vehículos, su recorrido en kilómetros, su rentabilidad, su uso, número de reparaciones y daños, entre otros.
TSD presenta toda esta información de manera estructurada en tablas las cuales pueden ser utilizadas para realizar diversos tipos de cálculos y en tiempo real.

7.2. Obtención de los datos.

Para obtener los datos, extraeremos la información directo de TSD, la misma que contendrá las tarifas históricas así como otras variables importantes las cuales se determinaron importantes dentro del análisis exploratorio, entre las cuales tenemos el número de unidades disponibles por día, también el número de unidades rentadas, la cantidad de reservas, el porcentaje de no show, la cantidad de contratos abiertos y contratos cancelados, a través de un análisis de correlación determinaremos la importancia de estas variables dentro de nuestro modelo de regresión.
Por otro lado, dado que existen varios modelos de vehículos, clasificamos los mismos dentro de las siguientes categorías:

- EDMR: Económico Pequeño 5 pasajeros (Kea picanto o Chevrolet Spark)
- CDMR: Compacto Sedan 5 pasajeros (Kia Soluto, Kia Rio)
- FFMR: SUV 5 pasajeros (Susuki Vitara, Chevrolet groove, Toyota Rize)
- MVAR: VAN 7 pasajeros (Kia Carnival, Hyundai Staria)
- LFAR: Suv Lujo (Toyota Fortuner, Chevrolet Tahoe

Ilustración 1 Modelo de recopilación de información



Fuente: Dataset

Elaborado por: Autor.

7.3. Limpieza y pre-procesamiento de datos.

Parte importante del proceso de recolección de los datos es las diversas bases de datos que obtenemos debido a las diferentes categorías de vehículos, sin embargo una ventaja que poseemos es que los datos que nos arroja el aplicativo son de manera estructurada por lo cual facilita la relación entre las diversas bases de datos, la estabilidad y la estructura de los datos nos permite generar un modelo de relación entre las bases de datos dado que siguen un mismo formato establecido (Oppel & Sheldon, 2009).

Por tal motivo el primer paso de nuestro proceso de pre-procesamiento será la unificación de nuestras diversas bases de datos por vehículos en una sola que contenga toda la información, dado que todas poseen las mismas variables y la

misma línea temporal se acoplaran bajo las mismas fechas sin embargo lo que cambiaría sería el modelo de vehículo y los valores del mismo.

7.4. Eliminación de variables.

Una vez definido la unificación de las diversas bases de datos procederemos a determinar que variables son aquellas que nos lanza por defecto TSD, dentro de estos definiremos cuales serán nuestras variables importantes, que columnas poseen valores en 0 y no aportan nada a nuestra investigación y también otras variables ya procesadas que nos ofrece TSD.

Antes de proceder a la eliminación de variables se detalla a continuación todas las variables que posee nuestro DATASET, se presenta a continuación una tabla donde se definirá cada variable, se dará un concepto de la misma y si procede o no a la eliminación de la misma.

Tabla 1 variables iniciales

Variable	Ejemplo	Tipo de variable	Subtipo de variable	Descripción	Eliminación
Date	1/10/20	Cuantitativa	Temporal	Fecha de definición del precio	No
Class	EDMR	Cualitativa	Ordinal	Clase de vehículo	No
Units On Rent	10	Cuantitativa	Continua	Unidades en Renta	No
MTD Units On Rent	10	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Units Sitting	11	Cuantitativa	Continua	unidades disponibles	No
MTD Units Sitting	11	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Useable Fleet	24	Cuantitativa	Continua	Unidades en uso	No
MTD Useable Fleet	24	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
In Maintenance	3	Cuantitativa	Continua	Unidades en mantenimiento	No
Non-Useable Fleet	0	Cuantitativa	Continua	unidades no disponibles	Si
MTD Non-Useable Fleet	0	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Total Units in Fleet	24	Cuantitativa	Continua	Total de unidades en la flota	No
Useable Fleet Utilization	41.67%	Cuantitativa	Continua	Utilización de la flota	Si
MTD Useable Fleet Utilization	41.67%	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Total Fleet Utilization	41.67%	Cuantitativa	Continua	Total de unidades utilizadas	Si
MTD Total Fleet Utilization	41.67%	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
MTD Avg Useable Fleet	24	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
MTD Avg Total Units	24	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
MTD Total Units	24	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
Reserved	0	Cuantitativa	Continua	Reservadas	No
MTD Reserved	0	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Reserved Avg Rate	0	Cuantitativa	Continua	precio de reserva	Si

MTD Reserved Avg Rate	0	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Rez Cancelled	0	Cuantitativa	Continua	Reservas canceladas	No
MTD Rez Cancelled	0	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
No Shows	0	Cuantitativa	Continua	Clientes que no se presentaron	No
MTD No Shows	0	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
% of No Shows	0.00%	Cuantitativa	Continua	Porcentaje de clientes no presentados	Si
MTD % of No Shows	0.00%	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
R/A Opened	1	Cuantitativa	Continua	Contratos Abiertos	No
MTD R/A Opened	1	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
Opened Avg Rate	44,64	Cuantitativa	Continua	Tarifa de contratos abiertos	No
MTD Opened Avg Rate	44,64	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Opened later Voided	0	Cuantitativa	Continua	Contratos abiertos cancelados	Si
MTD Opened later Voided	0	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
R/A Voided	0	Cuantitativa	Continua	Contratos cancelados	Si
MTD R/A Voided	0	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
R/A Closed	3	Cuantitativa	Continua	Contratos cerrados	SI
MTD R/A Closed	3	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
Closed Revenue Days	19	Cuantitativa	Continua	Extension de dias cerrados	SI
MTD Revenue Days	19	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
Time & Kilometers	758,65	Cuantitativa	Continua	Tiempo y kilometros de uso	SI
MTD Time & Kilometers	758,65	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Rate Charge	758,65	Cuantitativa	Continua	Total de la renta	SI
MTD Rate Charge	758,65	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Time & Kilometers + Selected	758,65	Cuantitativa	Continua	Tiempo y kilometros seleccionados	Si
MTD Time & Kilometers + Selected	758,65	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
Break Out	76,78	Cuantitativa	Continua	Otros	SI
MTD Break Out	76,78	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
Selected	758,65	Cuantitativa	Continua	Otros	SI
MTD Selected	758,65	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
Tax & Surcharge	108,22	Cuantitativa	Continua	Impuestos	SI
MTD Tax & Surcharge	108,22	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
Total Charges	943,65	Cuantitativa	Continua	Total de cargos	SI
MTD Total Charges	943,65	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
Daily Avg Time & Kilometers	39,93	Cuantitativa	Continua	Promedio de uso	SI
MTD Daily Avg Time & Kilometers	39,93	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
MTD Time & Kilometers per Unit	31,61	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	SI
MTD Daily Avg Selected	39,93	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
MTD Selected per Unit	31,61	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
Daily Avg Time & Kilometers + Selected	39,93	Cuantitativa	Continua	Promedio de uso y kilometros	Si
MTD Daily Avg Time & Kilometers + Selected	39,93	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si
MTD Time & Kilometers + Selected per Unit	31,61	Cuantitativa	Continua	MTD es el acumulativo de la variable anterior	Si

Elaborado por: Autor.

Como podemos observar en la tabla anterior TSD por defecto nos muestra 63 variables que son las etiquetas de nuestros datos, sin embargo muchos de esto son el resultado de operaciones de datos primarios, por ejemplo todas aquellas con “MTD” representan el acumulativo histórico de la variable subyacente, si estamos hablando de reservas MTD será la sumatoria en el tiempo de las reservas hasta la fecha, todas estas variables resultantes de un proceso de tratamiento de datos originales procederemos a eliminarlas ya que se tratar solo las variables fundamentales dentro de nuestro procesamiento de datos.

Por otro lado muchas de las columnas eliminadas tiene que ver con aquellas que tienen datos en 0, algunas de estas tienen más del 80% de datos superiores a 0, estas variables lejos de proporcionarnos información valiosa, tan solo distorsionarían nuestro estudio, un ejemplo de esto son los contratos cancelados por un arribo tardío, estos valores aparecen en 0 debido a que los agentes de venta no pueden cancelar las reservas, estas deben ser canceladas directamente por el bróker o por el cliente, por lo que estos valores casi siempre salen en 0.

7.5. Dataset a usar.

Al inicio se definió que existen 4 categorías de vehículos, el inconveniente es que de estas 5 categorías tres de estas no se rentan diariamente, TSD registra las tarifas y precios del número de contratos abiertos de la categoría de vehículos, desafortunadamente si no se abre ningún contrato o se ingresa una reserva TSD no registra un precio de renta, aunque este esté definido en los portales web, esto significaría una cantidad enorme de datos perdidos en las categorías de auto grandes, de lujo y VAN, ya que estos son destinados a un segmento específico de clientes los cuales rentan por ciertas temporadas o son poco usuales.

Bajo este precepto se decidió utilizar dos categorías de autos, las cuales se rentan siempre al menos una vez al día, lo que asegura una consistencia en nuestros datos, estas clases de vehículos son económicos y compactos.
Una vez que descargamos la base de datos de ambas categorías, procedemos a eliminar aquellas que no son de nuestra utilidad y unificamos ambas bases de datos en una sola.

7.5.1. Adición de variables.

Algo importante a considerar es que tenemos las fechas exactas de la definición de precios, aunque podemos obtener los datos desde el 2018, nuestro estudio estará enfocado al análisis y uso de datos desde el año 2021 hasta la actualidad, esto debido a que debido a la pandemia existen varios datos atípicos dentro del año 2020 y también varios datos perdidos dado que en ese año no ingresaban reservas por lo que los valores perdidos son considerables y afectaran a. nuestro modelo.

Las variables que se añadió al estudio tienen que ver con la temporalidad, de la fecha unificada, desglosamos los días de la semana, el número de semana y el mes, esto debido a que la temporalidad es un factor clave en la definición de precios, como se manifestó en el marco teórico, el precio depende de la temporada, en temporadas de alta demanda como fechas festivas los precios suelen ser superiores que en temporadas bajas.

Una vez eliminadas aquellas variables que serán inservibles en nuestro modelo utilizaremos varias librerías de Python que nos facilitarán el proceso de limpieza, así como de análisis e interpretación de resultados.

7.5.2. Valores perdidos.

Cargamos nuestra base de datos a Python la cual tendrá el nombre de DATASET.csv, importamos nuestras librerías entre las cuales utilizaremos:

- Numpy
- Pandas
- Matplotlib
- Seaborn

Desplegamos nuestro Dataset con df. head para visualizar si la base de datos se cargó correctamente.

Buscamos valores perdidos dentro de nuestro Dataset.

Python nos muestra que no existen valores perdidos dentro del Dataset, sin embargo en el despliegue podemos observar varias columnas con valores en 0, aparentemente estos datos deberían ser valores perdidos sin embargo el 0 es información valiosa para nuestro estudio, esto nos indica por ejemplo que no existen reservas, o que no existen reservas canceladas, para objeto de nuestro estudio esto es parte fundamental, la inexistencia de reservas o de autos rentados significaría una gradual disminución del precio para fomentar la generación de reservas, sin embargo procederemos a realizar el conteo de los valores en 0 de todas las variables para realizar un análisis más profundo.

Con count zeros, podemos determinar la sumatoria de todos los valores en 0 de nuestras variables, vemos que existen 0 en disponibles, esto significaría que en ciertas épocas todos estos autos están rentados y ninguno está disponible, tenemos 89 veces 0 en mantenimiento lo que significa que muchas veces existe al menos un auto en mantenimiento, 666 veces 0 en Reservadas, lo que significaría que existen fechas donde no existen reservas de estas categorías, En canceladas y en no show es donde tenemos más presencia de 0, esto debido a que prácticamente el no show no se da por la empresa sino que el bróker lo registra trimestralmente en la conciliación de reservas.

En donde debemos ser cuidadosos es en el conteo de valores en 0 en el precio, ya que esto significaría en primera instancia que ese día no se rentó esa categoría de auto, o que el precio no está definido, sin embargo, en nuestra variable 0 no existen valores faltantes ni tampoco valores en 0, por lo que consideramos que la base de datos es sólida en cuanto a valores perdidos.

Algo curioso es que existen 181 contratos en 0, eso dignificaría que aparentemente no debería haber precio en 181 celdas de precio debido a que si

no se abrió contratos el sistema no debería reflejar un precio, sin embargo los precios también están dados por las reservas, eso significaría que quizá se abrió el contrato pero por motivos de seguridad se tuvo que declinar o cerrar el contrato, esto sería por ejemplo si el cliente no cumpliera con alguno de los requisitos, el más común el no tener tarjeta de crédito, lo cual impediría que se celebre el contrato de arrendamiento.

7.6. Definición De Variables.

7.6.1. Variable dependiente.

La variable que buscamos explicar es el precio promedio o tarifa, el cual se ve afectado por las variables ya descritas anteriormente, los ingresos están conformados por todo aquello generado durante el periodo de renta, dado que la temporalidad es parte fundamental en el proceso de la renta, se tomó como referencia los pagos diarios realizados mas no el cobro de excedentes al cierre del contrato, estos excedentes pueden ser kilómetros extras, recargo por peajes o en su defecto horas extras.

