



UNIVERSIDAD
BICENTENARIA

INTELIGENCIA DE DATOS APLICADA AL ANÁLISIS COMERCIAL DE VEHÍCULOS USADOS EN ECUADOR

Almeida Coba, Christian David

Rivadeneira Fuel, Gabriel Alejandro



AUTORIDADES

Dr. Basilio Sánchez Aranguren
Presidente

Dr. Gustavo Sánchez
Rector

Dra. Mirian Regalado
Vicerrectora Académica

Dra. Zeyda Padilla
Vicerrectora Administrativa

Dra. Edilia Papa
Secretaria General



DECANATO DE INVESTIGACIÓN, EXTENSIÓN Y POSTGRADO

Dr. José Cordero
Decano

Abog. María T. Ramírez MSc.
Directora de Postgrado

Dra. Maite Marrero
Directora de Investigación

Dra. Yesenia Centeno
Responsable del Fondo Editorial



FACULTAD DE INGENIERÍA

Ing. Belkys Ramírez, MSc.

Decano de la Facultad de Ingeniería

Ing. José Luis Ceballos

Directora de la Escuela de Ingeniería en Sistema

Ing. Alexys Rodríguez, MSc.

Director de la Escuela de Ingeniería Eléctrica



COMITÉ EDITORIAL

Dra. Mirian Regalado Vicerrectorado Académico.

Dr. Ing. José Cordero Decanato de Investigación, Extensión y Postgrado

Dr. Ing. Cristina Rojas. Coordinadora de la Línea de Investigación Institucional Ciencia, Tecnología e Innovación Social, UBA-Venezuela

Dr. Víctor Hermoso. Coordinador de la Línea de Investigación Institucional Biodiversidad Ambiente y Salud. UBA-Venezuela

Ing. José Luis Ceballos. Directora de la Escuela de Ingeniería en Sistema

Ing. Alexys Rodríguez, MSc. Director de la Escuela de Ingeniería Eléctrica

Edición y Revisión General

Dra. Yesenia Centeno. FEUBA

1ra. Edición: diciembre, 2025

Lugar: Turmero, Venezuela

Depósito Legal: AR2025000261

ISBN: 978-980-455-041-6

Reservados todos los derechos
conforme a la Ley

Se permite la reproducción total o
parcial del libro siempre que se
indique expresamente la fuente

Fecha de aceptación: agosto, 2025

Fecha de publicación: diciembre, 2025

Diseño de Portada

Vicerrectorado de Información
y Comunicación

Ilustraciones:

[Freepik AI](#)

<https://es.123rf.com/>

ISBN: 978-980-455-041-6





©Fondo Editorial de la Universidad Bicentenaria de Aragua

Título: Inteligencia de Datos Aplicada al Análisis Comercial de Vehículos Usados en Ecuador

Almeida Coba, Christian David
Ing. Master en Visual Analytics and Big Dat
Universidad de Las Américas (UDLA): Quito, Ecuador
christian.almeida@udla.edu.ec
<https://orcid.org/0009-0003-5191-9711>

Rivadeneira Fuel, Gabriel Alejandro
Ing. Master en Visual Analytics and Big Dat
Universidad Internacional del Ecuador: Quito, Ecuador
garivadeneirafu@uide.edu.ec
<https://orcid.org/0009-0002-1361-3896>

**Serie Ingeniería,
Volumen 2 Número 5, 2025**

La Serie Ingeniería corresponde al Fondo Editorial anclada a la Facultad de Ingeniería de la Universidad Bicentenaria de Aragua (FEUBA), dirigida a la publicación de productos intelectuales resultantes de la Ingeniería, Inventiva, Innovación y Creatividad de estudiantes, docentes y público en general. Tiene como propósito divulgar las experiencias de interés para el desarrollo de la investigación, productividad y avances de la ciencia. Es una publicación semestral arbitrada por el sistema doble ciego.



PRÓLOGO

La transformación digital ha demostrado que los datos son, hoy en día, uno de los recursos más valiosos para comprender y anticipar el comportamiento de los mercados. En el ámbito automotriz, donde confluyen dinámicas económicas, sociales y tecnológicas, el análisis de datos se convierte en una herramienta esencial para generar conocimiento y tomar decisiones acertadas.

Este libro, *Inteligencia de Datos aplicada al Análisis Comercial de Vehículos Usados en Ecuador*, constituye una valiosa contribución académica y práctica a un campo poco explorado en nuestro país: el mercado de autos de segunda mano. A lo largo de sus páginas, los autores combinan rigurosidad científica con aplicaciones concretas, ofreciendo un recorrido que va desde la obtención de datos mediante técnicas de *web scraping*, hasta la construcción de modelos predictivos que permiten estimar el valor de un vehículo en función de sus características.

El lector encontrará aquí mucho más que cifras y gráficos. Este texto abre una ventana a la comprensión de un sector dinámico, estratégico y de gran impacto en la vida cotidiana de miles de ecuatorianos. La obra demuestra que la ciencia de datos no es exclusiva de grandes corporaciones o de mercados internacionales: también puede aplicarse, con creatividad y precisión, a problemáticas locales que requieren soluciones innovadoras.

Con un lenguaje claro, ejemplos pertinentes y un sólido respaldo metodológico, este libro se convierte en un referente para estudiantes, investigadores, empresarios y tomadores de decisiones interesados en el cruce entre tecnología y economía real. Sus hallazgos no solo orientan la compra y venta de vehículos, sino que también abren el camino hacia una cultura de decisiones basadas en evidencia, donde la información confiable se traduce en oportunidades.

Más allá de los aspectos técnicos, este estudio sienta las bases para una modernización del ecosistema digital en el país, impulsando a las plataformas de comercio electrónico a adoptar estándares de transparencia mucho más elevados. Al exponer la efectividad de los modelos de regresión, se incentiva la creación de herramientas de tasación en tiempo real que podrían integrarse directamente en los portales de venta, permitiendo que tanto el ciudadano común como las instituciones financieras cuenten con valores referenciales dinámicos. Este avance no solo reduce el riesgo de estafas o sobrepuestos, sino que también fomenta un mercado mucho más fluido y equitativo, donde la confianza del usuario se sustenta en la solidez de algoritmos validados y no en la especulación momentánea de los vendedores.

La obra se proyecta como una guía esencial para la formulación de políticas públicas relacionadas con la movilidad y el transporte sostenible en el contexto nacional. Esta obra no termina con la presentación de sus resultados, sino que inicia un diálogo necesario sobre cómo la ciencia de datos puede contribuir a una planificación urbana y económica más inteligente. Se invita al lector a recorrer este texto con la certeza de que cada página es un ejemplo de cómo la inteligencia de datos puede transformar realidades. Más que un estudio técnico, esta obra es una apuesta por la innovación y un aporte significativo al conocimiento aplicado en Ecuador.

Tabla de Contenidos

PRÓLOGO	5
Introducción.....	8
1. Encendiendo el Motor del Conocimiento	11
1.1 ¿Por Qué Estudiar el comercio de los Vehículos Usados en Ecuador?	12
Capítulo 2	19
2. El Camino Recorrido:	19
Panorama del Mercado Automotriz en Ecuador	19
Capítulo 3	28
3. Entre saberes y tecnología: el punto de partida.	28
Capítulo 4	34
4. Tras bambalinas: el engranaje metodológico.	34
Capítulo 5	44
5. Estudio y Gestión de los Datos	44
Capítulo 6	80
6. Explorando el comportamiento del mercado automotor ecuatoriano.....	80
Capítulo 7	85
7. Aprender del pasado para anticipar el valor de un vehículo.....	85
Capítulo 8	92
8. Resultados, Hallazgos y Aplicaciones.....	92
Capítulo 9	97
9. Conclusiones y Caminos por Explorar	97
Referencias	102
Anexos	104



INTRODUCCIÓN

En un mercado automotriz cada vez más competitivo, conocer el valor real de un vehículo de segunda mano se convirtió en una herramienta clave tanto para vendedores como para compradores. En este contexto, el presente trabajo analizó el mercado ecuatoriano de automóviles usados, tomando como fuente principal de información la página web Patio Tuerca Ecuador, reconocida por su amplia base de anuncios clasificados.

Para ello, se aplicó la técnica de web scraping, que permitió extraer de forma automática datos en formato HTML relacionados con diversas marcas, modelos, años de fabricación, kilometraje, precios y otras características de los vehículos. Posteriormente, la información se procesó y resumió mediante medidas estadísticas descriptivas, identificando patrones relevantes en la oferta disponible.

El estudio incorporó además un modelo de regresión para determinar la relación entre variables independientes como el año, el kilometraje, el tipo de combustible o la marca y la variable dependiente, el precio de venta. Los resultados revelaron, por ejemplo, que el año de fabricación y el kilometraje fueron los factores que más influyeron en el valor de comercialización, mientras que la marca mostró una incidencia significativa en ciertos segmentos del mercado. Finalmente, el uso del software libre R permitió realizar un tratamiento robusto de los datos y generar una referencia de precios ajustada a las características de cada vehículo, constituyendo así una herramienta útil para estimar valores de mercado y facilitar decisiones de compra y venta más informadas.

El estudio del mercado de vehículos usados en Ecuador se justifica por su enorme relevancia económica y social, al ser el principal mecanismo de acceso a la movilidad para gran parte de la población. No obstante, la persistente asimetría de información y la falta de análisis estadísticos rigurosos en este sector generan una incertidumbre que afecta a compradores y vendedores. La implementación de técnicas modernas de ciencia de datos, como el web scraping y el modelado predictivo, surge como una solución estratégica para dotar de transparencia a este mercado, permitiendo transformar registros digitales dispersos en conocimiento valioso para la toma de decisiones informadas y la formulación de políticas públicas.

La motivación de este proyecto trasciende el ámbito académico al buscar una aplicación práctica de herramientas de código abierto como R y RStudio en un entorno real y heterogéneo. Al sistematizar el análisis de variables críticas como la depreciación, el kilometraje y la ubicación geográfica, se pretende demostrar que la evidencia cuantitativa puede sustituir a las percepciones subjetivas en la valoración vehicular. Este enfoque no solo optimiza la comprensión de la oferta y la demanda actual, sino que también establece un precedente metodológico replicable que fomenta una cultura de análisis basada en datos dentro del ecosistema automotor nacional.

A través de la integración de objetivos exploratorios y predictivos, esta investigación se consolida como un laboratorio aplicado donde la tecnología y la economía convergen para identificar oportunidades y anticipar tendencias. El uso de modelos de regresión lineal múltiple permite cuantificar con precisión el impacto de

atributos técnicos en el precio de venta, reduciendo el margen de especulación y permitiendo que tanto consumidores como empresarios identifiquen los modelos con mejor valor de reventa. De este modo, el análisis no solo describe el estado actual de las transacciones, sino que ofrece una contribución concreta al desarrollo económico mediante la democratización de información procesada y verificada.

Este libro, fruto de la investigación realizada, no solo ofrece un panorama detallado del mercado de autos usados en Ecuador, sino que también brinda herramientas prácticas para comprender y anticipar las dinámicas de precio en un sector en constante cambio. Al recorrer sus páginas, el lector encontrará datos claros, análisis fundamentados y modelos aplicados que pueden convertirse en aliados estratégicos para tomar decisiones acertadas. Más que un estudio técnico, esta obra es una invitación a descubrir el valor que se esconde detrás de cada vehículo y a entender cómo la información, bien analizada, se transforma en conocimiento útil y accionable.

Al comprender con precisión la edad promedio del parque automotor y los niveles de recorrido de la flota usada, las autoridades pueden diseñar incentivos de renovación vehicular mucho más efectivos y segmentados por provincias. Es, en definitiva, un testimonio del talento analítico local aplicado al bienestar social, demostrando que el conocimiento riguroso es la herramienta más poderosa para transformar los desafíos de un sector estratégico en oportunidades de desarrollo para todo el Ecuador.

Este libro trasciende el rigor académico para ofrecer una visión humanizada y práctica de la tecnología aplicada a la cotidianidad nacional. Al demostrar que la ciencia de datos puede resolver problemáticas locales, los autores desmitifican la idea de que la inteligencia artificial o el machine learning son exclusivos de las potencias globales o las grandes corporaciones. El lector descubrirá cómo factores geográficos, estacionales y técnicos interactúan para definir el mercado en ciudades como Quito o Guayaquil, encontrando en cada gráfico y tabla un recurso accionable para optimizar decisiones comerciales. Es, en esencia, una invitación a transitar hacia una cultura de evidencia, donde la información verificada se convierte en el motor que impulsa la competitividad y la transparencia en el sector automotor ecuatoriano.