7.6.2. Variables independientes.

Las variables independientes, son todas aquellas que afectarían nuestra tarifa diaria las cuales serían:

Tabla 2 Variables del modelo.

Variable	Tipo	Subtipo	Detalle	Fuente
Date	Cuantitativa	Continua	Fecha donde se define el precio de la renta del auto, y la tarífa	Obtenidos directo de la empresa, CRM TSD.
Class	Cualitativa	Ordinal	Es la categoría del vehículo, en este caso sera económico o compacto	Obtenidos directo de la empresa, CRM TSD.
Dia	Cualitativa	Ordinal	Dia de la semana, se utiliza esta variable dado que en ciertos dias de la semana suele haber mayor demanda	Obtenidos directo de la empresa, CRM TSD.
Semana	Cuantitativa	Discreta	La cantidad de semanas que tenemos en el año, esto para poder hacer un análisis de temporadas de mayor afluencia	Obtenidos directo de la empresa, CRM TSD.
Mes	Cualitativa	Ordinal	El mes del año	Obtenidos directo de la empresa, CRM TSD.
Rentados	Cuantitativa	Discreta	Cantidad de autos que se encuentran rentados	Obtenidos directo de la empresa, CRM TSD.
Disponibles	Cuantitativa	Discreta	Cantidad de autos disponibles	Obtenidos directo de la empresa, CRM TSD.
Mantenimiento	Cuantitativa	Discreta	Cantidad de autos en mantenimiento	Obtenidos directo de la empresa, CRM TSD.
Totalflota	Cuantitativa	Discreta	Total de la flota en determinados autos	Obtenidos directo de la empresa, CRM TSD.
Reservadas	Cuantitativa	Discreta	Cantidad de unidades reservadas	Obtenidos directo de la empresa, CRM TSD.
Canceladas	Cuantitativa	Discreta	Cantidad de reservas canceladas	Obtenidos directo de la empresa, CRM TSD.
NoShow	Cuantitativa	Discreta	Cantidad de clientes que no se presentaron a tomar su reserva.	Obtenidos directo de la empresa, CRM TSD.
Contratos	Cuantitativa	Discreta	Cantidad de contratos abiertos	Obtenidos directo de la empresa, CRM TSD.

Fuente: Dataset

Elaborado por: Autor.

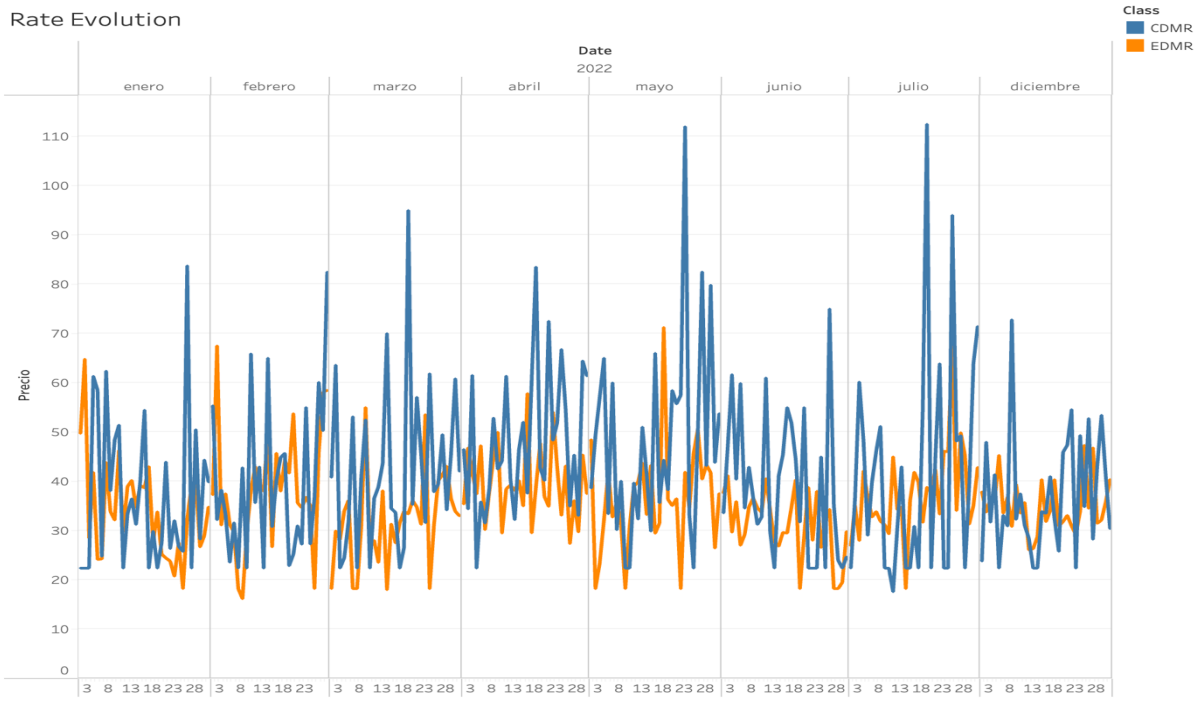
7.7. Visualización de Variables.

En este apartado se representará de manera gráfica todas nuestras variables, esto con el fin de determinar la importancia de nuestro estudio, la gráfica de las variables, nos permitirán ver a través de ilustraciones la problemática que aqueja a la compañía.

7.7.1. Rate Evolution.

El primer punto es determinar cómo se distribuye el precio, a través del tiempo, dado que tenemos observaciones por cada día a través del año optamos por utilizar Tableau para graficar la distribución del precio.
Dado que son varias observaciones optamos por utilizar solo al año 2022.

Ilustración 2 Evolución del precio



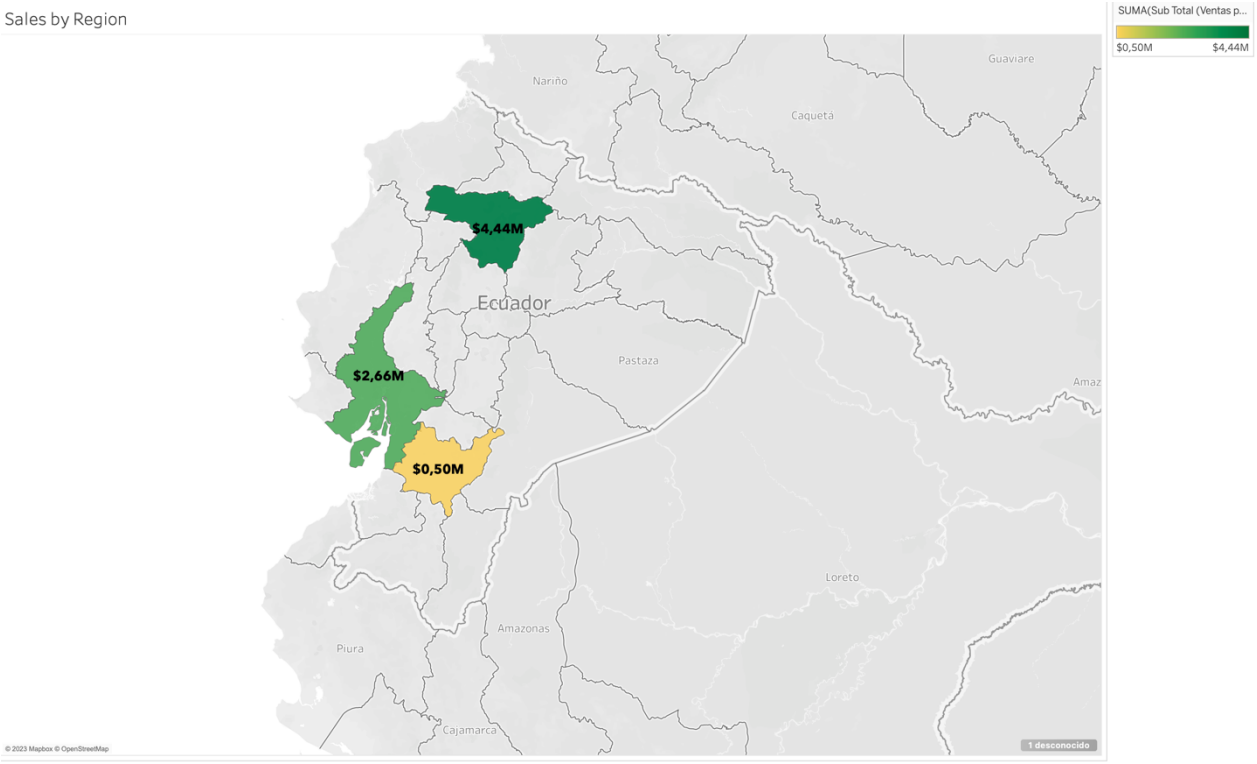
Fuente: Anexo 1, Rate Evolution
Elaborado por: Autor.

En la ilustración podemos observar cómo es que el precio se altera dependiendo de la temporada, teniendo picos en enero, marzo, junio y de diciembre a enero, en ambas categorías, la variabilidad del precio es lo que en primera instancia fue nuestro objeto de estudio, el determinar las mejores tarifas a futuro dependiendo de la oferta y la demanda es primordial.

7.7.2. Ventas Regionales.

La región también es parte importante en cuanto al incremento de ventas, dependiendo la región la empresa genera más o a su vez menos ingresos.

Ilustración 3 Ventas por región.



Fuente: Anexo 2, Sales Evolution

Elaborado por: Autor.

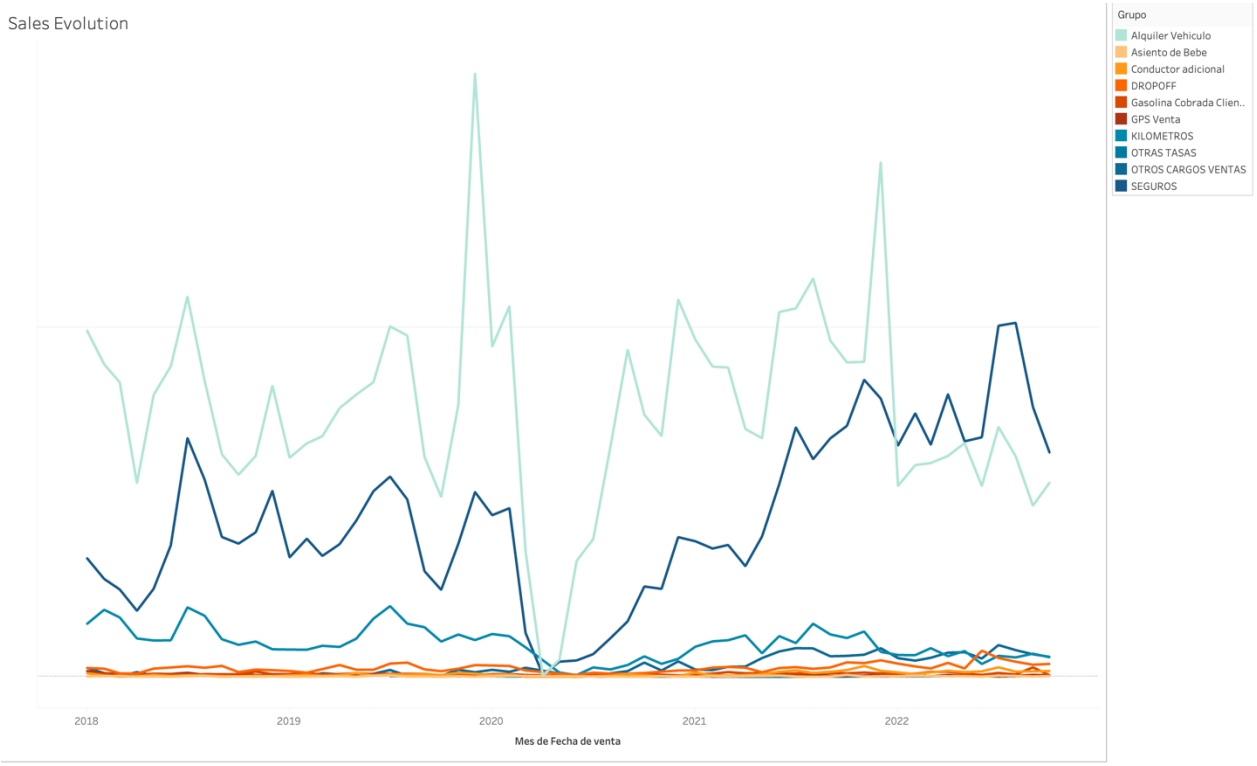
En la ilustración “Sales by región”, podemos observar que la región sierra genera aproximadamente el 65% de los ingresos a nivel nacional, y la región costa solo un 35%, esto debido a que solo en la ciudad de quito existen 3 agencias, mientras que en la ciudad de guayaquil existe una agencia.

La diferencia de precios y de agencias permiten también tener un mejor nivel de ingreso en cuanto a los drop offs y servicios adicionales, el incremento y mayor demanda en ciertos sectores también afectarían directamente el precio.