Capítulo 1

Encendiendo el Motor del Conocimiento



Fuente: [Free text-to-image generator](#) | [Freepik AI](#)

El estudio del mercado de vehículos usados en Ecuador se presenta como una necesidad fundamental para comprender un sector dinámico que impacta significativamente en la economía nacional y en la movilidad ciudadana. Dado que para una gran parte de la población la adquisición de un vehículo nuevo resulta financieramente inviable, el mercado de segunda mano se convierte en la principal alternativa de acceso al transporte privado. No obstante, este sector ha operado históricamente bajo asimetrías de información, donde las decisiones de compra y venta suelen basarse en percepciones subjetivas o recomendaciones informales, lo que resalta la urgencia de aplicar un enfoque cuantitativo y riguroso basado en la ciencia de datos para aportar transparencia y eficiencia a las transacciones.

La motivación central de esta investigación radica en transformar datos fragmentados de plataformas digitales en conocimiento estructurado y útil para compradores, vendedores y responsables de políticas públicas. A través del uso de herramientas tecnológicas avanzadas en el lenguaje de programación R, se busca demostrar que incluso mercados con altos niveles de informalidad pueden ser analizados sistemáticamente para identificar patrones de depreciación, marcas con mayor valor de reventa y tendencias regionales. Este proyecto no solo cumple con un propósito académico y profesional de aprendizaje aplicado, sino que también aspira a fomentar una cultura de toma de decisiones basada en evidencia dentro del ecosistema automotor ecuatoriano.

Para alcanzar estos fines, el estudio se estructura bajo una metodología exploratoria y predictiva que abarca desde la recolección automatizada de datos mediante técnicas de web scraping hasta la implementación de modelos de regresión lineal múltiple. El proceso incluye una etapa exhaustiva de limpieza y transformación de datos, un análisis estadístico descriptivo y la creación de visualizaciones que faciliten la interpretación de los hallazgos por parte de un público no especializado. De esta manera, el presente trabajo ofrece una radiografía detallada del mercado automotor de usados, proporcionando herramientas prácticas para estimar precios con base en atributos

objetivos y anticipar los movimientos de un sector clave para el desarrollo económico del país.

1.1 ¿Por Qué Estudiar el comercio de los Vehículos Usados en Ecuador?

El mercado de vehículos usados en Ecuador constituye un sector dinámico, diverso y en constante transformación, que impacta directamente tanto en la economía nacional como en la vida cotidiana de miles de ciudadanos. Según datos del Instituto Nacional de Estadística y Censos (INEC) y reportes de portales especializados como Patiotuerca, cada año se realizan decenas de miles de transacciones de compra y venta de autos de segunda mano, moviendo millones de dólares y generando un ecosistema económico que involucra a vendedores, compradores, intermediarios, concesionarios, talleres mecánicos, aseguradoras y entidades financieras. Este mercado es especialmente relevante porque para una gran parte de la población, adquirir un vehículo nuevo no es financieramente viable, lo que convierte al sector de usados en la principal puerta de acceso a la movilidad privada.

A pesar de su peso económico, el mercado de autos usados en Ecuador carece de investigaciones profundas que integren un enfoque cuantitativo y basado en datos. La mayor parte de las decisiones de compra-venta se realizan con base en percepciones personales, recomendaciones informales o anuncios publicados en portales web, lo que genera asimetría de información y deja espacio para la especulación de precios. Además, factores como la depreciación, la reputación de ciertas marcas y modelos, la variación de precios entre ciudades, el kilometraje y el estado mecánico del vehículo influyen de forma directa en el valor final, pero rara vez son evaluados con un análisis estadístico riguroso.

Analizar este mercado desde una perspectiva de ciencia de datos permite detectar patrones ocultos y generar información que beneficie a todos los actores involucrados. Por ejemplo, se pueden identificar los modelos más demandados en cada región, evaluar cómo evoluciona el precio de un mismo vehículo a lo largo del tiempo o estimar el impacto del kilometraje y la antigüedad en su valor de reventa. Asimismo, contar con información verificada y procesada reduce la incertidumbre para los compradores, fortalece la

confianza en las transacciones y puede incluso influir en políticas públicas relacionadas con movilidad, seguridad vial y sostenibilidad.

En un contexto donde el acceso a datos es un recurso estratégico, aplicar técnicas como *web scraping* para recopilar información directamente de plataformas de venta y combinarla con análisis estadístico y modelos predictivos es una herramienta poderosa. No solo permite describir el estado actual del mercado, sino también anticipar tendencias, identificar oportunidades y proponer mejoras que aumenten la transparencia y competitividad del sector. Estudiar los autos usados en Ecuador, por lo tanto, no es simplemente un ejercicio académico: es una contribución concreta al desarrollo económico y a la toma de decisiones más informadas por parte de consumidores, empresarios y autoridades.

1.2 Motivación del proyecto

Este proyecto nace de la inquietud por aprovechar el potencial de los datos abiertos y las herramientas digitales para comprender de manera más profunda un mercado cotidiano y altamente relevante: el de los vehículos usados en Ecuador. A diario, miles de transacciones se realizan a través de portales web, concesionarios y negociaciones directas, pero la información que respalda dichas operaciones suele estar fragmentada, incompleta o sujeta a percepciones personales. Al explorar plataformas como PatioTuerca, se descubre una fuente rica en datos sobre precios, modelos, marcas, kilometraje y otros atributos clave que, adecuadamente recopilados y analizados, pueden ofrecer una visión más clara y objetiva de este sector.

La motivación central es demostrar que, con las herramientas y metodologías correctas, incluso un mercado heterogéneo, poco estandarizado y con altos niveles de informalidad puede ser estudiado rigurosamente. Esto no solo permite generar conocimiento académico, sino que también aporta valor práctico a compradores, vendedores, investigadores y responsables de políticas públicas. Comprender cómo varían los precios según factores objetivos, identificar las marcas más estables en valor de reventa o detectar tendencias emergentes en la demanda, puede influir directamente en decisiones individuales y colectivas.

Existe también una motivación personal y profesional: **aprender haciendo**. Este trabajo se concibe como una oportunidad para poner en práctica habilidades de

programación en R, técnicas de *web scraping*, análisis estadístico, visualización de datos y modelado predictivo, dentro de un contexto real. No se trata únicamente de un ejercicio teórico, sino de un laboratorio aplicado en el que se enfrentan desafíos propios de proyectos de datos: limpieza de información, tratamiento de valores atípicos, selección de variables relevantes, evaluación de modelos y presentación de resultados comprensibles.

Finalmente, esta investigación busca servir como ejemplo para otros estudiantes, analistas o profesionales interesados en explorar el cruce entre ciencia de datos y economía real. El objetivo no es solo describir el estado actual del mercado automotriz de segunda mano en Ecuador, sino también inspirar a que se desarrollen más estudios de este tipo, ampliando la base de conocimiento disponible y fomentando una cultura de decisiones basadas en evidencia.

1.3 Objetivos del análisis

El estudio persigue tres grandes objetivos:

- Explorar las características del mercado de vehículos usados en Ecuador mediante datos obtenidos de plataformas digitales.
- Analizar de forma estadística y visual la oferta vehicular, considerando variables como marca, modelo, año, kilometraje y precio.
- Predecir el precio de un vehículo usado con base en sus atributos, a través de modelos de regresión.

Cada uno de estos objetivos está diseñado para generar conocimiento útil y aplicable tanto para compradores como para vendedores, analistas, investigadores o interesados en el cruce entre tecnología y mercado.

1.4 Enfoque metodológico

Este estudio se desarrolla bajo un enfoque cuantitativo y exploratorio, orientado a descubrir patrones, relaciones y tendencias dentro del mercado de vehículos usados en Ecuador a partir de datos reales obtenidos en línea. La investigación se sustenta en tres pilares: la recolección automatizada de datos mediante *web scraping*, el análisis estadístico con herramientas modernas de ciencia de datos y la construcción de visualizaciones y modelos predictivos que permitan transformar datos crudos en información clara y útil.

El trabajo metodológico se estructuró en varias etapas interdependientes:

1. Extracción de datos

Se utilizó la técnica de web scraping para recopilar información directamente desde la plataforma PatioTuerca, uno de los principales portales de compraventa de vehículos en Ecuador. La elección de esta fuente respondió a su alta representatividad, volumen de datos y variedad de atributos disponibles (marca, modelo, año, kilometraje, precio, ciudad, entre otros).

2. Limpieza y transformación de datos

Una vez recolectada la información, se procedió a depurar la base de datos eliminando duplicados, corrigiendo inconsistencias, gestionando valores faltantes y filtrando valores atípicos que pudieran distorsionar el análisis. Además, se homogenizaron formatos y unidades para garantizar la coherencia de las variables.

3. Análisis exploratorio de datos (EDA)

Se aplicaron técnicas estadísticas descriptivas para comprender la distribución y comportamiento de cada variable, identificar correlaciones y detectar patrones preliminares. Esta fase sirvió como base para decidir qué variables incluir en los modelos posteriores.

4. Visualización de resultados

Se desarrollaron gráficos y paneles que facilitan la interpretación de tendencias, comparaciones y relaciones. Las visualizaciones no solo buscan transmitir hallazgos, sino también democratizar la comprensión de los datos, haciéndolos accesibles para públicos no especializados.

5. Modelado predictivo

Se implementó un modelo de regresión lineal múltiple con el fin de estimar el precio de un vehículo en función de sus características clave. Este modelo fue evaluado para garantizar su robustez, verificando supuestos como la ausencia de multicolinealidad y la homocedasticidad de los residuos.

6. Presentación e interpretación de resultados

Finalmente, se sintetizaron los hallazgos en conclusiones prácticas y recomendaciones, orientadas a compradores, vendedores y analistas interesados en entender las dinámicas de este mercado.

Para todo el proceso se emplearon R y RStudio, aprovechando librerías como rvest para el scraping, dplyr y tidyr para manipulación de datos, ggplot2 para visualización y caret para modelado. Esta elección responde tanto a la versatilidad de R en el manejo de datos como a su condición de software libre, lo que lo hace accesible para estudiantes, investigadores y profesionales sin barreras de licencia.

Con este enfoque, la investigación no solo describe un caso concreto de análisis de mercado, sino que también ofrece un modelo metodológico replicable en otros contextos y sectores.

1.5 Estructura del libro

El presente libro ha sido diseñado para guiar al lector de forma progresiva y ordenada, partiendo de una visión panorámica del mercado de vehículos usados en Ecuador hasta llegar a los aspectos más técnicos y especializados del análisis de datos y modelado predictivo. La secuencia de capítulos no solo sigue un orden lógico, sino que también busca mantener el interés y facilitar la comprensión, incluso para quienes se aproximan por primera vez a la ciencia de datos aplicada al sector automotor.

La obra se compone de nueve capítulos principales y varios apéndices complementarios, organizados de la siguiente manera:

Capítulo 2: Presenta un análisis del contexto del mercado automotor ecuatoriano, con énfasis en el segmento de autos de segunda mano, sus actores principales, mecanismos de comercialización y las dinámicas que lo caracterizan.

Capítulo 3: Revisa las investigaciones previas y las herramientas tecnológicas más relevantes para el análisis de mercados, incluyendo el uso de R, técnicas de web scraping y visualización de datos.

Capítulo 4: Describe la metodología utilizada en el proyecto, desde la planificación y selección de herramientas hasta el diseño de cada etapa del análisis.

Capítulo 5: Detalla el proceso de gestión de datos, explicando cómo se extrajo la información de la plataforma PatioTuerca, cómo se depuró y cómo se estructuró para el análisis posterior.

Capítulo 6: Se adentra en el análisis estadístico y la visualización de datos, explorando patrones, comparaciones entre marcas, modelos y años, así como representaciones gráficas que facilitan la interpretación de resultados.

Capítulo 7: Explica la construcción de un modelo predictivo de precios basado en regresión lineal, evaluando las variables más influyentes y mostrando ejemplos de predicciones en escenarios reales.

Capítulo 8: Integra y sintetiza los hallazgos más relevantes, identifica patrones de comportamiento en el mercado y presenta recomendaciones prácticas para compradores, vendedores y analistas.

Capítulo 9: Cierra con conclusiones generales, reflexiona sobre las limitaciones del estudio y propone líneas futuras de investigación que podrían enriquecer y ampliar el análisis.

Finalmente, los apéndices aportan un valor añadido para el lector más técnico o investigador. También incluye el artículo académico derivado del estudio, que condensa la investigación en un formato formal y validado por pares. El código completo en R, comentado y listo para ser ejecutado, de forma que otros puedan replicar, adaptar o ampliar el análisis en diferentes contextos.

En conjunto, esta estructura permite que el libro funcione tanto como una guía práctica para proyectos similares como un documento de referencia para comprender mejor un mercado específico a través del poder de los datos. Por ello que, la estructura progresiva del libro asegura que el lector comprenda desde el contexto macroeconómico hasta la implementación técnica de modelos de regresión lineal. Los objetivos planteados buscan explorar, analizar y predecir el comportamiento del mercado de forma integral, ofreciendo tanto resultados interpretables como el código fuente necesario para su replicación. De esta manera, la investigación se consolida como un laboratorio aplicado donde la tecnología y la economía convergen para identificar oportunidades, anticipar tendencias y proponer mejoras concretas en la competitividad y transparencia del comercio de vehículos de segunda mano.

Capítulo 2

El Camino Recorrido:

**Panorama del Mercado
Automotriz en Ecuador**



Fuente: [Free text-to-image generator](#) | [Freepik AI](#)

El estudio del mercado de vehículos usados en Ecuador se justifica por su enorme relevancia económica y social, al ser el principal mecanismo de acceso a la movilidad para gran parte de la población. No obstante, la persistente asimetría de información y la falta de análisis estadísticos rigurosos en este sector generan una incertidumbre que afecta a compradores y vendedores. La implementación de técnicas modernas de ciencia de datos, como el web scraping y el modelado predictivo, surge como una solución estratégica para dotar de transparencia a este mercado, permitiendo transformar registros digitales dispersos en conocimiento valioso para la toma de decisiones informadas y la formulación de políticas públicas.

La motivación de este proyecto trasciende el ámbito académico al buscar una aplicación práctica de herramientas de código abierto como R y RStudio en un entorno real y heterogéneo. Al sistematizar el análisis de variables críticas como la depreciación, el kilometraje y la ubicación geográfica, se pretende demostrar que la evidencia cuantitativa puede sustituir a las percepciones subjetivas en la valoración vehicular. Este enfoque no solo optimiza la comprensión de la oferta y la demanda actual, sino que también establece un precedente metodológico replicable que fomenta una cultura de análisis basada en datos dentro del ecosistema automotor nacional.

A través de la integración de objetivos exploratorios y predictivos, esta investigación se consolida como un laboratorio aplicado donde la tecnología y la economía convergen para identificar oportunidades y anticipar tendencias. El uso de modelos de regresión lineal múltiple permite cuantificar con precisión el impacto de atributos técnicos en el precio de venta, reduciendo el margen de especulación y permitiendo que tanto consumidores como empresarios identifiquen los modelos con mejor valor de reventa. De este modo, el análisis no solo describe el estado actual de las transacciones, sino que ofrece una contribución concreta al desarrollo económico mediante la democratización de información procesada y verificada.

2.1 Contexto del mercado de autos de segunda mano

El ser humano presenta diversas necesidades y deseos que, en función de su capacidad adquisitiva, se transforman en demandas concretas. Estas necesidades han sido ampliamente estudiadas y jerarquizadas en la conocida pirámide de Maslow, que contempla cinco niveles: fisiológicas, seguridad, sociales, autoestima y autorrealización. Dentro de las necesidades vinculadas a la seguridad y a la movilidad, el transporte ocupa un lugar esencial, convirtiéndose en un elemento clave para el desarrollo personal y social. En este sentido, la adquisición de un vehículo ya sea nuevo o usado, constituye una meta importante para muchas personas (Turrienzo, 2016).

En Ecuador, el sector automotriz ha experimentado un notable incremento en la compra y venta de vehículos usados, fenómeno observable en diferentes ciudades del país. La comercialización de autos de segunda mano no solo responde a necesidades prácticas de movilidad, sino que también refleja patrones económicos, disponibilidad de financiamiento y tendencias culturales que hacen de este mercado un campo de estudio relevante.

Según datos del Instituto Nacional de Estadística y Censos (INEC), el parque automotor ecuatoriano creció un 57% en un periodo de apenas cinco años, pasando de 1'226.349 vehículos en 2010 a 1'925.368 en 2015 (INEC, 2015). Este crecimiento sostenido se explica tanto por la incorporación de autos nuevos como por el dinamismo del mercado de segunda mano, donde la renovación de flotas y el cambio de preferencias motivan la circulación constante de vehículos usados.

El Informe Anuario de Transporte del INEC revela que entre 2008 y 2016 el número de vehículos matriculados mostró un incremento constante año tras año, lo que confirma la importancia que tiene para la sociedad ecuatoriana contar con un medio de transporte propio (INEC, 2016). Este mismo informe destaca que, aunque Pichincha no es la provincia con mayor población, concentra una proporción significativa de matriculaciones, superada en habitantes por Guayas, pero con una fuerte presencia en el mercado automotor (INEC, 2016).

Tabla 1

Número de habitantes por provincia

Provincia	Número de habitantes
Guayas	3.645.483
Pichincha	2.576.287
Manabí	1.369.780
Los Ríos	778.115

Nota. Fuente INEC, (2016)

El mercado de vehículos usados en Ecuador ha sido históricamente una alternativa estratégica para quienes buscan un equilibrio entre precio, calidad y funcionalidad. Ante el incremento sostenido de los precios de los vehículos nuevos y las restricciones que enfrentan ciertos sectores de la población para acceder a crédito automotriz, los autos usados se posicionan como una opción más asequible y adaptable a diferentes realidades económicas.

Además, este mercado refleja la diversidad geográfica y socioeconómica del país. En zonas rurales y agrícolas, la preferencia suele inclinarse hacia camionetas y vehículos robustos, capaces de soportar caminos irregulares y transportar carga. En contraste, en áreas urbanas densamente pobladas, prevalecen los autos compactos, de bajo consumo y fáciles de maniobrar y estacionar.

En años recientes, incluso el segmento de segunda mano ha experimentado un cambio en las prioridades del consumidor. Se observa un interés creciente por vehículos con mejor eficiencia de combustible, menores emisiones contaminantes y, en algunos casos, tecnologías de asistencia al conductor. Esta evolución revela un consumidor más consciente y exigente, influenciado tanto por factores económicos como por consideraciones medioambientales y tecnológicas.

En consecuencia, estudiar el mercado de autos usados en Ecuador no solo permite entender cómo se satisfacen necesidades básicas de transporte, sino también cómo estas decisiones de compra se entrelazan con cambios sociales, tecnológicos y

ambientales, configurando un sector dinámico y con un alto potencial para ser optimizado mediante el análisis de datos.

2.2 Plataformas de comercialización digital

El auge de plataformas digitales como PatioTuerca, OLX Autos y Encuentra24 ha transformado radicalmente la forma en que se comercializan los vehículos usados en Ecuador. Estas herramientas actúan como mercados virtuales que permiten a los compradores comparar precios, modelos, año de fabricación, kilometraje, estado de conservación y ubicación geográfica de forma inmediata, lo que les brinda la posibilidad de tomar decisiones más informadas. Para los vendedores, ya sean particulares o concesionarios, estas plataformas amplían el alcance del mercado, permitiendo que un vehículo publicado en una ciudad pueda ser visto y adquirido por compradores de cualquier región del país.

Esta digitalización del proceso de compraventa ha generado nuevas oportunidades y desafíos. Entre las oportunidades destacan la ampliación del alcance geográfico, la posibilidad de segmentar audiencias según preferencias y la recolección de datos masivos para análisis de mercado. Sin embargo, también se presentan retos importantes, como la necesidad de estandarizar la información publicada, garantizar la veracidad de los datos (especialmente en lo relacionado con kilometraje y antecedentes del vehículo), y comprender la dinámica de visibilidad que imponen los algoritmos y sistemas de posicionamiento dentro de estas plataformas.

Los sitios web especializados no solo cumplen una función de catálogo, sino que han evolucionado hacia experiencias de usuario personalizadas, donde el sistema recomienda vehículos basándose en búsquedas previas, historial de interacción y patrones de navegación. Esta capa de personalización, combinada con la masificación del acceso a internet y la expansión del uso de smartphones, ha multiplicado el volumen de transacciones que se gestionan total o parcialmente en línea.

En consecuencia, cada publicación en estas plataformas se convierte en un registro de datos estructurado y valioso, con potencial para ser analizado estadísticamente: precios, características técnicas, ubicación, tiempo que tarda en venderse, y variables

asociadas a la percepción del valor. Para proyectos como el presente, estas bases de datos abiertas representan un insumo fundamental para modelar el comportamiento del mercado y generar predicciones con base en evidencia real.

2.3 Tendencias y dinámicas actuales

Dentro del mercado automotor ecuatoriano se identifican varias tendencias y comportamientos que moldean la dinámica de compra y venta:

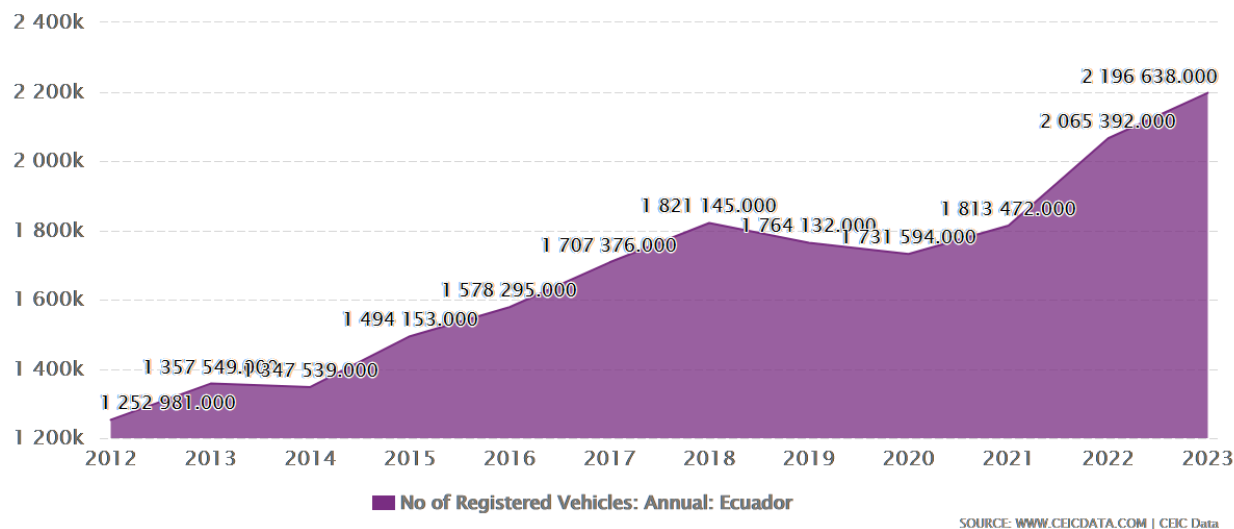
- Alta rotación de modelos populares: Marcas como Chevrolet, Kia, Hyundai y Toyota se mantienen entre las más buscadas, especialmente en versiones que han demostrado durabilidad y facilidad de mantenimiento, lo que favorece su valor de reventa.
- Creciente oferta de vehículos seminuevos certificados: Impulsada principalmente por concesionarios y casas comerciales que ofrecen revisiones técnicas previas y garantías limitadas, buscando diferenciarse del mercado informal y brindar mayor confianza al comprador.
- Aumento de la conciencia ecológica: Aunque incipiente, existe un interés progresivo por modelos híbridos y eléctricos, incluso en el segmento de segunda mano. Este fenómeno se ve estimulado por la preocupación medioambiental y las políticas de exoneración de impuestos para este tipo de vehículos.
- Fluctuación de precios influenciada por factores externos: El valor de los vehículos se ve afectado por variables macroeconómicas como el tipo de cambio, la disponibilidad y costo de repuestos, las políticas de importación, así como la inflación interna.
- Incorporación de tecnologías de validación de historial: Herramientas que verifican el kilometraje real, los accidentes previos o el número de propietarios, las cuales inciden directamente en la valoración y en la confianza del comprador.

Estos elementos conforman un entorno altamente dinámico, donde la información publicada en línea se convierte en un activo estratégico tanto para vendedores como para investigadores. Analizar grandes volúmenes de estos datos permite construir una

radiografía detallada del mercado, identificar patrones de comportamiento y anticipar movimientos de oferta y demanda, aportando así un valor práctico al estudio del sector.