7.7.3. Evolución de las ventas.

Otro aspecto fundamental a topar en nuestro análisis descriptivo es la evolución no solo de las tarifas en la web, sino también de las ventas que tiene la empresa, esto debido a que la distribución a través del tiempo puede darnos pistas de porque es importante la temporalidad dentro de la predicción del precio.

Ilustración 4 Evolución de las ventas.



Fuente: Anexo 2 Sales Evolution

Elaborado por: Autor.

En la ilustración Sales evolutión, podemos observar cómo los ingresos de la compañía al igual que en la evolución de las tarifas tienen incrementos acordes a la temporalidad, el pico menos sustancial es obviamente en la pandemia del 2020, el grafico representa la evolución de los ingresos de todos los servicios que la compañía oferta.

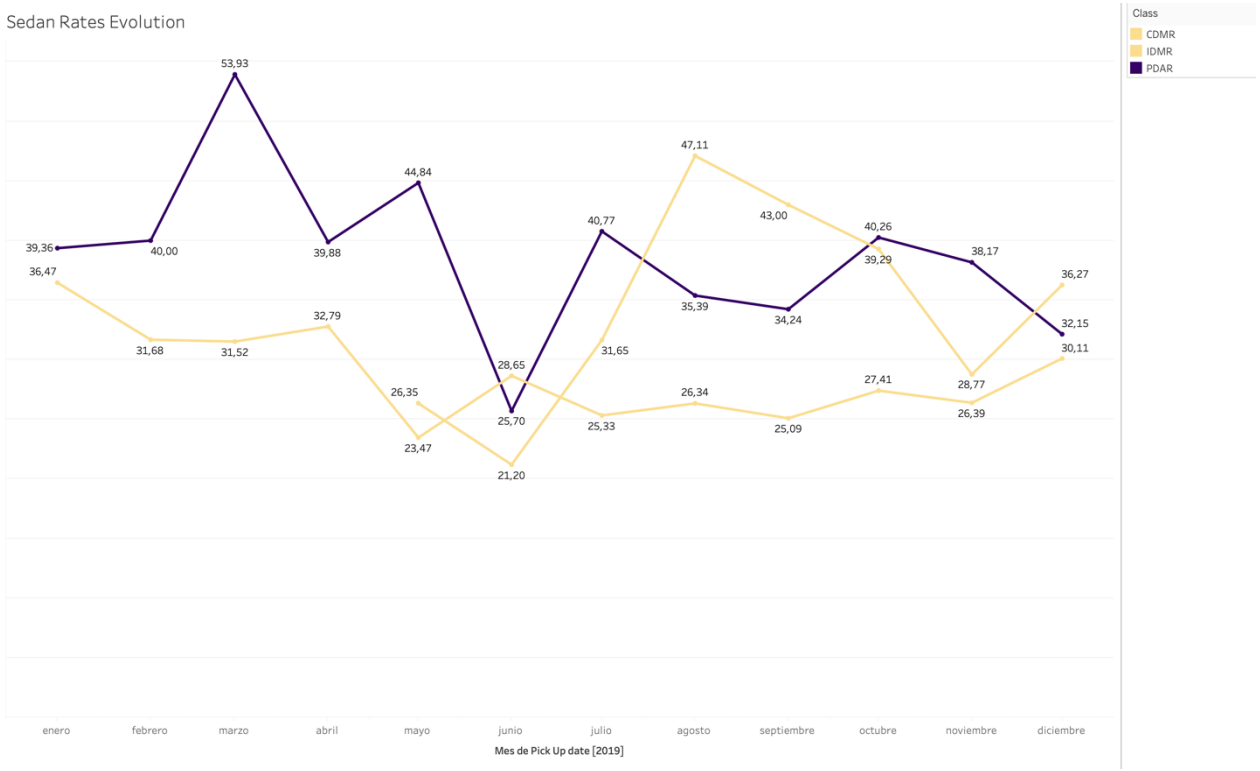
Los picos más altos se dan a finales e inicios de año, donde la venta de seguros está por debajo del valor de la renta de los vehículos siendo un valor considerable.

La temporalidad nos muestra que en los picos de alta demanda se incrementan los ingresos, sin embargo, esto no es concordante con la ilustración del precio de la renta, esta parece decaer en los meses de alta demanda, es el motivo por el cual, este estudio busca incrementar tales tarifas para tener un incremento en los ingresos.

7.7.4. Evolución de las tarifas de autos compactos 2019.

La evolución de las tarifas a través del tiempo es un análisis de que debe ser planteado para explicar la problemática de la variabilidad de los precios.

Ilustración 5 Sedan Rates Evolution



Fuente: Anexo 1, Rate Evolution

Elaborado por: Autor.

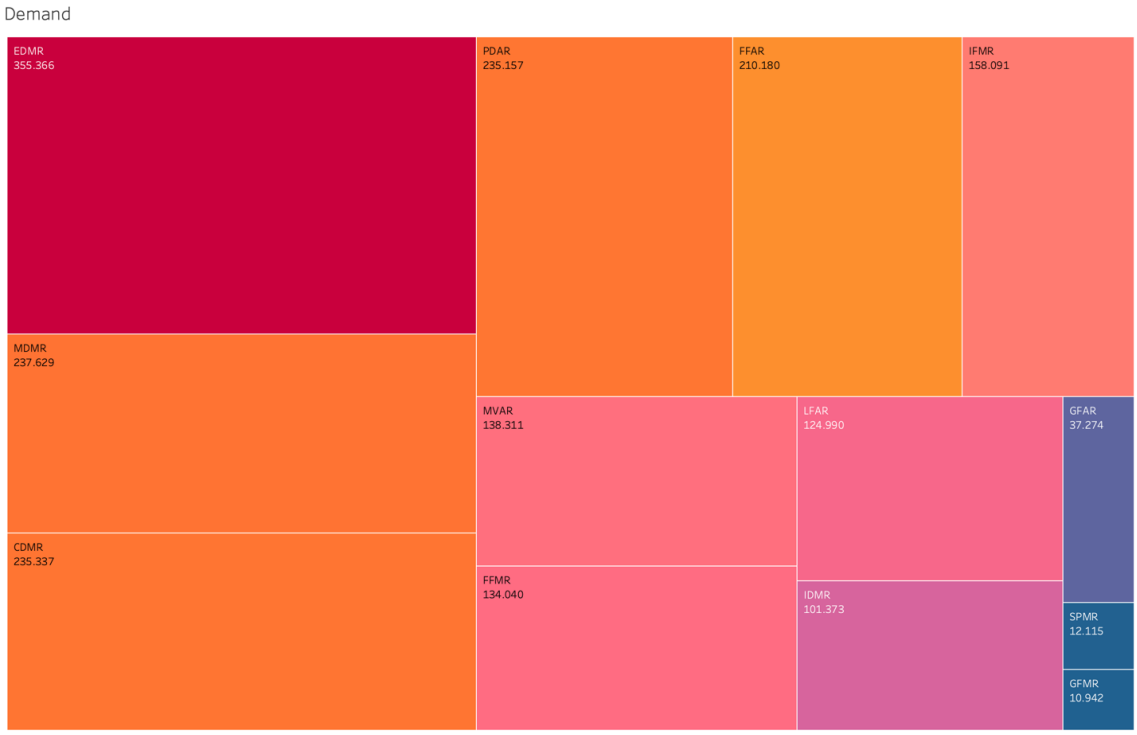
En la ilustración sobre la evolución de las tarifas de los autos sedan compactos, tomamos como referencia al año 2019 dado que a pesar de ser pre pandemia la

variabilidad de los precios era existente y sin concordancia con las temporadas de alta demanda, podemos observar que en el mes de junio, las tarifas de los autos están en los precios mínimos, siendo este mes el de mayor demanda a nivel anual, esto significaría que en estas temporadas lo que incremento el nivel de ingresos son en si solo fueron la venta de upsales, pero la cantidad de ingresos en el negocio principal que es la renta de autos se vio bastante mermada.

7.7.5. Demanda de autos.

En un principio de planteo la predicción de tarifas a través del precio promedio diario del número de rentas, sin embargo, algo que no se tomó en consideración fue los diferentes modelos de autos que existen en la flota.

Ilustración 6 Demanda de autos



Fuente: Anexo 1, Rate Evolution

Elaborado por: Autor.

En la ilustración sobre demanda, podemos observar la concentración de pedidos de los autos, estos destacan entre los compactos y los económicos, motivo por

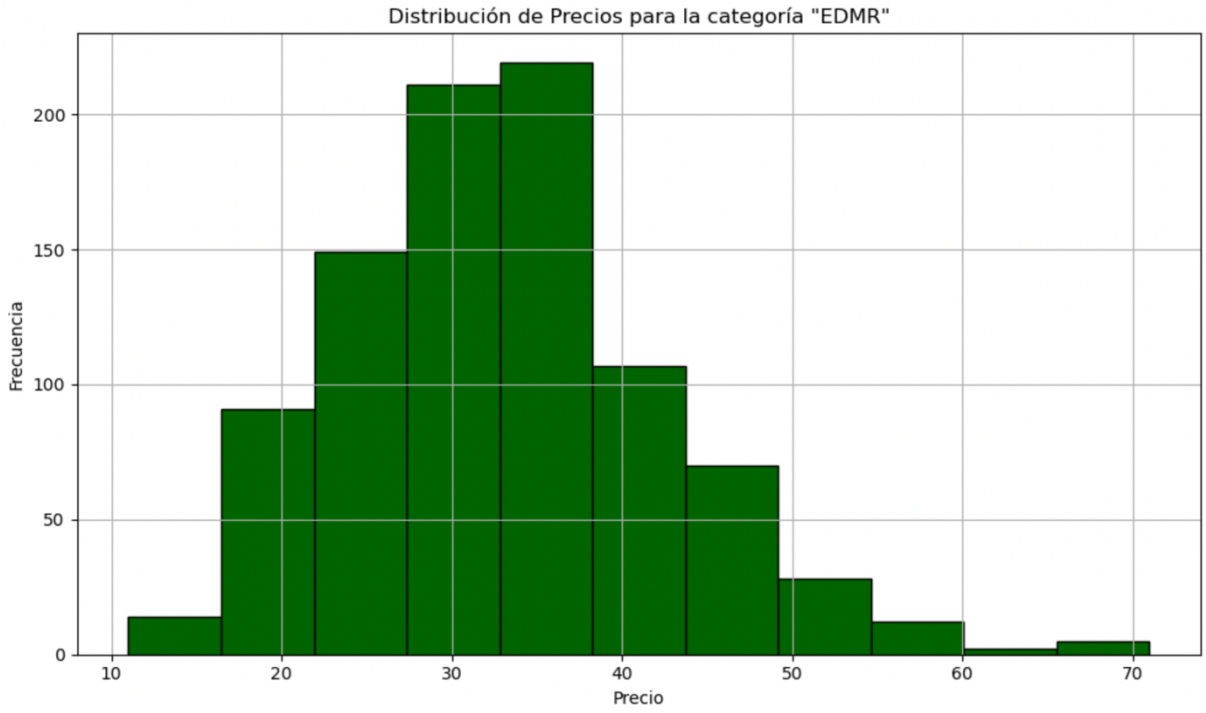
el cual haremos énfasis en estas dos categorías de autos, porque a su vez al ser más demandado existe mucha más data disponible para la generación del modelo de predicción, cosa que no sucede con las categorías de autos de lujo, por ejemplo.

7.7.6. Distribución Del Precio.

Para realizar la distribución del precio optaremos por utilizar Python, y la biblioteca, pandas y Matplotlib.

7.7.7. Distribución del precio autos “EDMR”

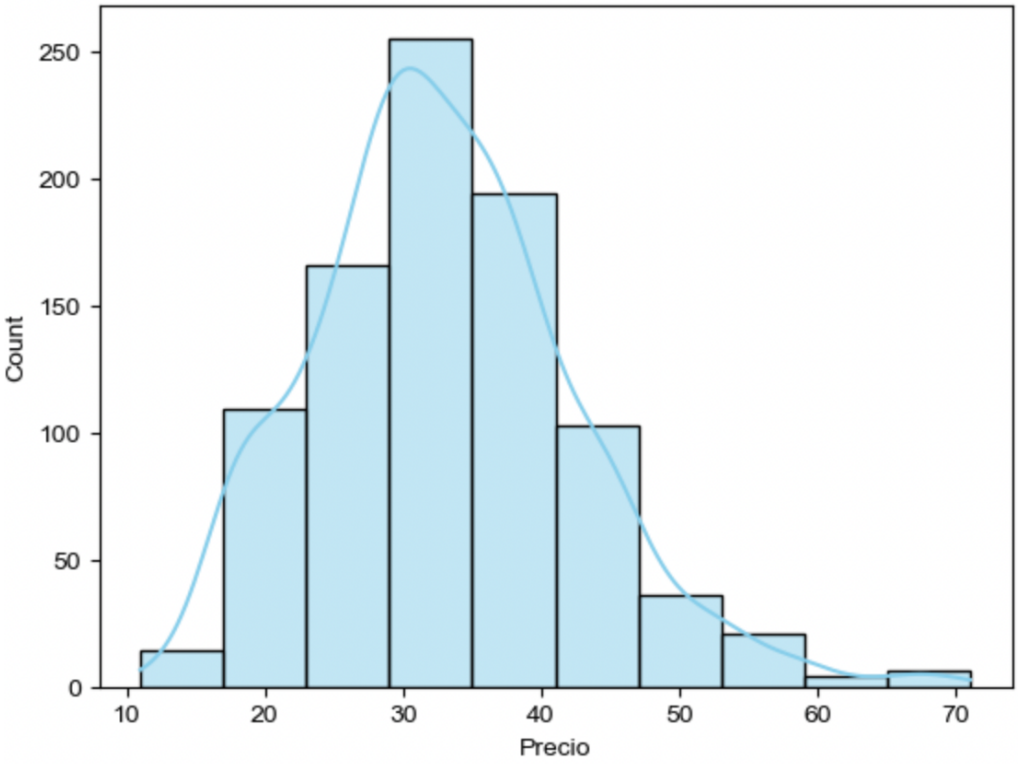
Ilustración 7 Distribución del precio.