Figura 1

Número de vehículos motorizados matriculados en Ecuador (2019–2023)



Nota. Fuente CEIC – Indicator: Number of Registered Vehicles

Según el Instituto Nacional de Estadística y Censos (INEC), el parque automotor matriculado en Ecuador muestra un crecimiento sostenido en los últimos años:

Tabla 2

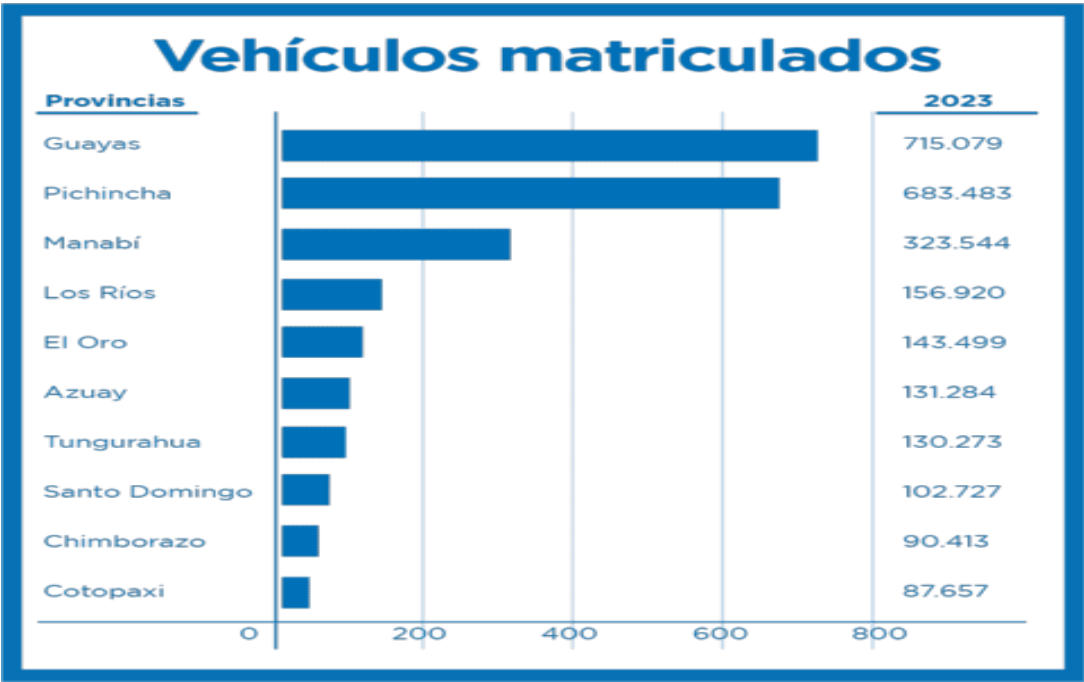
Vehículos matriculados en Ecuador

Año	Vehículos matriculados
2019	2 311 960
2020	2 361 175
2021	2 535 853
2022	2 880 910
2023	3 065 967

Nota. Fuente CEIC – Indicator: Number of Registered Vehicles

Este aumento continuo refuerza la idea de que el parque vehicular no solo crece en cantidad, sino que cada vehículo representado en estas cifras—nuevo o usado—es una posible unidad para análisis estadístico y modelado dentro del mercado. Esto posiciona al sector automotor como un terreno fértil para estudiar dinámicas de oferta, demanda y comportamiento económico.

Figura 2
Vehículos matriculados por provincia en Ecuador (2023)



Nota. Fuente: INEC – Estadísticas de Transporte, Tabulados 2023.

La provincia de Guayas lidera en número absoluto de matrículas, con 715 079 vehículos registrados en 2023, seguida de Pichincha con 683 483 —una variación interanual del 20,10% en esta última. Manabí, Los Ríos y El Oro también registraron cifras destacadas, mientras que Azuay mostró una caída notable en el número de vehículos matriculados. Estos datos reflejan dinámicas relevantes del mercado:

1. Alta concentración regional: La mayor masa vehicular se encuentra en Guayas y Pichincha, lo que convierte estas provincias en áreas prioritarias para estrategias de mercado y modelado de precios.

2. Variaciones significativas: Pichincha creció más del 20 % en comparación con 2022, lo que sugiere una revitalización del mercado vehicular en esa región.

3. Tendencias heterogéneas: Zonas como Azuay muestran una disminución notable, lo cual puede deberse a factores específicos como migración, cambios económicos o políticas locales.

El mercado de vehículos de segunda mano en Ecuador se ha consolidado como una respuesta estratégica a las necesidades de movilidad y seguridad de la población, fundamentado en un crecimiento sostenido del parque automotor que superó los tres millones de unidades en 2023. Esta dinámica no solo refleja la búsqueda de alternativas económicas frente a los costos de los autos nuevos, sino que también evidencia una especialización geográfica donde provincias como Guayas y Pichincha concentran la mayor actividad comercial. La evolución del sector muestra un consumidor más exigente que prioriza la eficiencia, la durabilidad de marcas tradicionales y, de forma incipiente, la sostenibilidad ambiental a través de tecnologías híbridas y eléctricas.

La transformación digital ha sido el catalizador fundamental en esta evolución, desplazando los métodos tradicionales hacia plataformas virtuales que democratizan el acceso a la información y amplían el alcance geográfico de las transacciones. El uso de estos mercados digitales ha permitido estandarizar variables críticas como el kilometraje, el año de fabricación y el estado técnico, reduciendo las asimetrías de información entre compradores y vendedores. Esta digitalización convierte cada interacción en un registro de datos valioso, facilitando una toma de decisiones más informada y permitiendo la personalización de la experiencia de usuario mediante algoritmos de recomendación.

Por tanto, la integración de grandes volúmenes de datos provenientes de registros oficiales y plataformas de comercialización ofrece una oportunidad sin precedentes para el análisis estadístico y el modelado de precios. El sector se presenta como un entorno altamente dinámico donde factores macroeconómicos, políticas de importación y el surgimiento de vehículos seminuevos certificados configuran un ecosistema complejo. En última instancia, entender estas tendencias permite no solo interpretar el comportamiento actual de la oferta y la demanda, sino también anticipar movimientos futuros en un mercado que es pilar del desarrollo económico y social del país.

Capítulo 3

Entre Saberes y Tecnología: el Punto de Partida.



Fuente: [Free text-to-image generator](#) | [Freepik AI](#)

El estudio del mercado automotor en Ecuador se inserta en un contexto global marcado por la innovación tecnológica y la transición hacia una movilidad sostenible. La literatura reciente destaca cómo la introducción de vehículos eléctricos e híbridos, respaldada por políticas ambientales internacionales, ha comenzado a permear el mercado local, influyendo incluso en el segmento de autos de segunda mano. Este panorama se ve complementado por investigaciones previas que subrayan la importancia del mantenimiento vehicular en la seguridad vial y el impacto de las políticas energéticas nacionales, como la eliminación de subsidios a los combustibles, las cuales han reconfigurado las preferencias de los consumidores y la estructura de costos del transporte en el país.

A nivel nacional, el estado del arte revela una industria robusta compuesta por ensambladoras, comercializadoras y una vasta red de servicios complementarios que sostienen la economía del sector. No obstante, las investigaciones académicas identifican desafíos persistentes, tales como la asimetría de información entre compradores y vendedores y la coexistencia de mercados formales e informales en regiones clave como Pichincha. Estos estudios previos enfatizan la necesidad de mecanismos de control más eficientes y de una mayor transparencia en variables críticas como el kilometraje real, el historial de siniestralidad y la procedencia de las unidades, factores que determinan el valor de reventa y la confianza en las transacciones.

En este escenario, el avance de las herramientas digitales y técnicas de recolección automática, como el web scraping, surge como una solución tecnológica estratégica para el análisis de mercados dinámicos. La capacidad de extraer masivamente datos actualizados sobre precios y características técnicas desde portales especializados permite mitigar la falta de información y optimizar la toma de decisiones. En consecuencia, la integración de estas referencias tecnológicas con el análisis de las políticas arancelarias y comerciales vigentes proporciona una base sólida para comprender la evolución del parque automotor ecuatoriano y anticipar las tendencias que definirán la oferta y demanda de vehículos usados en el corto y mediano plazo.

3.1 Estado del Arte y Referencias Tecnológicas

En los últimos años, tanto a nivel nacional como internacional, se han desarrollado diversos estudios vinculados al mercado automotor, impulsados por la incorporación de nuevas tecnologías y por políticas gubernamentales que han influido de manera directa en la comercialización de vehículos.

3.1.1 Avances tecnológicos en el sector automotor

En el plano tecnológico, el mercado global ha experimentado la introducción de automóviles con características innovadoras, como los vehículos eléctricos e híbridos. Estos modelos, considerados una alternativa sostenible, contribuyen a la disminución de emisiones contaminantes y, al mismo tiempo, representan un ahorro económico significativo para los usuarios debido a su bajo consumo de combustibles fósiles o, en algunos casos, su total exención de estos (Pelayo, 2013).

La tendencia hacia la movilidad sostenible ha sido respaldada por políticas de incentivo en varios países, aunque su adopción en Ecuador aún es incipiente. No obstante, el interés por este tipo de vehículos, incluso en el mercado de segunda mano, ha crecido en los últimos años, configurando un segmento emergente que podría modificar la estructura de la oferta y la demanda en el corto y mediano plazo.

3.1.2 Cambios en la política energética y comercial ecuatoriana

En el contexto nacional, se han producido transformaciones relevantes en la política energética y en la regulación del comercio exterior de automotores. Durante el mandato del expresidente Lenín Moreno, el Gobierno ecuatoriano eliminó parcialmente el subsidio a las gasolinas “súper” y “extra” equivalentes a las denominadas internacionalmente “premium” y “regular”. Esta medida generó un incremento inmediato en los precios de estos combustibles y provocó un debate público sobre su impacto en la economía de los hogares y en la estructura de costos del transporte (Ministerio de Industrias y Productividad, 2018).

Adicionalmente, la política arancelaria ha establecido tasas diferenciadas de importación según el origen del vehículo. Por ejemplo, los automóviles provenientes de Europa pagan impuestos más bajos que los fabricados en otras regiones, con el fin de promover la diversificación del parque automotor y facilitar acuerdos comerciales.

3.1.3 Estudios previos en el ámbito nacional

La literatura académica nacional incluye investigaciones centradas en diversos aspectos del mercado automotor. Pelayo (2013) analizó el mantenimiento vehicular realizado por los propietarios y su incidencia en la seguridad de conducción, identificando que las prácticas preventivas y correctivas influyen directamente en la vida útil y el desempeño de los vehículos.

Desde un enfoque empresarial, la Asociación de Empresas Automotrices del Ecuador (2018) reporta que el país cuenta con cuatro ensambladoras de vehículos, 1.271 empresas dedicadas a la comercialización de automóviles nuevos y usados, y 3.126 empresas vinculadas a actividades complementarias como talleres mecánicos, centros de lavado, comercios de repuestos y servicios de reparación. Estos datos reflejan la magnitud del sector y su relevancia en la economía nacional.

3.1.4 Mercado formal e informal de vehículos usados

En lo que respecta a la comercialización, Quimbita (2016) estudió el mercado de vehículos usados en el Distrito Metropolitano de Quito, centrándose en la asimetría de información entre compradores y vendedores. Su investigación subraya la necesidad de considerar variables como el historial de mantenimiento, kilometraje real y siniestralidad previa para tomar decisiones informadas.

Por su parte, Ramírez (2017) abordó el mercado informal de compra y venta de vehículos usados en ferias y patios de la provincia de Pichincha. El autor identificó que estas actividades no registradas afectan la recaudación tributaria y propuso mecanismos de control más eficientes. Asimismo, resaltó el crecimiento sostenido de la comercialización, la amplia diversidad de precios y las facilidades de financiamiento que

han incentivado el acceso a automóviles por parte de distintos segmentos de la población.

3.1.5 Web scraping como herramienta para el análisis del mercado

Otro campo de estudio relevante para el presente trabajo es el uso del web scraping como técnica de recolección automática de información publicada en plataformas digitales. Condori (2014) propuso un modelo de web scraping para obtener datos actualizados en internet con notificaciones automáticas en dispositivos móviles, optimizando así la disponibilidad de información para los usuarios.

Baena (2018) destaca que esta metodología puede aplicarse en sectores como el financiero, el asegurador y el gubernamental, aprovechando información pública alojada en páginas web, foros, redes sociales y portales de datos abiertos. En el contexto automotor, su implementación permitiría a empresas, investigadores y organismos reguladores acceder a datos relevantes sobre precios, características técnicas y tendencias de mercado, sin incurrir en altos costos tecnológicos.

3.1.6 Síntesis y perspectivas

Si bien las investigaciones revisadas ofrecen un panorama amplio del sector automotor ecuatoriano, el dinamismo del mercado y la evolución de las políticas públicas obligan a mantener un monitoreo constante. La rápida aparición de nuevas tecnologías, cambios regulatorios y variaciones en el comportamiento del consumidor hacen que las variables que hoy condicionan la comercialización de vehículos de segunda mano puedan diferir significativamente en el futuro. Por ello, se recomienda la realización de estudios periódicos que permitan actualizar la comprensión del mercado y anticipar tendencias.

La revisión del estado del arte demuestra que el mercado automotor ecuatoriano se encuentra en un proceso de transición influenciado por innovaciones tecnológicas globales y ajustes en las políticas públicas nacionales. La introducción de vehículos híbridos y eléctricos, junto con la eliminación de subsidios a los combustibles y la aplicación de aranceles diferenciados, ha reconfigurado las preferencias del consumidor y la estructura de costos del transporte. Esta evolución exige que los actores del mercado no solo consideren variables tradicionales de compra, sino también factores de eficiencia energética y sostenibilidad que antes tenían un impacto marginal en la valoración de los bienes.

Por otro lado, la literatura académica subraya la persistencia de desafíos estructurales, como la asimetría de información y la alta informalidad en la comercialización de vehículos usados, especialmente en centros urbanos como Quito. Los estudios previos coinciden en que la falta de registros verificables sobre mantenimiento, kilometraje y siniestralidad genera incertidumbre y distorsiona la recaudación tributaria. Ante esta problemática, surge la necesidad de implementar soluciones que permitan transparentar el sector y brindar mayor seguridad jurídica y técnica tanto a compradores como a vendedores, reduciendo los riesgos asociados a las transacciones informales.

Al respecto, en el uso de herramientas tecnológicas avanzadas como el web scraping se posiciona como un pilar fundamental para la modernización del análisis de mercado en el país. La capacidad de recolectar y procesar datos masivos de plataformas digitales de forma automatizada permite superar las barreras de información y monitorear tendencias de precios en tiempo real sin incurrir en costos elevados. En conclusión, la integración de estas técnicas de recolección de datos con el conocimiento de las dinámicas económicas y regulatorias actuales permite construir una visión integral y actualizada del sector, facilitando la toma de decisiones estratégicas basadas en evidencia empírica.

Capítulo 4

Tras Bambalinas: el Engranaje Metodológico



Fuente: [Free text-to-image generator](#) | [Freepik AI](#)

La metodología de este proyecto se fundamenta en un diseño de investigación cuantitativo, descriptivo y predictivo, estructurado para analizar sistemáticamente el mercado de vehículos de segunda mano en Ecuador. A través de un enfoque integral, se combina la extracción automatizada de datos con el análisis estadístico avanzado, asegurando que cada etapa del proceso sea reproducible y transparente. La elección de herramientas de código abierto permite no solo una gestión eficiente de los recursos, sino también la creación de un flujo de trabajo sólido que abarca desde la captura inicial de información hasta la generación de conocimiento estratégico aplicable al sector automotriz.

El núcleo operativo del estudio se centra en el uso de R y RStudio, empleando librerías especializadas como rvest para el proceso de web scraping desde la plataforma PatioTuerca. Este procedimiento técnico permite transformar el contenido HTML del sitio web en bases de datos estructuradas, las cuales son sometidas a rigurosos protocolos de limpieza y depuración mediante dplyr y plyr para eliminar duplicados e inconsistencias. Este tratamiento previo es esencial para garantizar la calidad del análisis estadístico descriptivo posterior y la fiabilidad de las visualizaciones desarrolladas con ggplot2, las cuales facilitan la identificación de patrones de oferta y demanda en las principales ciudades del país.

En el componente predictivo se desarrolla mediante modelos de regresión lineal múltiple, diseñados para estimar el precio de venta en función de atributos clave como la marca, el modelo, el año y el kilometraje. Para asegurar la validez técnica de estas predicciones, se aplican pruebas diagnósticas específicas, incluyendo tests de multicolinealidad y heterocedasticidad, utilizando paquetes como lmtest y mctest. Este rigor metodológico permite no solo describir el estado actual del mercado, sino también construir herramientas de estimación precisas que reducen la incertidumbre en la valoración de vehículos usados, beneficiando tanto a compradores como a vendedores dentro del ecosistema digital.

4.1 Metodología del Proyecto

La metodología de este proyecto se estructuró de forma integral, combinando técnicas de recolección automatizada de datos, análisis estadístico y modelado predictivo con un enfoque sistemático y reproducible. El propósito fue garantizar que cada fase, desde la obtención de la información hasta la interpretación de los resultados, se ejecute de manera coherente y alineada con los objetivos del estudio. La selección de herramientas de código abierto no solo optimiza recursos, sino que también fomenta la transparencia y la replicabilidad del proceso. De esta manera, se estableció una base metodológica sólida que permitió comprender con precisión el comportamiento del mercado de vehículos de segunda mano en Ecuador y generar estimaciones confiables sobre su dinámica de precios.

4.1.1 Diseño del Estudio

Este proyecto tuvo como propósito principal el análisis del mercado de vehículos de segunda mano en el Ecuador, utilizando técnicas de extracción automática de datos y herramientas estadísticas avanzadas. El enfoque adoptado fue de tipo cuantitativo, descriptivo y predictivo, y se basó en la utilización de herramientas de código abierto para todo el ciclo de procesamiento de datos: desde la captura automatizada hasta el modelado estadístico.

La fuente principal de datos fue el sitio web www.patiotuerca.com.ec, un portal especializado en la venta de automotores en el país. Su posicionamiento digital, su especialización en vehículos usados y su volumen constante de publicaciones lo convierten en una fuente confiable para este tipo de análisis. El estudio contempló las siguientes fases:

- Selección y análisis de la fuente de datos
- Desarrollo del script para extracción de datos (web scraping)
- Procesamiento y depuración de datos
- Análisis estadístico descriptivo y exploratorio
- Aplicación de modelos de regresión para predicción de precios
- Visualización de resultados y hallazgos clave
- Evaluación y validación de los modelos construidos

4.2 Objetivos del Proyecto

A continuación se presentan los objetivos planteados en el proyecto del cual surge el presente libro.

4.2.1 Objetivo General

Realizar un análisis de las ventas de vehículos de segunda mano en el Ecuador mediante técnicas automatizadas de extracción de datos y modelos estadísticos predictivos, con el fin de estimar el precio de los vehículos en función de sus características.

4.2.1.1 Objetivos Específicos

- Capturar datos desde una página web en formato HTML dedicada a la venta de vehículos en diferentes ciudades del Ecuador.
- Establecer un script utilizando RStudio y R para la automatización eficiente del proceso de web scraping desde la plataforma www.patiotuerca.com.ec.
- Limpiar y depurar los datos obtenidos para su posterior análisis.
- Obtener medidas estadísticas descriptivas que resuman las principales variables.
- Aplicar modelos de regresión para identificar relaciones entre variables y construir predicciones del precio de venta.
- Predecir el precio de un vehículo usado en función de atributos como marca, modelo, año, kilometraje y ciudad de comercialización.
- Evaluar los resultados generados para convertir los datos procesados en conocimiento aplicable para diferentes actores del mercado automotriz.

4.3 Herramientas Utilizadas: R, RStudio y Tecnologías Asociadas

Para llevar a cabo el estudio, se utilizaron tecnologías y paquetes del entorno de programación R, que permiten trabajar con grandes volúmenes de datos de forma eficiente y reproducible:

Tabla 3*Descripción de las herramientas utilizadas*

Librería	Descripción
dplyr	Manipulación de datos
plyr	Manipulación de datos
rvest	Obtención de datos del sitio web
lmtest	Test de heterocedasticidad
sandwich	Corregir el test de heterocedasticidad
mctest	Test de multicolinealidad
leaps	Ayuda a visualizar el mejor modelo
ggplot2	Ayuda a graficar los datos

Nota. Fuente: Elaborado por el autor.

Además, se utilizó RStudio como entorno de desarrollo, por su interfaz intuitiva, manejo eficiente de objetos y soporte para desarrollo reproducible.

4.4 Etapas del Análisis

A continuación, se detallan las etapas metodológicas desarrolladas en este proyecto:

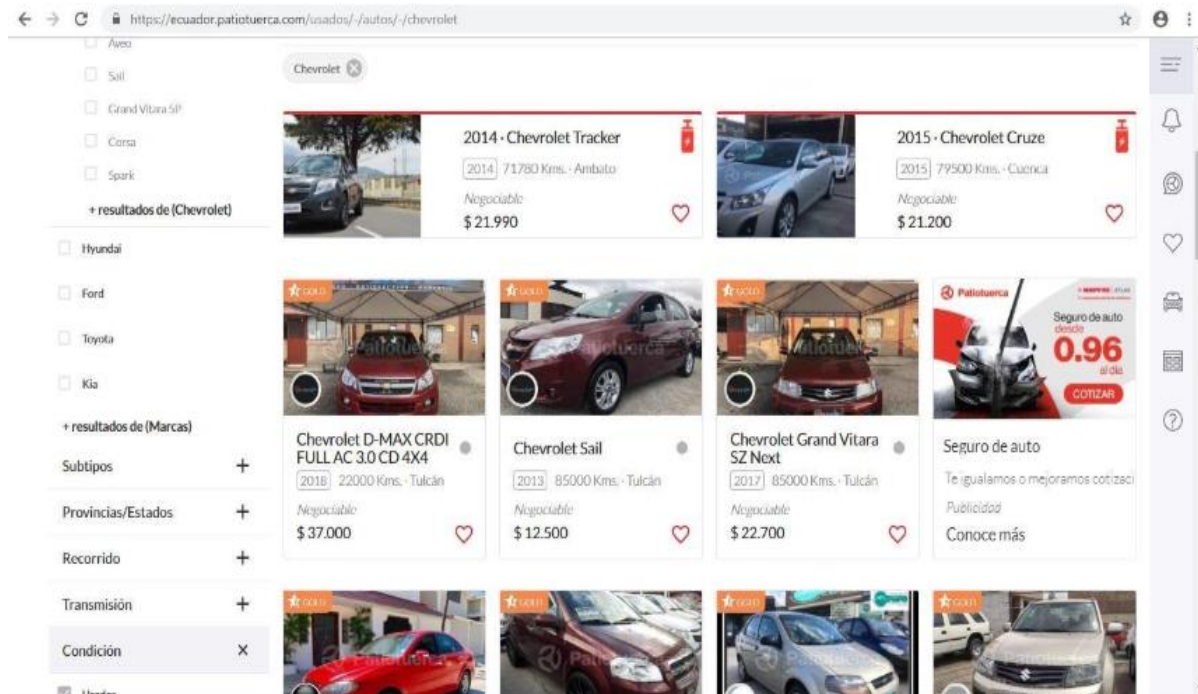
4.4.1 Selección de Fuente de Datos

Se eligió el sitio web <https://ecuador.patiotuerca.com/usados/-/autos>, por ser el más relevante en Ecuador en términos de tráfico y especialización en la venta de vehículos de segunda mano. La selección consideró factores como el SEO (posicionamiento en buscadores), especificidad del contenido y estructura HTML amigable para web scraping.

El eje principal de la empresa es el anuncio y comercialización de autos nuevos y seminuevos en las diferentes provincias del Ecuador, cualquier persona puede anunciar la venta de su vehículo en esta página, siguiendo los términos y condiciones de la misma. A continuación, se presenta la visualización de autos seminuevos tomando como ejemplo la búsqueda de un vehículo de marca Chevrolet.

Figura 3

Sitio web de Patio Tuerca Ecuador



Nota. Fuente: <https://ecuador.patiotuerca.com/usados/-/autos>

4.4.2 Web Scraping Automatizado

Para la captura de datos se realizó mediante la técnica para la extracción de información y datos de las webs. Actualmente existe varias técnicas para realizar el web scraping, existen softwares y extensiones de navegadores, solo por nombrar algunos en el mercado se encuentran (Web Scraper, Dexi.io, entre otros).

Se desarrolló un script en R utilizando el paquete rvest para extraer datos clave de cada publicación, tales como:

- Marca
- Modelo
- Año de fabricación
- Kilometraje
- Ciudad
- Precio

Este proceso se diseñó para ser escalable y replicable, permitiendo futuras actualizaciones de datos.