Elaborado por: Autor.

Con la distribución del precio de la categoría de autos económicos, las tarifas van desde los 15 USD hasta los 70 USD como máximo, existe una gran concentración de datos en los 30 y 50 USD, por lo que son las tarifas más populares en los sitios web, es lógico debido a que estos precios son los que se ofrecen en counter directamente a los clientes que no tienen reserva.

Ilustración 8 Línea de tendencia del precio.



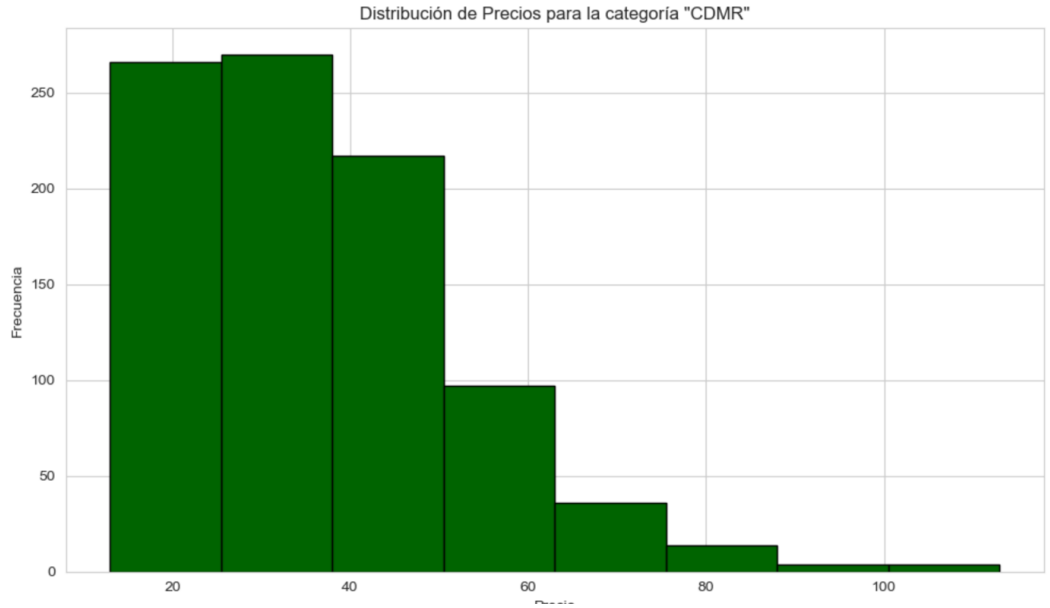
Elaborado por: Autor.

La distribución no parece tener una cola pesada sin embargo posee una asimetría positiva, entenderíamos que nuestra moda se ubicaría en la parte izquierda del precio, lo que significaría que existen más tarifas bajas que altas.

7.7.8. Distribución de CDMR.

Nuevamente utilizamos las librerías antes mencionadas, obtenemos los siguientes resultados:

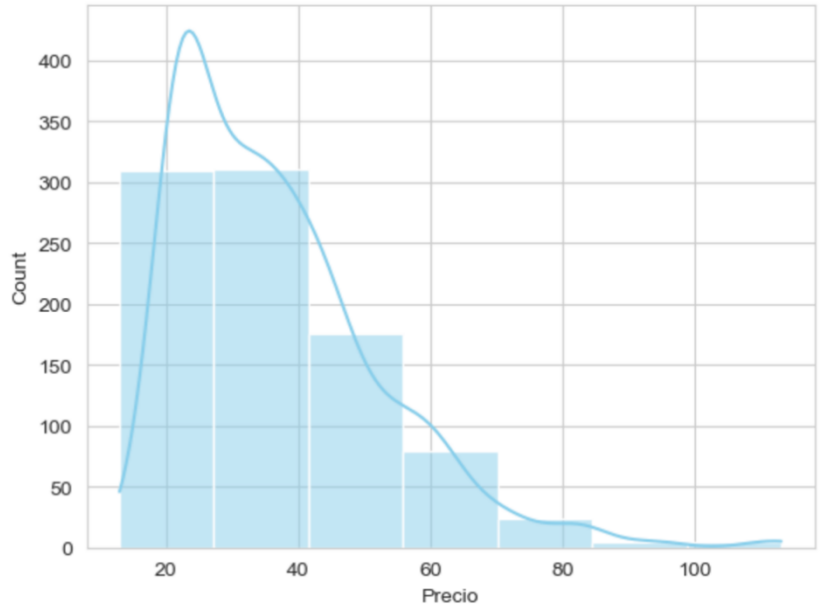
Ilustración 9 Distribución de autos CDMR



Elaborado por: Autor.

En nuestro gráfico de distribución de la categoría compactos podemos visualizar que existe una gran concentración de coches que tienen cerca de 60 a 70 USD diarios en el valor del auto, sin embargo, las tarifas pueden ir desde los 20 USD hasta los 100 USD.

Ilustración 10 línea de distribución CDMR



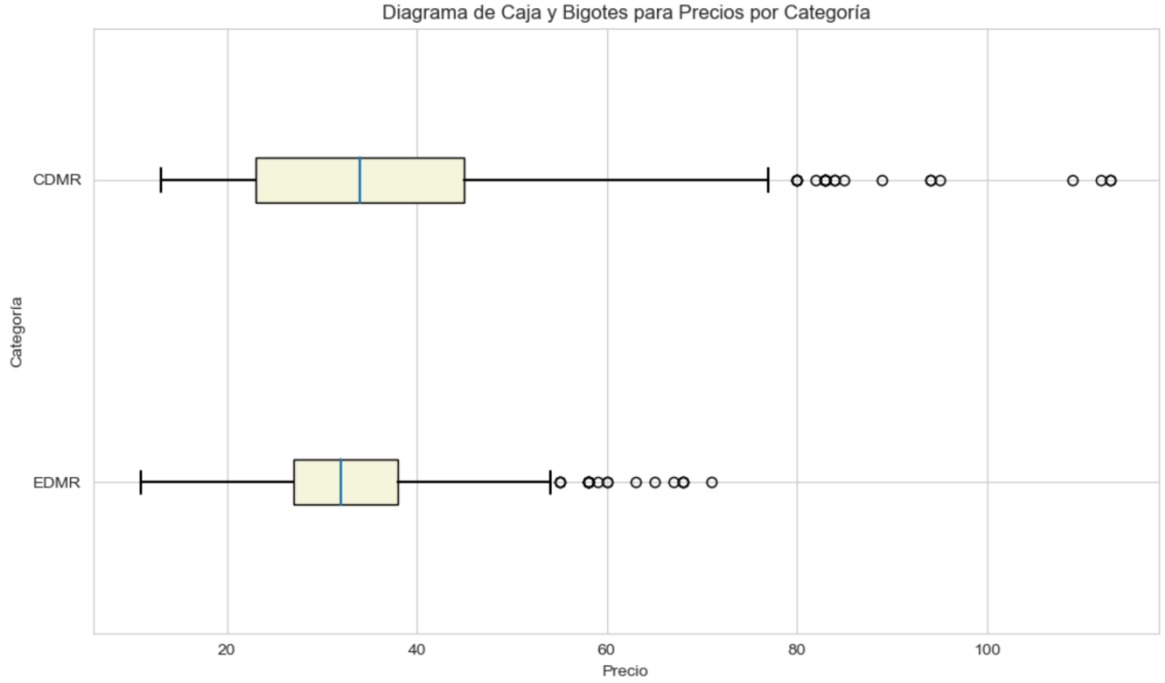
Elaborado por: Autor.

Adicionalmente nuestro gráfico, parece tener una cola pesada y una asimetría negativa, nuevamente las tarifas son más bajas que altas.

7.7.9. Diagrama de caja y bigotes, por categoría.

Finalmente, para entender de mejor manera la concentración de los precios y datos atípicos utilizaremos la herramienta de bigotes y cajas en con la librería Matplotlib y pandas, con el siguiente código obtener nuestro resultado.

Ilustración 11 Diagrama de caja y bigotes.



Elaborado por: Autor.

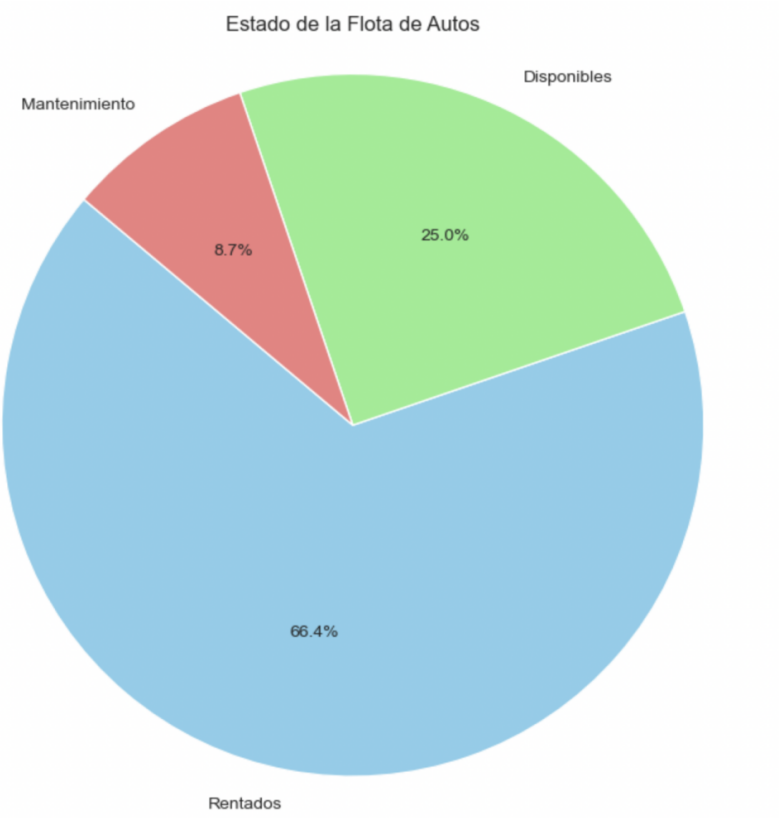
Identificamos que en la categoría económicos la concentración de precio se da en los 40 USD no existen datos fuera de los rangos intercuartiles, por lo que podemos afirmar que no existen datos atípicos, por otro lado en las categorías compacto la concentración del precio se da en los 60 USD y tenemos que algunos datos superan más de los 100 USD esto se debe a que algunas reservas son realizadas por los agentes en counter, estas tarifas suelen tener precios mucho más altos que los se encuentran en sitios web.

7.7.10. Utilización de la flota.

Otra parte fundamental de la problemática es incrementar la utilización de la flota esto se consigue con una excelente estrategia de precios que es lo que este estudio plantea, generando reservas cuando no exista mucha demanda e incrementando precios cuando sí.

Para describir la utilización de la flota utilizaremos las unidades en mantenimiento, rentadas y disponibles de estas dos categorías.

Ilustración 12 Utilización de la flota



Elaborado por: Autor.

En nuestro período de análisis solo el 66,4% se ha utilizado en la flota, de estas dos categorías de autos, teniendo un 25% de unidades disponibles y un 9% en mantenimiento, el objetivo sería incrementar el porcentaje de unidades rentadas.

7.8. Analíticos Descriptivos.

A través de Python optamos por utilizar describe, para visualizar lo descriptivos de nuestras variables numéricas, teniendo el siguiente resultado.