Es importante destacar que para realizar el scraping se debe tener en cuenta los siguientes parámetros:

- Establecer la página web, es decir se debe colocar el url, para ser más claros la dirección web, que comienza con http.
- Posteriormente se seleccionaron las etiquetas que se obtuvieron de la página en esta parte se obtiene el css.
- A continuación, se extrae el texto que se encuentra en la página.
- Por último, se convierte en una tabla con estructura de datos.
- A continuación, se explicará mediante un ejemplo obtenido de la “Guía para principiantes de Web Scraping con R (usando Rvest) con ejemplos prácticos” (Romero,2018).

Figura 4

Invocar la Librería

```
#invocar la librería
library("rvest")

#Asignamos la url
url <-
"https://www.amazon.com.mx/b/ref=s9_acsd_hfnv_hd_bw_bAcAgpX_ct_x_ct00
_w?_encoding=UTF8&node=9725407011&pf_rd_m=AVDBXBAVVSXLQ&pf_rd_s=merch
andised-search-
4&pf_rd_r=Q47SQG6GBGGECT1SXNZY&pf_rd_t=101&pf_rd_p=8cc61471-874f-
5f7f-96d7-e2b634466523&pf_rd_i=9725377011"
```

Nota. Código.

Figura 5

Establecer la url para webscraping

```
#Obtenemos código html de la página web
pagina_web <- read_html(url)
```

Nota. Código.

Figura 6

Establecer la Clase y los Nodos

```
#Asignamos la clase
css_producto <- "a.a-link-normal.s-access-detail-page.s-color-
twister-title-link.a-text-normal"

#Obtenemos el código html que contiene el nombre del producto
producto_html <- html_nodes(pagina_web,css_producto)
producto_texto <- html_text(producto_html)

#mostramos los datos, al final presionar enter
productos_texto
```

Nota. Código.

Figura 7

Obtención de los textos del código

```
#clase css del producto
css_precio <- "span.a-size-base.a-color-price.s-price.a-text-bold"
#obtenemos el contenido de la clase en código html
precio_html <- html_nodes(pagina_web,css_precio)
#limpiamos el código para obtener el texto
precio_texto <- html_text(precio_html)
```

Nota. Código.

Figura 8

Eliminación de caracteres en los datos

```
#Eliminamos el signo de peso
precio_limpio <- gsub("$","",precio_texto)
#Eliminamos la coma
precio_limpio <- gsub(",","",precio_limpio)
#Transformamos a numérico
precio_numerico <- as.numeric(precio_limpio)
```

Nota. Código.

4.4.3 Limpieza y Estructuración de Datos

Los datos brutos extraídos del sitio web fueron sometidos a un proceso de limpieza:

- Eliminación de duplicados
- Conversión de tipos de datos
- Corrección de valores nulos o inconsistentes
- Estandarización de nombres de variables y categorías
- Para ello se utilizaron funciones de los paquetes dplyr y plyr.

4.4.4 Análisis Estadístico Descriptivo

Se calcularon medidas de tendencia central, dispersión, frecuencias y distribuciones para conocer el comportamiento de las variables clave

4.4.5 Modelado Predictivo

Con el objetivo de predecir el precio de los vehículos usados, se aplicaron modelos de regresión lineal múltiple. Para ello se utilizó la función `lm()` de R, complementada con los paquetes:

- `lmtest`: para evaluar la significancia de los coeficientes
- `sandwich`: para corrección de heterocedasticidad
- `mctest`: para detectar multicolinealidad
- `leaps`: para seleccionar el mejor subconjunto de variables predictoras

4.4.6 Visualización de Resultados

El paquete `ggplot2` fue utilizado para crear representaciones gráficas que faciliten la comprensión de patrones y relaciones:

- Dispersión entre precio y kilometraje
- Comparaciones de precios por marcas y modelos
- Mapas de calor por ciudad

4.4.7 Evaluación y Validación del Modelo

Se realizaron pruebas de validación de los modelos construidos mediante análisis de residuos, ajuste de R^2 y error cuadrático medio (RMSE), con el objetivo de medir la eficacia predictiva y corregir posibles sesgos.

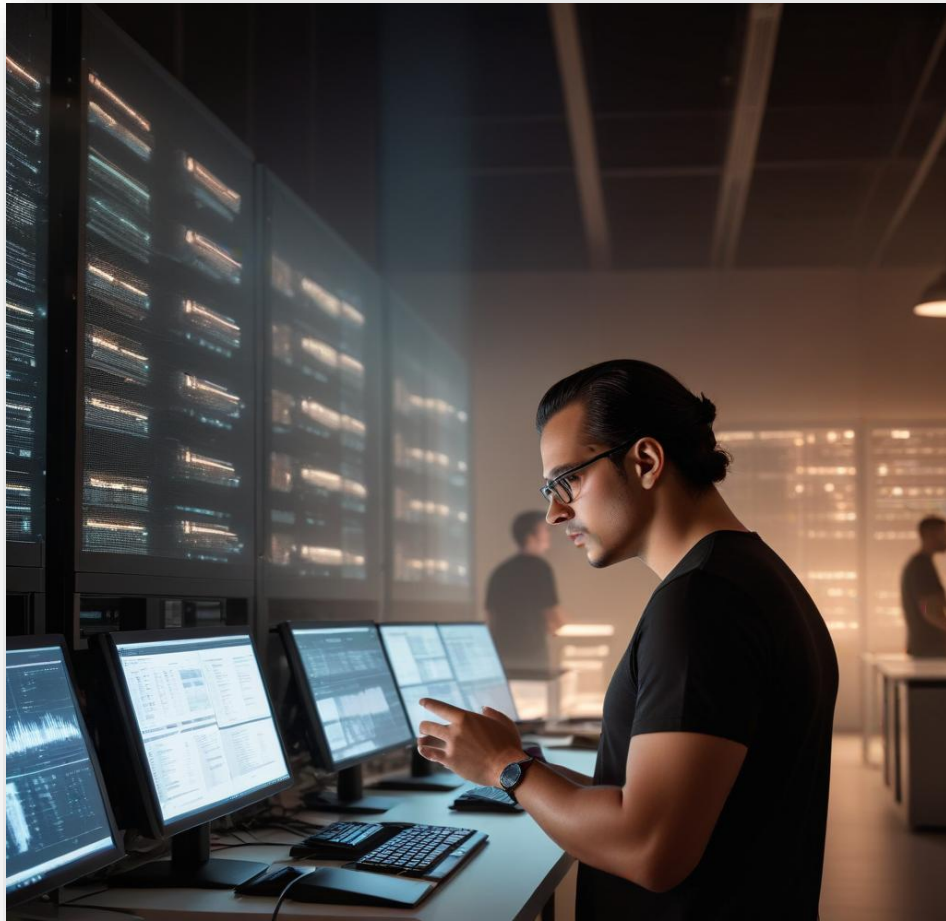
La metodología aplicada en este proyecto garantiza un análisis riguroso y sistemático del mercado automotor ecuatoriano mediante el uso de herramientas de código abierto que aseguran la transparencia y replicabilidad del estudio. Al integrar el entorno de programación R con librerías especializadas, se logró transformar información no estructurada de la web en una base de datos robusta y confiable. Este enfoque sistemático permitió transitar desde la obtención técnica de datos mediante web scraping hasta la interpretación de tendencias complejas, fundamentando cada etapa en principios estadísticos sólidos.

El desarrollo del proceso técnico, que incluyó la limpieza profunda de datos y la gestión de inconsistencias, fue esencial para la construcción de modelos predictivos de alta precisión. La aplicación de regresiones lineales múltiples, validadas mediante tests de heterocedasticidad y multicolinealidad, permitió identificar cómo variables críticas como el kilometraje, el año y la marca influyen directamente en la valoración de los vehículos. De este modo, la metodología no solo cumplió con una función descriptiva, sino que proporcionó una herramienta analítica capaz de estimar precios con fundamento científico.

Se evidencia que, la estructura metodológica adoptada permitió convertir grandes volúmenes de registros digitales en conocimiento estratégico para los diversos actores del sector. La combinación de análisis exploratorio, modelado estadístico y visualización de datos facilitó la identificación de patrones de oferta y demanda en las principales ciudades del país. Este marco de trabajo establece un precedente para futuras investigaciones en mercados dinámicos, demostrando que la integración de la ciencia de datos con el análisis económico real es fundamental para reducir la incertidumbre y mejorar la toma de decisiones en el comercio de vehículos usados.

Capítulo 5

Estudio y Gestión de los Datos



Fuente: [Free text-to-image generator](#) | [Freepik AI](#)

El proceso de construcción de una base de datos robusta para el mercado automotor ecuatoriano comienza con la transición técnica desde la interfaz visual de una página web hacia una estructura de datos organizada en R. Mediante el uso del lenguaje HTML y herramientas especializadas como el paquete rvest, es posible capturar de forma automática variables esenciales como la marca, el modelo, el año, el kilometraje y el precio de venta. Esta metodología de extracción no solo garantiza que cada publicación sea registrada de manera uniforme en columnas coherentes, sino que también democratiza el acceso al análisis de datos para usuarios que no poseen una formación profunda en ingeniería de software, permitiendo que la información pública se convierta en un activo estratégico para la investigación económica.

La solidez del análisis posterior depende directamente de la calidad de la depuración y el tratamiento estadístico aplicado a los más de 43,000 registros obtenidos. Durante esta fase, se identificaron y corrigieron inconsistencias críticas, como errores de digitación en precios y kilometrajes inverosímiles, además de eliminar registros de vehículos nuevos para centrar el estudio exclusivamente en el mercado de segunda mano. La integración de librerías adicionales como dplyr y ggplot2 facilitó no solo la limpieza de espacios y caracteres innecesarios, sino también la identificación visual de valores atípicos, asegurando que las medidas de tendencia central y la segmentación por marcas reflejen la realidad comercial del país, donde firmas como Chevrolet, Hyundai y Renault lideran el volumen de transacciones.

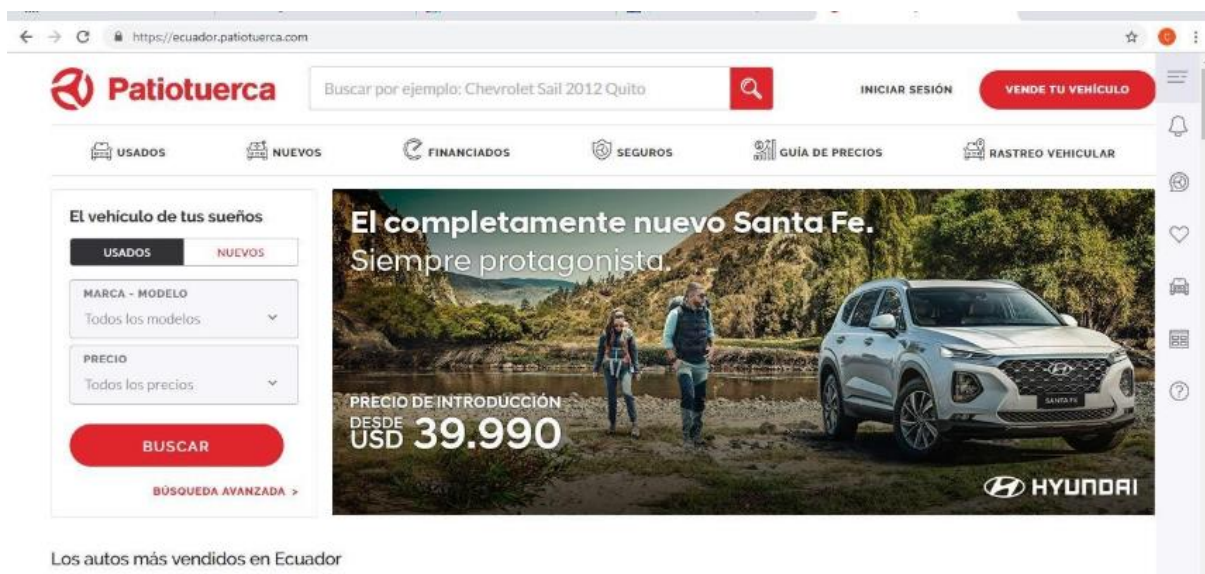
Finalmente, la aplicación de modelos de regresión lineal múltiple permitió transformar la base de datos en una herramienta predictiva capaz de estimar el valor de un vehículo con un nivel de explicación de hasta el 75%. Al validar supuestos estadísticos como la homocedasticidad y la ausencia de multicolinealidad, se garantizó la fiabilidad de las predicciones, demostrando que variables como el modelo y el kilometraje tienen un impacto diferenciado según la gama del automóvil. Este flujo metodológico, que concluye con la capacidad de predecir precios en función del recorrido y la antigüedad, proporciona una base científica para reducir la especulación y mejorar la transparencia en la comercialización de vehículos usados en todo el territorio nacional.