Ilustración 13 Descriptivos de la variable.

```
df.describe()
```

	Semana	Rentados	Disponibles	Mantenimiento	Totalflota	Reservadas	Canceladas	NoShow	Contratos	Precio
count	1816.000000	1816.000000	1816.000000	1816.000000	1816.000000	1816.000000	1816.0	1816.00000	1816.000000	1816.000000
mean	24.584802	19.708150	7.263767	2.234031	37.292952	1.205396	0.0	0.01707	3.160242	34.962004
std	14.979232	9.791471	4.867425	1.380609	15.736259	1.339460	0.0	0.12957	2.599891	13.085908
min	1.000000	3.000000	0.000000	0.000000	20.000000	0.000000	0.0	0.00000	0.000000	11.000000
25%	12.000000	13.000000	4.000000	1.000000	24.000000	0.000000	0.0	0.00000	1.000000	25.000000
50%	23.000000	18.000000	6.000000	2.000000	44.000000	1.000000	0.0	0.00000	3.000000	33.000000
75%	37.000000	22.000000	10.000000	3.000000	52.000000	2.000000	0.0	0.00000	4.000000	41.000000
max	53.000000	54.000000	27.000000	9.000000	83.000000	10.000000	0.0	1.00000	18.000000	113.000000

Elaborado por: Autor.

La media de precios es de 34 USD diarios, aproximadamente se abren 3 contratos diarios, tenemos un total de 1816 observaciones por variable, existe una gran dispersión de datos en cuando a las variables precio y total de la flota.

7.8.1. Gráfico de correlación.

Ilustración 14 Gráfico de correlación

```
df = df.drop("Canceladas", axis=1)
```

```
df.corr().style.background_gradient(cmap='coolwarm')
```

	Semana	Rentados	Disponibles	Mantenimiento	Totalflota	Reservadas	NoShow	Contratos	Precio
Semana	1.000000	-0.033035	0.021959	-0.073732	0.024743	0.003374	0.027216	0.002728	0.030172
Rentados	-0.033035	1.000000	-0.011574	0.191887	0.707267	0.473439	0.136385	0.664378	-0.085017
Disponibles	0.021959	-0.011574	1.000000	-0.013454	0.397401	0.066221	0.029549	0.153743	-0.050013
Mantenimiento	-0.073732	0.191887	-0.013454	1.000000	0.403365	0.110745	0.057735	0.138745	0.024920
Totalflota	0.024743	0.707267	0.397401	0.403365	1.000000	0.397936	0.106445	0.553916	0.017782
Reservadas	0.003374	0.473439	0.066221	0.110745	0.397936	1.000000	0.189310	0.670854	-0.096746
NoShow	0.027216	0.136385	0.029549	0.057735	0.106445	0.189310	1.000000	0.147253	-0.033412
Contratos	0.002728	0.664378	0.153743	0.138745	0.553916	0.670854	0.147253	1.000000	0.022867
Precio	0.030172	-0.085017	-0.050013	0.024920	0.017782	-0.096746	-0.033412	0.022867	1.000000

Elaborado por: Autor.

Determinamos una correlación con relación positiva fuerte entre las variables "Rentados" y "Totalflota" (0.707267), lo que sugiere que a medida que aumenta la cantidad de autos rentados, aumenta el tamaño total de la flota.

También podemos observar una correlación positiva regular entre las variables "Rentados" y "Contratos" (0.664378), esto es un indicativo que cuando tenemos más autos rentados, existen más contratos.

Determinamos la aparición de la correlación negativa moderada entre el precio y la cantidad de reservas (-0.096746), lo que sugiere que a medida que aumenta el precio, disminuyen las reservas.

Finalmente encontramos una correlación negativa moderada entre el precio y la cantidad de unidades rentadas (-0.085017), lo que sugiere que a medida que aumenta el precio, disminuye la cantidad de autos rentados.

7.9. Selección del modelo estadístico.

En este apartado definiremos que modelo estadístico se acopla de mejor manera a nuestro objetivo general, en base a nuestro marco teórico pudimos identificar

algunos modelos dentro los cuales se determinó que la regresión lineal y un random forest son los candidatos idóneos para la implicación de nuestro estudio.

Estos mismo fueron aplicados directamente sobre estudios de la variabilidad del precio de los boletos aéreos en aeropuertos internacionales (Montiel, 2018), por otro lado, la regresión lineal fue de echo la estrategia primordial en cuando se desea predecir el valor de las acciones de Apple (Broncano, 2022).

7.9.1. Modelo de Regresión

Los modelos de regresión se utilian tradicionalmente para determinar la influencia de las variables independientes sobre la variable dependiente (Lejarza, 2018), en neustro modelo de random forest el modelo de regresión se utiliza para tener un mayor acercamiento de nuestras variables predictora y su influencia sobre la variable que estamos modelando.

El modelo de regresión en un random forest utiliza el principio de los minimos cuadrados en cada arbol de desición, donde trata de determinar a la ecuación de regresión al minimizar la suma de los cuadrados de las distancias verticales entre los valores reales de Y y los valores pronosticados de Y (Lind, Marchal, & Wathen, 2008., pág. 477).

En el Random forest utilizamos el MSE quen es la sumatoria de los mínimoc cuadrados de cada arbol divididos por el numero total de observaciones, la formula es la siguiente:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y_i}\right)^2$$

donde:

- n es el número de observaciones.
- Yi son los valores reales.
- *y^i* son las predicciones del modelo.

El MSE a su vez nos da un claro referente de la calidad del modelo, entre mas bajo sea el mismo representa que la calidad del modelo no es buena y si es alto indica que el modelo tiene un buen comportamiento prediciendo los datos buscados.

7.9.2. Random forest.

El random forest es una herramienta muy poderosa la cual se utiliza tanto en problemas de regresión como de clasificación, esta herramienta es idónea debido a que nuestro estudio se enfoca en un análisis de regresión.

Esta herramienta es bastante útil considerando que utiliza varios árboles de decisión para realizar predicciones, recordemos que el árbol de deposición es una representación gráfica de todas las consecuencias o resultado posibles de la toma de una decisión (Lind, Marchal, & Wathen, 2008., pág. 764).

Adicionalmente en consideración de que también poseemos variables cualitativas, el random forest también permite trabajar con este tipo de variables, en las cuales tenemos la clase del auto, por ejemplo.

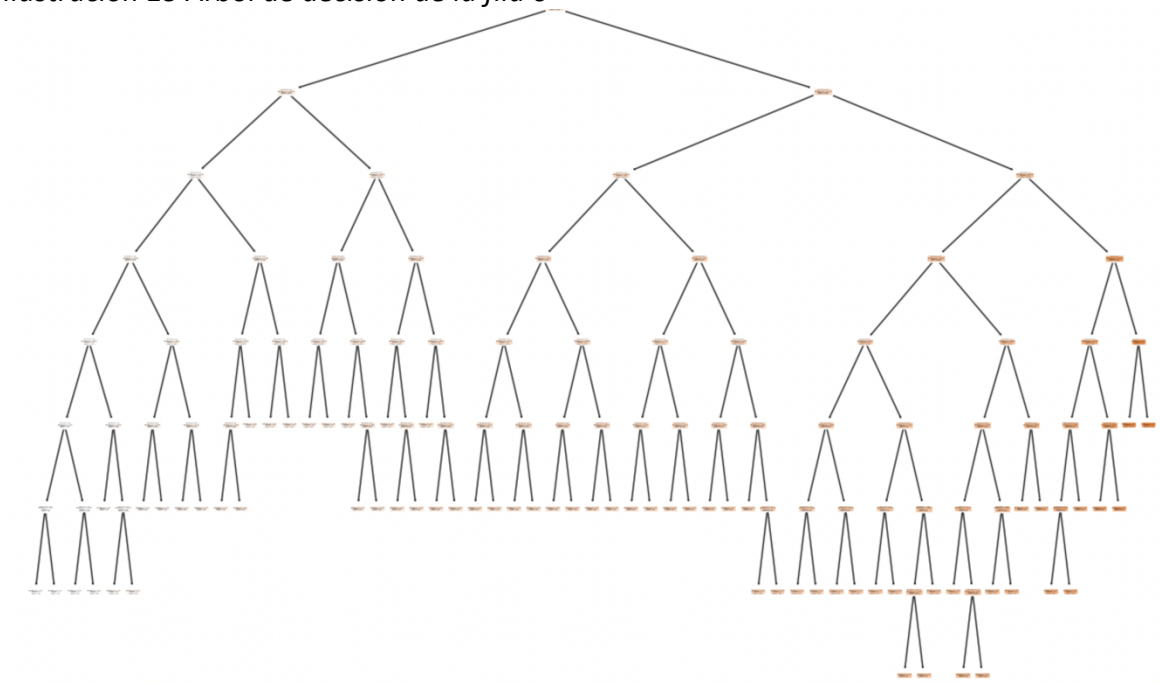
El random forest utiliza varios conjuntos de árboles de decisión al ser un modelo de aprendizaje supervisado el mismo utiliza cierta parte de nuestros datos para entrenamiento y otra parte de los mismos para prueba.

7.9.3. Árbol de decisión.

Dado que nuestro random forest es básicamente un conjunto de árboles de decisión utilizamos Python para graficar el primer árbol en este caso de la línea

0 para poder visualizar grosso modo el conjunto de decisiones que llegaran a un resultado.

Ilustración 15 Árbol de decisión de la fila 0



Fuente: (Europcar, 2022)

Elaborado por: Autor.

En este ejemplo, estamos tomando el primer árbol del bosque aleatorio (índice 0) y utilizando plot_tree de la biblioteca sklearn. tree para graficarlo.

Aquí podemos apreciar todas las decisiones posibles existentes en cuanto a las variables independientes con el fin de poder predecir el precio.

Este es tan solo un árbol de decisión que, si bien es excelente al momento de realizarse por individual, esto genera un sobreajuste en los datos.

Por este motivo se plantea la opción del random forest como medio de predicción de precios donde evitemos el sobreajuste, pero contemos con la precisión y la poca variabilidad de un árbol de decisiones, conformado nuestro resultado en base al resultado de varios árboles de decisión con diferentes muestras las

cuales a la larga otorgarán el peso a nuestro resultado, y aquellos árboles que supongan una alta variabilidad, tendrán poco peso en el resultado final.

8. RESULTADOS

8.1. Análisis del modelo estadístico.

El modelo a usar dentro de nuestro estudio es un análisis de regresión lineal a través de un random forest, el random forest o arboles aleatorios, nos permite encontrar la predicción optima del precio a través del conjunto de árboles de decisión que nos permitirán tener una excelente aproximación del precio para fechas futuras, considerando todas aquellas variables que incluimos dentro del estudio.

Parte importante del árbol aleatorio es que podemos utilizar variables cuantitativas que categorizaremos, estas variables son por ejemplo la clase, dado que trabajamos con dos clases diferentes de autos en este estudio, también las fechas como el día semana o mes, ya que en el análisis descriptivo se demostró la importancia de la temporalidad en la definición del precio.

Otro aspecto fundamental es que podemos medir la precisión y la variabilidad de nuestros datos utilizando métricas como el Error cuadrático medio y el coeficiente de determinación, esto nos permitirá saber que tan bueno es nuestro modelo al momento de predecir precios de fechas futuras con los datos que proporcionaremos para entrenamiento y para muestra.

Tradicionalmente los árboles de decisión tienen bastante precisión, pero mucha variabilidad, lo que solucionamos al crear un conjunto de árboles aleatorios con diferentes muestras que permitan tener un mejor ajuste de nuestro modelo (random forest).

Otro aspecto importante para considerar es la cantidad en la cual vamos a generalizar estos árboles, para esto determinaremos el número de estimadores a realizar, esto con el fin de optimizar recursos, lo que buscamos básicamente es saber el número de estimadores necesarios para tener el MSE más bajo posible, esto lo realizaremos a través de la determinación de hiperparametros en nuestro modelo.

8.2. Modelo de regresión a través de Random Forest.

Una de las ventajas del random forest es que posee todas las bondades de un árbol de decisión, pero al generar aleatoriamente varios árboles de decisión solventan el problema de sobreajuste que podría tener un modelo basado en un árbol de decisión el cual es sensible a cualquier cambio en el modelo de entrenamiento esta alta variación se soluciona con la implementación de los bosques aleatorios.

Para entrenar un modelo de Random Forest en Python, primero realizamos algunas transformaciones a los datos, como la codificación de variables categóricas y la selección de las características a utilizar en el modelo.

En nuestro caso con labelencoder en Python podemos codificar las variables, que serían class, día y mes, Label encoder automáticamente transforma a las variables en numéricas categóricas, en nuestro caso los features serán todas aquellas variables que describen el comportamiento de nuestra variable dependiente, y será precio.

Asignaremos a las variables temporales números en este caso la leyenda quedaría de la siguiente manera:

Tabla 3 Tabla de variables categorizadas.

Variable	Tipo	Numero
Class	EDMR	1
	CDMR	0
Dias	Lunes	1
	Martes	2
	Miercoles	3
	Jueves	4
	Viernes	5
	Sabado	6
	Domingo	7
Meses	Enero	1
	Febrero	2
	Marzo	3
	Abril	4
	Mayo	5
	Junio	6
	Julio	7
	Agosto	8
	Septiembre	9
	Octubre	10
	Noviembre	11
	Diciembre	12

Fuente: (Andagoya, 2023)

Elaborado por: Autor.