5.1 De la web al análisis: construyendo una base sólida

Al realizar la búsqueda de autos usados, muestra los vehículos con la información de marca del vehículo, el modelo, el año de su fabricación, la ciudad en la que se está comercializando, el kilometraje y su precio al cual se está comercializando, todas las publicaciones van acompañadas por sus imágenes. Toda la información mencionada anteriormente es la que se extrajo de los vehículos de manera automática para construir la base de datos y con la cual se trabajó posteriormente para el análisis y predicción de datos. Lo que se va extrajo de los vehículos que se encuentran en la página web, es información que constan en todas las publicaciones, por lo que no permitió tener numero distinto de observaciones en las diferentes columnas que conformaron la base de datos.

Figura 9

Página patio tuerca Ecuador



Nota. Fuente: <https://ecuador.patiotuerca.com/>

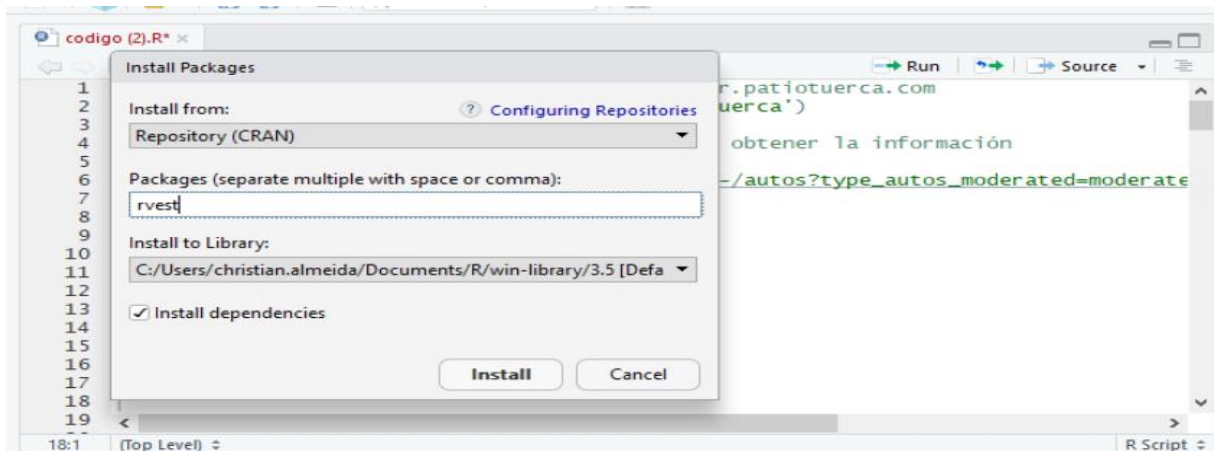
La página web está en formato Html, que es un lenguaje para la elaboración de páginas web las cuales significa HyperText Markup Language, traducción al lenguaje español significa lenguaje de marcas de hipertexto, es un formato simple y general que se encuentra entre los más utilizados para el desarrollo de páginas web.

Para realizar la extracción de la información en R, fue necesario instalar el paquete “rvest” el cual facilitó la obtención a personas que no son expertos en la programación o

que sus estudios no tienen afinidad con la carrera de sistemas o software. Después de instalar la librería como muestra la figura 10.

Figura 10

Instalación de la librería Rvest

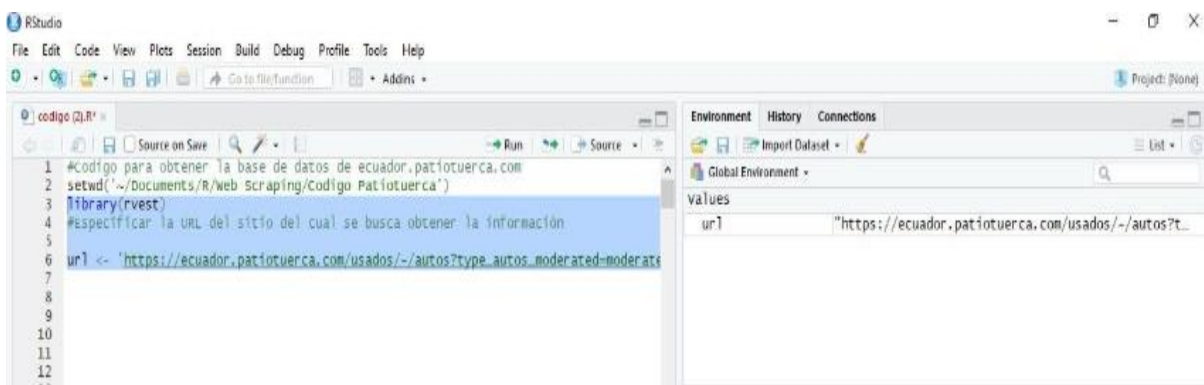


Nota. Fuente: <https://ecuador.patiotuerca.com/>

Posterior a la instalación del paquete “rvest”, hay que llamar al paquete para que se active, se realiza mediante el comando library(rvest) con esto se ingresa la url donde se encuentra la información extraída mediante el web scraping como se puede observar en la figura. Se puede evidenciar que el código ha sido correctamente ejecutado y que se cuenta con la url del sitio web en este caso de patio tuerca Ecuador como una variable en la parte superior derecha que se denomina url.

Figura 11

Verificación de paquete y pagina web



Nota. Fuente: <https://ecuador.patiotuerca.com/>

Después de tener cargado la página web en R, se procede con la extracción de las variables de los automóviles, se empezará por la marca de vehículo, el cual, mediante el marcador establecido en la página web, 'card-info.card-content .module.tittle h2', se podrá realizar el código, además a la variable contiene algunos caracteres que deben eliminarse como espacios y puntos; por último, se le transforma en una variable de texto, mediante la opción carácter.

Figura 12

Verificación de paquete y pagina web

```
url <- 'https://ecuador.patiotuerca.com/usados/-/autos?type_autos_moderated=moderated'
#Leer el código HTML del website
webpage <- read_html(url)
#Datos de marca del auto
badge_data_html <- html_nodes(webpage, '.card-info.card-content .module.tittle h2')
badge_data <- html_text(badge_data_html)
head(badge_data)
badge_data <- gsub("\n", "", badge_data)
badge_data <- gsub(" ", "", badge_data)
head(badge_data)
badge_data <- gsub(".", "", badge_data)
badge_data <- as.character(badge_data)
```

Nota. Código fuente.

Luego de obtener la primera variable de los vehículos, continua la obtención de las siguientes variables modelo de los automóviles, año de fabricación o procedencia, kilometraje que tiene y por último el precio que se está comercializando. Todas las variables deben ser depuradas ya que contienen espacios o algún carácter que no es necesario, la variable de modelo de vehículo se la transforma a texto al igual que la de marca, las otras variables en cambio se las convierte a numéricas, como se puede observar en el siguiente código.

Figura 13

Código de fuente 2

```
#Datos del modelo del auto
model_data_html <- html_nodes(webpage, '.card-info.card-content .module.tittle h2')
model_data <- html_text(model_data_html)
head(model_data)
model_data <- gsub("\n", "", model_data)
model_data <- gsub("'", "", model_data)
head(model_data)
model_data <- word(model_data, -1)
model_data <- as.character(model_data)

#Datos del año del auto
year_data_html <- html_nodes(webpage, '.latam-secondary-text .year')
year_data <- html_text(year_data_html)
head(year_data)
year_data <- as.numeric(year_data)
head(year_data)

#Datos del kilometraje del auto
km_data_html <- html_nodes(webpage, '.latam-secondary-text:nth-child(4)')
km_data <- html_text(km_data_html)
head(km_data)
km_data <- gsub("\n", "", km_data)
km_data <- gsub(" ", "", km_data)
km_data <- gsub(".", "", km_data)
km_data <- gsub(".*", "", km_data)
km_data <- as.numeric(km_data)
head(km_data)

#Datos del precio del auto
price_data_html <- html_nodes(webpage, '.left.share-value')
price_data <- html_text(price_data_html)
head(price_data)
price_data <- gsub("\n", "", price_data)
price_data <- gsub("'", "", price_data)
price_data <- substring(price_data, 2, 7)
```

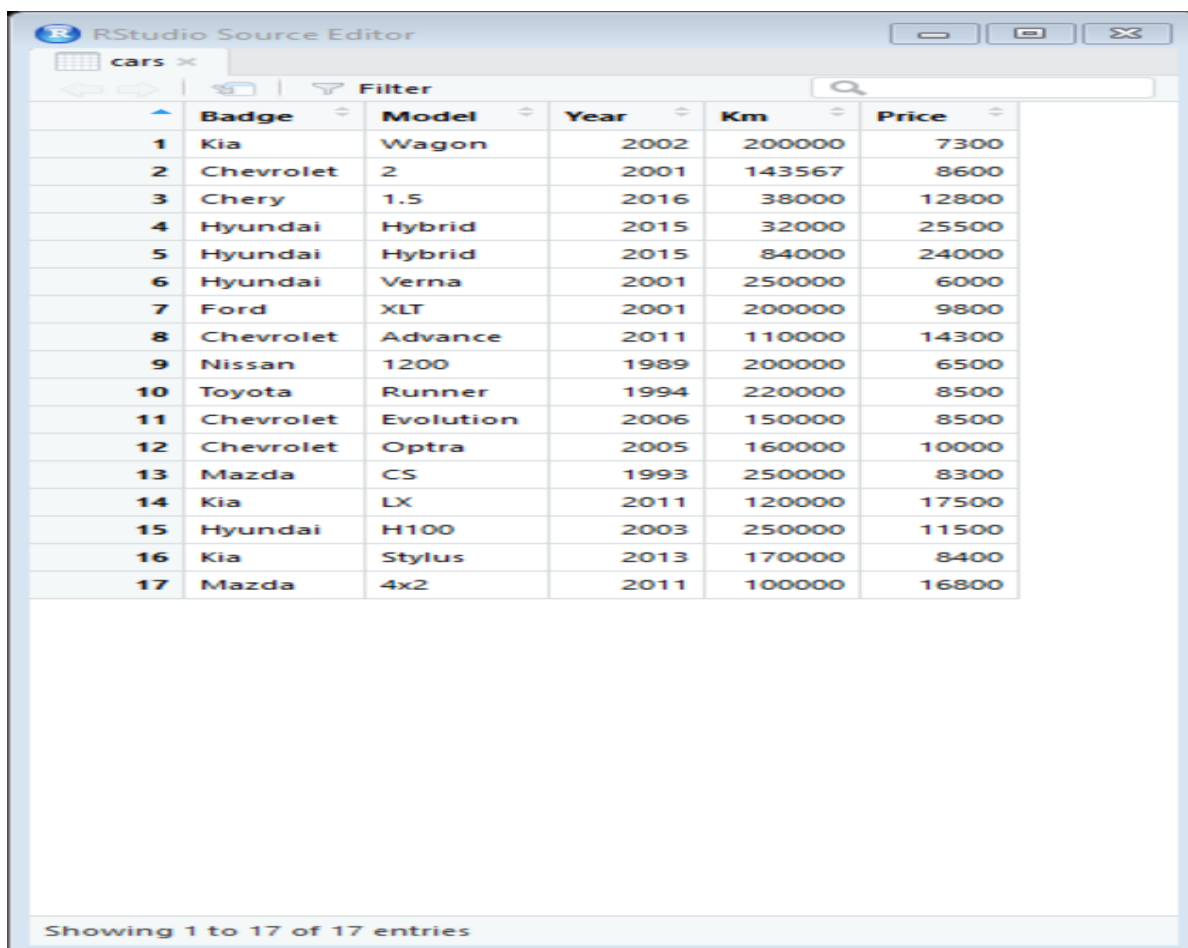
Nota. Fuente: <https://ecuador.patiotuerca.com/>

El proceso de extracción automatizada permitió capturar de manera estructurada la información contenida en la interfaz inicial del portal, transformando el código HTML en una matriz de datos organizada. Como se observa en la figura 14, el resultado de este procedimiento refleja la consolidación de variables críticas como la marca, el modelo, el año y el precio de cada vehículo listado. Esta etapa es fundamental para el éxito del proyecto, ya que garantiza que la materia prima del análisis sea veraz y represente fielmente las condiciones reales de oferta en el mercado digital ecuatoriano al momento de la captura.