8.3. Interpretación de Resultados.

8.3.1. Entrenando el modelo con todo el Dataset.

En primera instancia entrenamos todo nuestro modelo con nuestro Dataset, esto significa que no optamos por dividir las muestras en conjuntos de datos de entrenamiento y otros de prueba, con esto buscamos indicadores como el coeficiente de determinación o a su vez el error cuadrático medio.

Usualmente utilizamos como muestra todo nuestro Dataset cuando tenemos una muestra significativamente pequeña, en ese sentido al poseer cerca de 1800 observaciones por columna optamos por encontrar estos dos indicadores e interpretar sus resultados.

Tabla 4 R2 y MSE con todo el Dataset

Indicador	Random Forest
R2	0,8912
MSE	18,607

Fuente: (Andagoya, 2023)

Elaborado por: Autor.

Bajo estos resultados entendemos primeramente que el R2 que es un indicador de que también está funcionando nuestro modelo a la hora de predecir datos, esto básicamente mide la variabilidad de nuestra variable precio y la precisión del modelo al tratar de explicar esa variabilidad, entre más se acerca a uno el modelo tiende a realizar mejores predicciones, ya que entre más se acerca explica de mejor manera la variabilidad de nuestros datos.

En este caso en particular nuestro resultado de R2 es 0,89 es casi el 90% lo que significaría que nuestro predice en un 90% la variabilidad de nuestros datos en cuanto al precio, siendo este un excelente indicador de nuestro modelo.

Por otro lado, el MSE[2] nos ayuda a determinar en qué medida nuestro modelo posee calidad, de predicción, este mide la variabilidad que existe entre la predicción planteada por nuestro modelo contra nuestros datos reales, lo que usamos para entrenar el modelo.

A través de la diferencia entre la predicción del precio y el precio real, se elevan al cuadrado esto con el objetivo de eliminar signos con símbolo negativo los cuales pueden afectar nuestro modelo, y finalmente divide entre el total de observaciones realizando así un promedio de estas, en nuestro caso en

[2] Mean Square Error

específico seria la diferencia entre las observaciones del precio predichas y las observaciones del precio real de la data, elevadas al cuadrado y promediadas.

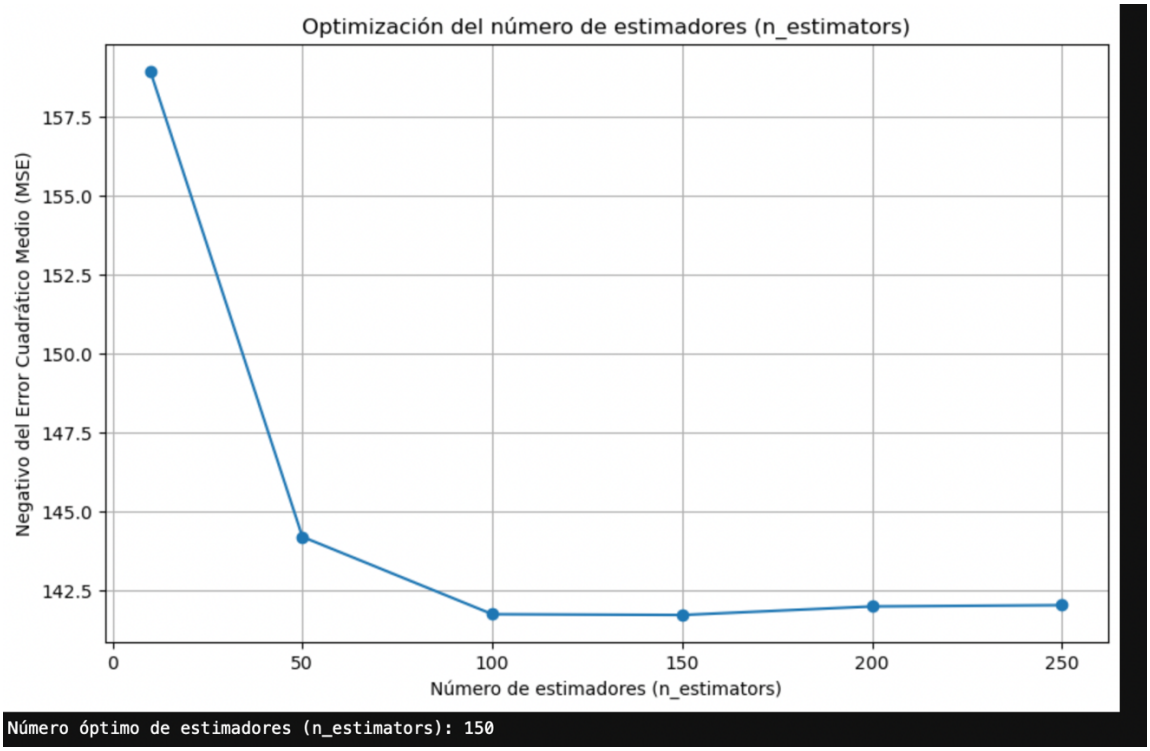
Inversamente proporcional al R2 nuestro MSE entre más bajo sea mejor indicativo será de nuestro modelo ya que indicaría que la variabilidad es menor o más bien que están desviadas por 18,6 unidades promedio elevadas al cuadrado, en este caso en específico para nuestro modelado de precios es un resultado aceptable, pero no optimo considerando el precio de los autos y como se vio en el análisis exploratorio del precio promedio de los EDMR y CDMR cuyo valor más bajo era 20 y el más alto era 100 USD.

8.4. Modelo de random forest con datos de prueba y entrenamiento.

A pesar de que nuestro modelo existente en base al uso de todo el Dataset, nos proporcionó buenos indicadores, este modelo puede presentar un sobreajuste, esto debido a que al no tener datos de entrenamiento el modelo conoce de antemano el 100% del Dataset, esto podría implicar que el modelo esta sobre ajustado y que explica demasiado bien las predicciones y su variabilidad debido a que conoce el Dataset entero, por este motivo optaremos por utilizar un 20 % de la data y el 80% restante para prueba, también plantearemos algunos indicadores importantes para evaluar nuestro modelo.

8.4.1. Total, de estimadores.

Ilustración 16 Optimización del total de estimadores.



Fuente: (Andagoya, 2023)

Elaborado por: Autor.

Antes de iniciar nuestro análisis de Random Forest, al ser este un conjunto de árboles aleatorios, nuestro total de estimadores son en escancia la cantidad de árboles que debemos usar para que nuestro modelo este compuesto y sea integro.

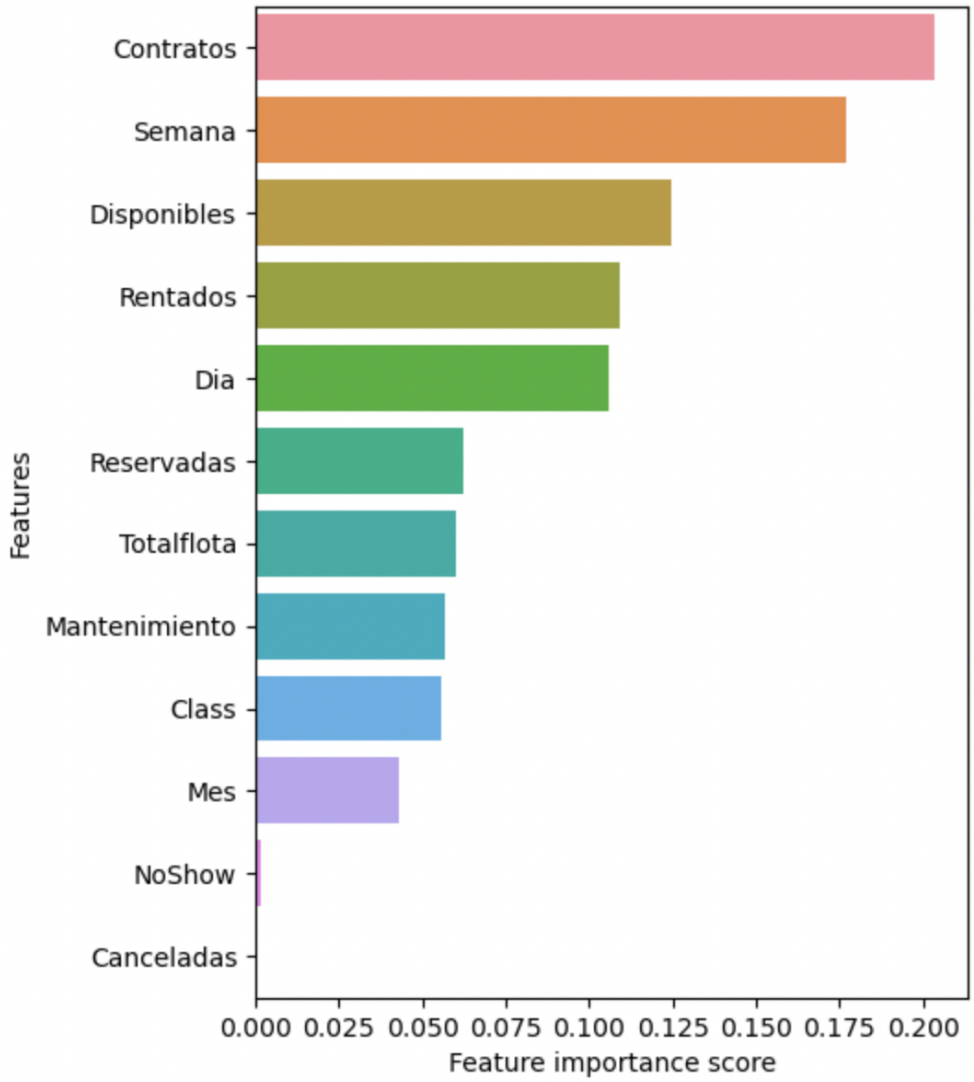
Usualmente se pensaría que, a un mayor número de estimadores, nuestro modelo estaría ensamblado de mejor manera, sin embargo, existen un numero de árboles desde a partir de los cuales el incremento en la cantidad no mejor significativamente el modelo, esto es porque su aleatoriedad no aporta ya información significativa a nuestro modelo.

En este caso según el grafico del total de estimadores, podemos observar que la cantidad de árboles optima debe ser 150, ya que a partir de aquí la información que aportarían más arboles no sería de gran ayuda.

8.4.2. Feature importance

Ahora bien, a través del uso del total de nuestro Dataset algo relevante a tomar en cuenta es la cantidad de relevancia que tienen nuestras variables dentro de nuestro modelo de predicción, para esto graficamos el future importance con el siguiente resultado.

Ilustración 17 Feature importance.



Fuente: (Andagoya, 2023)

Elaborado por: Autor.

En la ilustración anterior podemos determinar que la variable que tiene más peso o influencia en nuestro modelo son la cantidad de contratos, seguido de la semana en la que se encuentre definido el precio, esto es lógico debido a que en el marco teórico definimos que la temporalidad es importante al momento de incrementar o disminuir el precio, de igual manera a mayor unidades rentadas es lógico que el precio suba porque escasea la oferta, sin embargo aquella que tiene más peso es la cantidad de contratos, para un mejor entendimiento de la variables importantes se explica una tabla a continuación.

Tabla 5 Feature Importance

Variable	Peso
Contratos	0.203460
Semana	0.176833
Disponibles	0.124656
Rentados	0.109277
Dia	0.105997
Reservadas	0.062280
Totalflota	0.060289
Mantenimiento	0.056688
Class	0.055825
Mes	0.043208
NoShow	0.001486
Canceladas	0.000000

Fuente: (Andagoya, 2023)

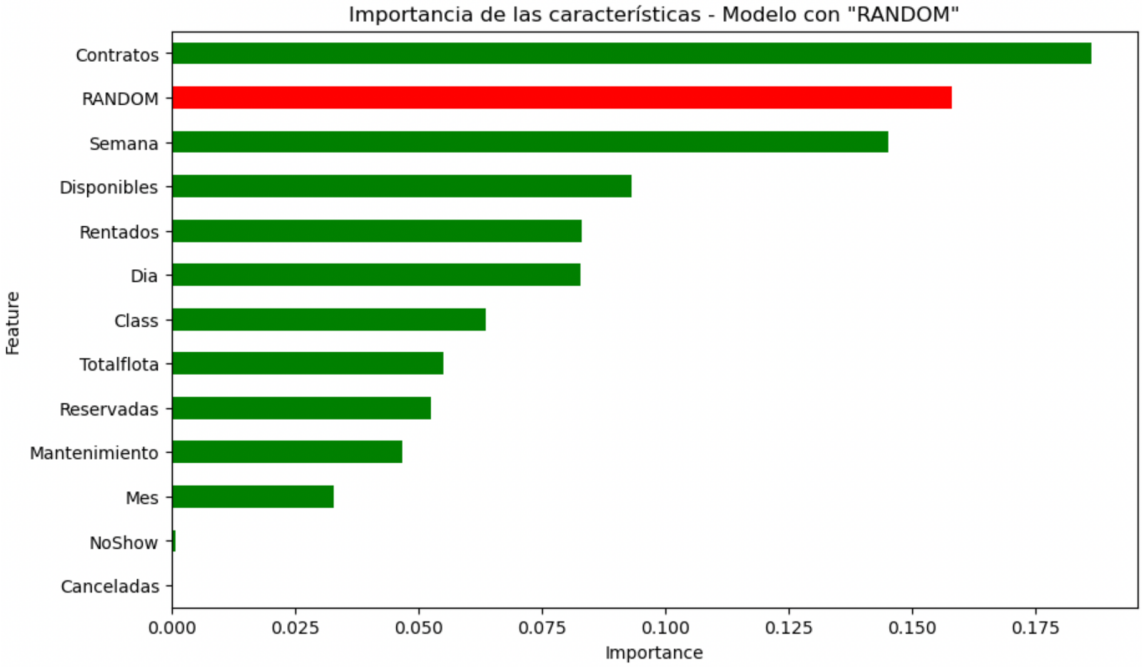
Elaborado por: Autor.

En la tabla podemos aprecia que el valor máximo es 0,2 el cual corresponde a la variable contratos, y el mínimo es 0 el cual corresponde a las reservas canceladas, podemos inferir que canceladas no tienen ningún peso debido a que las reservas canceladas son inexistentes.

8.4.3. Feature importance con una función aleatoria.

A pesar de determinar cuál es el peso de las variables importantes dentro del modelo, al trabajar en arboles aleatorios es probable que alguna de estas características radique su importancia en la aleatoriedad, para esto añadimos una característica aleatoria, la cual nos permitiría diferenciar entre las características realmente importantes y aquellas cuya importancia radica en la aleatoriedad.

Ilustración 18 Feature importance con una característica aleatoria



Fuente: (Andagoya, 2023)

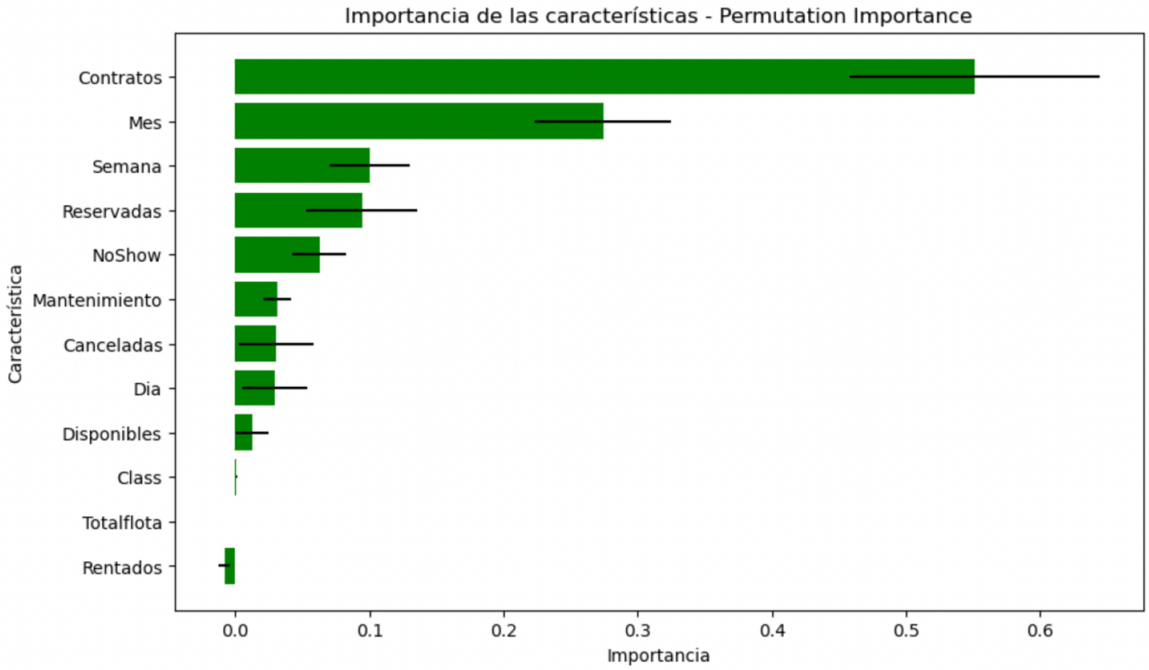
Elaborado por: Autor.

En el grafico con la característica aleatoria incluida podemos concluir que solo la variable contrato tiene su peso en importancia realmente y las demás variables que se encuentran debajo de la característica Random, deben parte de su importancia a la aleatoriedad.

8.4.4. Feature importance a través de la permutación.

Una permutación es básicamente la cantidad de veces que podemos combinar varios elementos de un conjunto de una única manera cada vez, en este caso lo que hará en nuestro modelo es disminuir drásticamente los valores de todas nuestras características esto de manera aleatoria y en base a este cambio medir el rendimiento del modelo y la importancia de sus variables con el cambio.

Ilustración 19 Features importance con permutación



Fuente: (Andagoya, 2023)

Elaborado por: Autor.

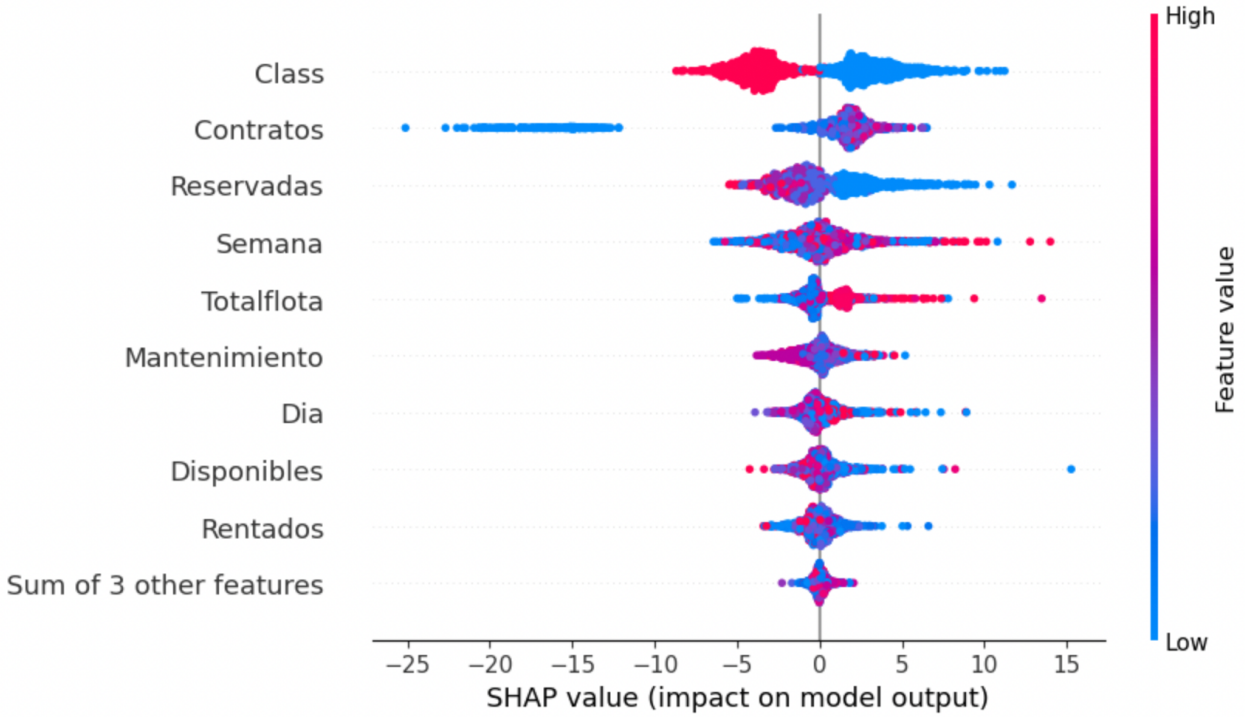
En nuestro grafico podemos determinar que nuevamente la cantidad de contratos es nuestra variable influyente o con más peso en el modelo y la línea que interseca la barra nos da un concepto gráfico de la variabilidad, que en este caso la desviación estándar más alta se da en el día y la cantidad de reservas

canceladas, esto podría sugerir que a pesar de ser variables no tan influyentes poseen un alto nivel de variabilidad.

8.4.5. Feature importance a través del método Shap.

El método SHAP[3] es básicamente un modelo que utiliza la teoría de juegos (Contreras, 2022) es útil dentro del aprendizaje supervisado, básicamente nos ayuda a comprender de mejor manera como nuestras características de entrada (variables independientes) contribuyen en cierta medida a nuestro modelo, utiliza la teoría de juegos haciendo la suposición que nuestras características del modelo son jugadores y en base a esto cada jugador aporta de diferente manera al juego.

Ilustración 20 Feature importance con el modelo SHAP



Fuente: (Andagoya, 2023)

Elaborado por: Autor.

[3] SHapley Additive exPlanations **Fuente especificada no válida.**

Para entender de mejor manera el gráfico, las características se encuentran dentro del eje Y los puntos están distribuidos del centro a la izquierda y a la derecha, cuando los puntos caen en la izquierda significa que la variable impacta de una manera negativa a nuestro modelo y cuando caen en la derecha es todo lo contrario, el “coolwarm” de los puntos representa la importancia de nuestra variable, entre más rosado la característica es más importante.

En este estudio, observamos que la clase posee un gran impacto negativo en nuestro modelo de predicción, los puntos están divididos uniformemente esto podría deberse a que tenemos dos clases de autos una podría impactar de manera negativa y con gran importancia y otra de manera positiva, pero con poca importancia.

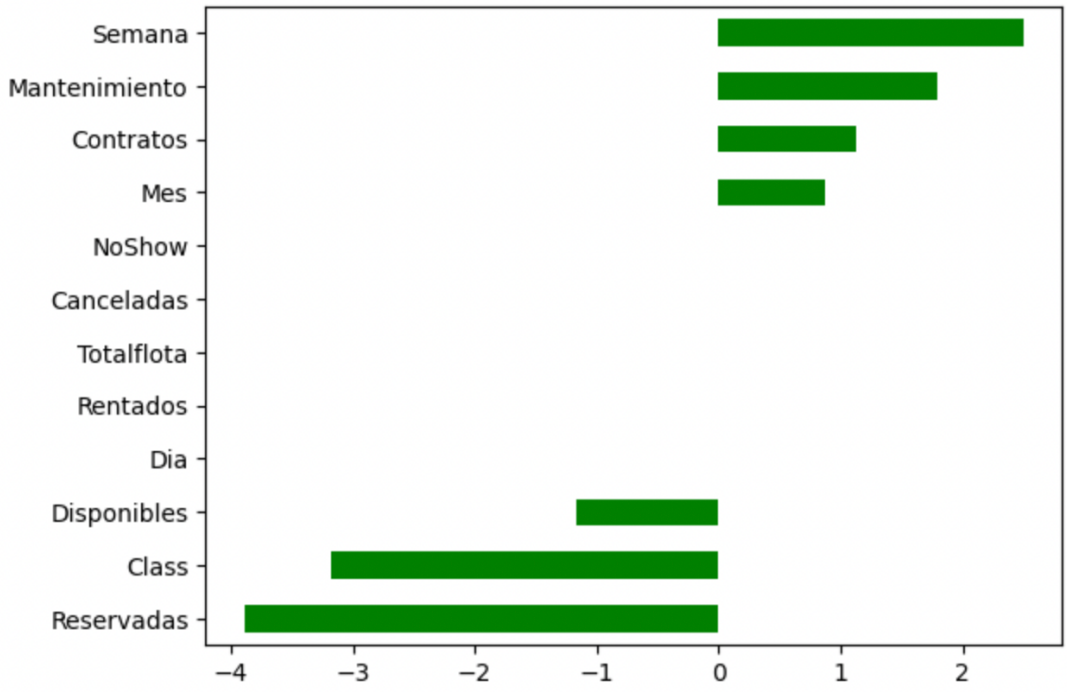
También observamos que la cantidad de contratos impacta de gran manera y positiva a nuestro modelo de predicciones, al igual que el total de la flota, que no tiene una gran concentración de observaciones en el lado derecho sin embargo estas son de gran importancia positiva.

8.4.6. Feature importance a través de la ruta del árbol.

Como se sabe el random forest es un conjunto de árboles aleatorios, cada árbol posee una ruta especifica que lleva a un determinado resultado, el conjunto de estos nos permite tener una predicción de varios árboles en conjunto.

A través de la ruta de cada árbol podemos definir de igual manera la importancia de nuestras características, a través de nuestros datos de entrenamiento

Ilustración 21 Feature importance a través de la ruta del árbol.



Fuente: (Andagoya, 2023)

Elaborado por: Autor.

A través de este gráfico podemos determinar la cantidad de contribución que existe en cada Feature característica, a través de la diferencia entre el valor que predice el modelo y el valor de cada nodo o hoja en nuestro modelo de random forest.

El grafico nos muestra que el número de unidades reservadas afecta de manera negativa a las predicciones de nuestro modelo, y por otro lado el numero de la semana afecta de manera positiva a las predicciones de nuestro modelo.

8.4.7. MSE y R2 con el modelo entrenado.

Una vez determinada la importancia de nuestras características, así como el número de árboles a usar y como impactan las características dentro de nuestro modelo de predicción de precios, se presenta estos dos indicadores en base al

modelo entrenado, considerando que usando toda la data el modelo podría estar sobre ajustado, los resultados son los siguientes.

Tabla 6 MSE y R2 con el modelo entrenado

Indicador	Random Forest
MSE	139,9047991
R2	0,121483549

Fuente: (Andagoya, 2023)

Elaborado por: Autor.

Con nuestro modelo entrenado tenemos un R2 de 0,12 lo que indicaría que el rendimiento del modelo al momento de medir la variabilidad de las predicciones es tan solo de un 12% lo cual es extremadamente bajo, por otro lado, el MSE es de 139 lo que indica una alta variabilidad entre las predicciones y los datos reales, esto indicaría un pobre desempeño de nuestro modelo de predicción de precios lo que tendría inferencias relevantes sobre nuestros objetivos.

9. DISCUSIÓN DE LOS RESULTADOS Y PROPUESTA DE SOLUCIÓN

9.1. Implicaciones para la organización.

En base a los resultados expuestos el modelo de determinación de precios parece ser inviable a pesar de recopilar información relevante de la influencia de variables externas para la codificación del precio, las cuales han sido verificadas en estudios posteriores de su influencia, parece ser que el modelo de precios no predice con éxito las tarifas dinámicas.

9.2. Implicaciones comerciales.

La principal implicación comercial seria la inexactitud de la definición del precio, el modelo determino que existen variables como la clase y la cantidad de contratos abiertos que influyen directamente del precio, sin embargo, como se planteó en el análisis exploratorio, el precio ha sido definido por la compañía de manera errónea y aleatoria, basándose en parámetros inexistentes o a su vez definiendo el precio de manera reactiva cuando los cambios ya debían haber sucedido.

Esta aleatoriedad indicaría que el modelo no explica la variable precio porque esta a su vez nunca fue definida en la parte gerencial en base a variables relevantes como la temporalidad o la oferta y demanda de autos, esto significaría que el precio que se define no tiene ninguna relación con estas variables por lo que erróneamente fue definido diariamente sin tomar en consideración factores externos importantes en el incremento de la utilidad.

9.3. Implicaciones financieras.

Otro problema fundamental surge en la implicación financiera de la mala definición del precio, y la dependencia del talento humano, si bien en nuestro análisis exploratorio definimos que el 45% de los ingresos son de la venta de seguros, mas no de la renta de autos esto significaría que al no tener una buena definición de precio, el giro de negocio principal de la empresa pasaría a ser secundario, impidiéndole desarrollar su máximo potencial financiero.

1.1. Estrategias.

9.3.1. Estrategia de definición del precio.

La principal estrategia en cuanto a la definición del dinamyc pricing es utilizar a las características obtenidas en el estudio, como base para la definición de nuestros precios desde hoy en adelante, el problema fundamental con la

creación del modelo es que el precio no tenía relación con las variables influyentes debido a que el precio está mal definido por la parte gerencial.

Una vez que tomemos como referencias las variables como temporalidad, demanda y oferta para definir el precio que son aquellas con un peso significativo y positivo en la predicción del precio, podríamos entrenar de mejor manera nuestro modelo de predicción permitiéndonos acércanos al dinamyc pricing que utilizan las aerolíneas, en primera instancia se buscaría incrementar la participación del servicio de renta de autos en cuanto a su ocupación en el nivel de ventas de la empresa.

9.3.2. Definición de precios ancla en counter.

Otra estrategia fundamental que nos permitió este estudio es plantear la creación del Dynamic pricing no solo para los bróker de internet si no también definir que gran parte de los precios definidos en reserva son balk ups (cliente sin reserva) esto permite incrementar el ingreso de la renta debido a que el precio definido en counter es superior dado que son reservas de último minuto, el Dynamic pricing no solo se limita a bróker sino que puede ser una herramienta fundamental para los agentes de venta.

9.3.3. Aprovechamiento de la flota.

El modelo también determino que las variables que incluyen el aprovechamiento de la flota, tales como las unidades en mantenimiento, o el total de la flota, a través de nuestro modelo en SHAP vimos que estas variables tienen un impacto positivo y significativo dentro de la predicción de precios, esto permite de alguna manera incrementar el aprovechamiento de la flota, en este proyecto se hizo hincapié en dos categorías de vehículos que demostró que el aprovechamiento de la flota tiene un impacto significativo sobre el precio, esto a su vez permite que la empresa tome mejores decisiones sobre el tipo de unidades que incorporan en la flota, y su desempeño en la predicción del precio de las mismas.

9.3.4. Aprovechamiento de la data futura.

El definir las variables importantes nos permite tomar acción en curso sobre la recopilación de nuevos datos que en el futuro permitirá entrenar de mejor manera nuestro modelo de predicción de precios, esta recopilación podría incluir datos ajenos al control de la compañía, como por ejemplo la definición del precio de la competencia, la estrategia es recopilas nueva data basada en las variables mencionadas.

9.4. Innovación y desarrollo.

SI bien el modelo de precios actual parece no predecir correctamente precios futuros, esto se debe mayoritariamente a que nunca ha existido una politica correcta de precios, consecuentemente aunque existan variables que definan el precio, al no estar este bien definido el modelo no podra predecir de manera correcta nuestra variable dependiente.

Sin embargo dado que todas estas tarifas son publicas, y dada la limitación que tenemos de la data la propuesta de innovación que se plantea es utilizar las tarifas de los brokers como punto de partida, ya que estas tarifas son dinamicas y son tomadas como referencia para la competencia.

9.4.1. Web Scraping como medio de obtención de información.

El internet está repleto de información que se genera constantemente, creamos y compartimos información cada segundo con diferentes finalidades, muchas de estas comerciales (Subirats, 2019).

Sin embargo, mucha de esta información no se encuentra a manera de datos estructurados, de echo más del 80% de los datos que podemos encontrar en el internet, son datos no estructurados (Barbosa, 2022), y en nuestro caso la información no la encontramos en tablas que podamos descargar fácilmente como nuestra anterior herramienta TSD.

Toda la información de brokers se encuentra representada a través de gráficos a los cuales se les asignan atributos como precio, cilindraje del motor o inclusive

aire acondicionado. Encontramos pues entonces información valiosa de la competencia como Álamo, quien fue aquellos fundadores del Dynamic pricing en la industria turística.

Tener como referencia estas tarifas significaría tener información confiable de los demás competidores para definir de mejor manera el precio, sin embargo, como se mencionó anteriormente esta información no se encuentra de manera estructurada, para lo cual nos valdremos de una herramienta denominada web scraping

La definición literal del web scraping en español sería algo como raspado web (Subirats, 2019), básicamente es una herramienta de programación que puede desarrollarse en diversos lenguajes de programación, lo que hace es que a través de código se obtiene información específica de diferentes sitios web de manera automática a datos estructurados.

Por ejemplo, podríamos generar el código para obtener el precio, lo modelos y la compañía que oferta el servicio de manera automatizada y en tiempo real, esto a su vez limitaría mucho el enfoque reactivo que tienen dentro de la empresa ya que tradicionalmente se recopilan los precios de manera manual, traspasándolo a una hoja de Excel y toda esta información se pierde.

Sin embargo, con esta herramienta podríamos definir el precio promedio de las diferentes categorías y adaptarlos a los datos existentes en nuestra base sobre las variables influyentes.

Esto ayudaría enormemente a la empresa sobre todo en cuanto a la limitación de información de los vehículos que no son tan demandados y de igual manera permite tener una clara referencia de la competencia limitándonos no tan solo a. nuestra información sino también a la de la competencia.

9.4.2. Limitaciones

Parte de las limitaciones está el hecho de que este proceso automatizado depende de un grupo consolidado de backend que pueda generar el código para automatizar el proceso, en primera instancia la única limitación seria económica,

posteriormente se necesitaría un equipo de front end para que la generación de la información sea digerible y entendible al momento de definir precios.

Por otra parte, otra limitación sería la privacidad de la información si bien toda esta información es publica, es probable que la competenciaia tenga rigurosos estandares sobre el tratamiento de su información en redes, ya que esta no se utilizaria con fines de compora y venta sino mas bien con fines empresariales para la toma de desiciones.

10. CONCLUSIONES.

Nuestro modelo define a las variables contratos, clase de auto y a su vez la temporalidad como aquella que influencia de manera positiva y fuerte dentro de la creación de nuestro modelo de predicción de precios a través de un random forest, con esto cumplimos nuestro primer objetivo específico el cual se basaba en definir la relevancia de las características dentro de la definición de precio.

Nuestro modelo determina u contextualiza los fundamentos del dinamyc pricing poniendo en evidencia las variables importantes en la determinación del precio, esto con los datos de dos modelos de autos en específico lo cuales son altamente demandados.

A través del estudio de investigación se determina como herramienta fundamental el Random forest como modelo de regresión para predicción de precios, ya que este nos permite determinar el precio a través de un conjunto de árboles aleatorios cuyos resultados le dan más peso a las predicciones importantes, al mismo tiempo el modelo posee la bondad de tener una baja variabilidad y evita el sobreajuste de los árboles de decisión individuales y la variabilidad de los mismos frente a los mínimos cambios.

11. RECOMENDACIONES.

Se recomienda a la empresa cambiar el método de definición del precio, ya que en cierta medida la definición actual del mismo funciona de manera retrograda, u entorpece la creación de modelos de predicción de precios más efectivos tomando en cuenta que la definición actual es reactiva mas no proactiva, esto distorsiona el modelo debido a que los datos de precio que se selecciona en una fecha en especifica corresponden a valores que cambian constantemente y que erróneamente se definen una vez que el precio ha cambiado.

Se recomienda por otro lado basar la definición del precio en las características antes mencionadas, mas no en tan solo la tarifa mínima que es como se actúa hoy en día definiendo el precio en base a el mínimo de la competencia para ser competitivos en precios, esto limita la capacidad de la empresa de incrementar sus ingresos y sujeta a la empresa a la dependencia de venta de productos complementarios a la renta de vehículos.

Se recomienda de igual manera destinar fondos para la creación del algoritmo de machine learning a traves del web scraping, con el fin de poder obtener información fidedigna para la definición del precio.

12. REFERENCIAS

- Marshall, A. (1931). *Principios de Economía.* París: El consultor bibliográfico.
- Garzon, P., Lozada, D., & Monroy., L. (2018). *Methodology for pricing using price elasticity of demand. Case type: automotive parts sector.* Bogota.: CENES.
- Collin, J. G. (2015). *Contabilidad de Costos.* Ciudad de México: Mc Graw Hill.
- Kotler, P., & Keller, K. (2016). *Dirección de Marketing.* Atlacomulco: Pearson.
- Poundstone, W. (1995). *Priceless, The Myth of Fair Value (and How to Take Advantage of It).* New York: Hill and Wang.
- Neagle, T., & Muller., G. (2018). *THE STRATEGY AND TACTICS OF PRICING.* New York: Routledge.
- Montiel, M. (2018). *Estudio de factores que influyen en la compra de boletos de avion por internet.* Monterrey: TEC de Monterrey.
- Tomas, F., & Vega., M. (2020). *ESTRATEGIAS DE FIJACIÓN DE PRECIOS EN EL TRANSPORTE AÉREO DE PASAJEROS.* Sevilla.
- Broncano, R. (2022). *Application of machine learning techniques for apple stock price prediction.* Lima: Revista de investigación de sistemas e informática.
- Medina, A. (2020). *Predicción de precios de vivienda en la ciudad de Medellin.* Medellin: UNIR.
- Porter, M. (1980). *ompetitive Strategy: Techniques for Analyzing Industries and Competitors.* New York: Simon & Schuster.
- Castro, M. (2009). ACERCA DE LAS ECONOMÍAS DE ESCALA, EL TAMAÑO Y LA LOCALIZACIÓN DE INVERSIONES. *Redalyc*, 30.

- Europcar. (2022). *Europcar.* Obtenido de Europcar.com: https://www.europcar.es/EBE/module/render/Nuestra-Historia
- Oppel, A., & Sheldon, R. (2009). *Fundamentos de SQL.* Santa Fe: Mc Graw Hill.
- Sandoval, L. J. (2018). ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS. *REVISTA TECNOLÓGICA ITCA*, 40.
- Lind, D. A., Marchal, W. G., & Wathen, S. A. (2008.). Estadistica aplicada a los negocios y economía. *Mc Graw Hill*, 864.
- Contreras, V. M. (2022). *UN MODELO PREDICTIVO INTERPRETABLE PARA LA ESTIMACIÓN DEL INGRESO MONETARIO DE CLIENTES BANCARIOS BASADO EN XGBOOST Y SHAP.* Concepción: UNIVERSIDAD DE CONCEPCIÓN FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS.
- Subirats, L. (2019). *Web Scraping.* Catalunya.
- Barbosa, J. (2022). *Datos estructurados, ¿qué son y cómo se crean?* Marketeros.