La obtención de estos registros primarios constituye el primer paso hacia una comprensión profunda de las tendencias comerciales, permitiendo identificar a simple vista la densidad de publicaciones y la variedad de la flota disponible. Al sistematizar esta captura, se elimina el error humano asociado a la recolección manual y se establece un flujo de trabajo escalable que puede aplicarse a miles de anuncios adicionales. De esta forma, los datos presentados no son solo una lista de vehículos, sino la base técnica necesaria para alimentar los modelos estadísticos que determinarán los patrones de valoración y comportamiento en el sector de usados.

Figura 14

Datos de la primera página de la web.



	Badge	Model	Year	Km	Price
1	Kia	Wagon	2002	200000	7300
2	Chevrolet	2	2001	143567	8600
3	Chery	1.5	2016	38000	12800
4	Hyundai	Hybrid	2015	32000	25500
5	Hyundai	Hybrid	2015	84000	24000
6	Hyundai	Verna	2001	250000	6000
7	Ford	XLT	2001	200000	9800
8	Chevrolet	Advance	2011	110000	14300
9	Nissan	1200	1989	200000	6500
10	Toyota	Runner	1994	220000	8500
11	Chevrolet	Evolution	2006	150000	8500
12	Chevrolet	Optra	2005	160000	10000
13	Mazda	CS	1993	250000	8300
14	Kia	LX	2011	120000	17500
15	Hyundai	H100	2003	250000	11500
16	Kia	Stylus	2013	170000	8400
17	Mazda	4x2	2011	100000	16800

Showing 1 to 17 of 17 entries

Nota. Fuente: <https://ecuador.patiotuerca.com/>

Como se puede observar la ejecución ha sido correcta y se ha tenido en la base de vehículos de la primera página del sitio web de patio tuerca Ecuador, para obtener la base completa del sitio web, primero se debe conocer cuál es el número de la última página, la cual es la 2.179.

Lo segundo que se debe considerar es realizar el código para que continúe con la obtención de los datos de las siguientes páginas, el cual se puede realizar mediante un “for”, con el que permite realizar repeticiones para diferentes procesos por lo que es mejor alternativa.

Figura 15

Extracción de datos

```
For (i in 1:2179){
  baseURL <- 'https://ecuador.patiotuerca.com/usados/-/autos?type_autos_moderated=mod
  completeURL <- paste(baseURL,i)
  webpage <- read_html(completeURL)
  #Datos de marca del auto
  badge_data_html <- html_nodes(webpage, '.card-info.card-content .module.tittle h2')
  badge_data <- html_text(badge_data_html)
  head(badge_data)
  badge_data <- gsub('\n', '', badge_data)
  badge_data <- gsub(' ', '', badge_data)
  head(badge_data)
  badge_data <- gsub(" .*", "", badge_data)
  badge_data <- as.character(badge_data)

  #Datos del modelo del auto
  model_data_html <- html_nodes(webpage, '.card-info.card-content .module.tittle h2')
  model_data <- html_text(model_data_html)
  head(model_data)
  model_data <- gsub('\n', '', model_data)
  model_data <- gsub(' ', '', model_data)
  head(model_data)

  model_data <- word(model_data, -1)
  model_data <- as.character(model_data)

  #Datos del año del auto
  year_data_html <- html_nodes(webpage, '.latam-secondary-text .year')
  year_data <- html_text(year_data_html)
  head(year_data)
  year_data <- as.numeric(year_data)
  head(year_data)
```

Nota. Fuente: <https://ecuador.patiotuerca.com/>

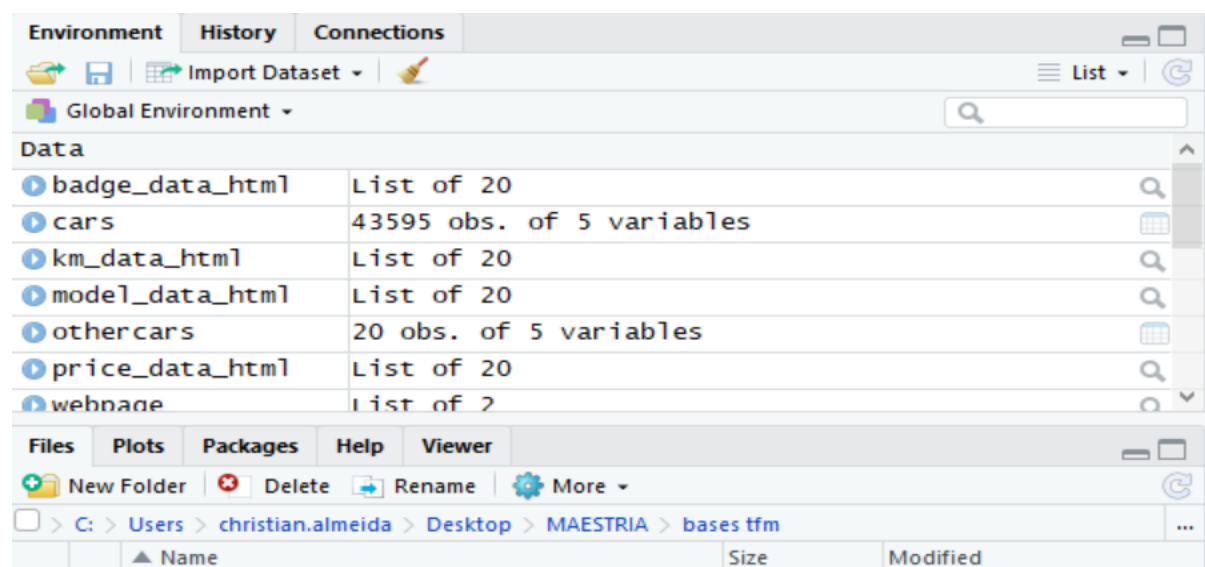
Con este proceso se ha completado el script para realizar el web scraping y conseguir los datos de manera automática, la descarga de datos toma entre 35- 50

minutos dependiendo del procesador del computador y de la velocidad y tráfico del internet, tomando en cuenta que son 2.179 paginas es mínimo con el tiempo que se tomaría al realizarlo manualmente y los errores que existirían en la digitación.

Como se observa en la figura 16, se puede ver que el código ha terminado y que la base, que tiene como nombre “cars”, la cual está ubicada en la parte superior derecha de Rstudio, contiene 43595 registros, que en este proceso se interpretaría que existen esa cantidad de registros o de publicaciones de venta de vehículos que se encuentran en la página de patio tuerca en que se están comercializando en todo el Ecuador.

Figura 16

Base con número de observaciones y columnas

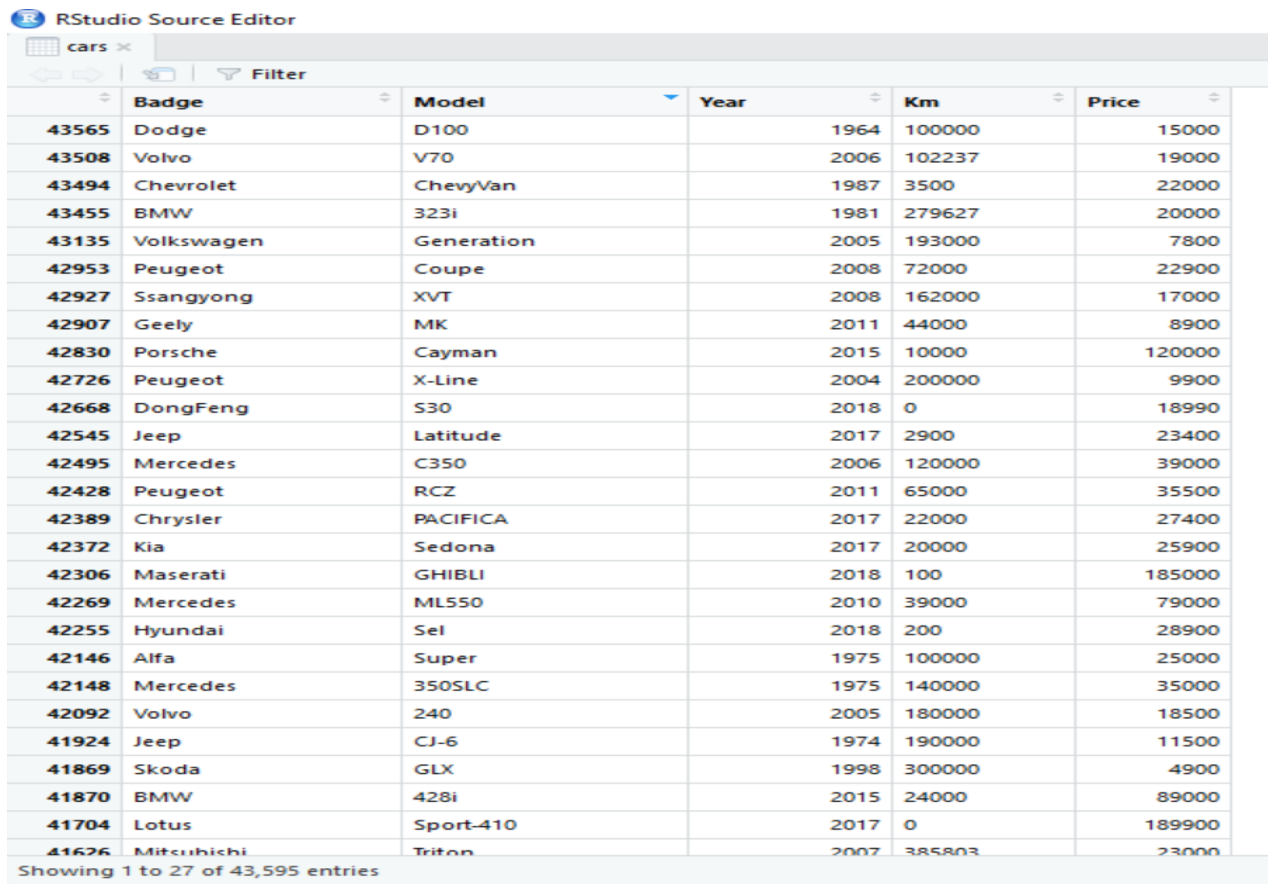


Nota. Fuente: <https://ecuador.patiotuerca.com/>

Si se da click en la palabra cars se va a visualizar la base de datos obtenida, para verificación se puede observar en la figura 17 los datos sobre los 43.000 registros para verificación que la base esta descargada y los datos en cada una de las columnas.

Figura 17

RStudio



RStudio Source Editor

cars x

Filter

	Badge	Model	Year	Km	Price
43565	Dodge	D100	1964	100000	15000
43508	Volvo	V70	2006	102237	19000
43494	Chevrolet	ChevyVan	1987	3500	22000
43455	BMW	323i	1981	279627	20000
43135	Volkswagen	Generation	2005	193000	7800
42953	Peugeot	Coupe	2008	72000	22900
42927	Ssangyong	XVT	2008	162000	17000
42907	Geely	MK	2011	44000	8900
42830	Porsche	Cayman	2015	10000	120000
42726	Peugeot	X-Line	2004	200000	9900
42668	DongFeng	S30	2018	0	18990
42545	Jeep	Latitude	2017	2900	23400
42495	Mercedes	C350	2006	120000	39000
42428	Peugeot	RCZ	2011	65000	35500
42389	Chrysler	PACIFICA	2017	22000	27400
42372	Kia	Sedona	2017	20000	25900
42306	Maserati	GHIBLI	2018	100	185000
42269	Mercedes	ML550	2010	39000	79000
42255	Hyundai	Sel	2018	200	28900
42146	Alfa	Super	1975	100000	25000
42148	Mercedes	350SLC	1975	140000	35000
42092	Volvo	240	2005	180000	18500
41924	Jeep	CJ-6	1974	190000	11500
41869	Skoda	GLX	1998	300000	4900
41870	BMW	428i	2015	24000	89000
41704	Lotus	Sport-410	2017	0	189900
41626	Mitsubishi	Triton	2007	385803	23000

Showing 1 to 27 of 43,595 entries

Nota. Base descarga completa.

Como se visualiza en la figura 17, la base de datos con la que se trabajará, a continuación, se describirá las columnas:

- Columna 1: marca del vehículo.
- Columna 2: modelo del vehículo.
- Columna 3: año de fabricación.
- Columna 4: recorrido del automóvil la medida en que se encuentra es en kilómetros.
- Columna 5: precio de comercialización, la unidad del precio se establece en dólares americanos.

Además de la librería *rvest*, que se utilizó para realizar la obtención de datos desde la página de internet, también se utilizó las librerías *stringr*, *dplyr*, *plyr*, *lmtest*, *sándwich*,

mctest y ggplot2. La librería “strigr” fue descargada y utilizada para quitar espacios y algunos caracteres que al momento de obtener los datos aparecen y que no son necesarios o que se deben retirar para depurar la base.

El paquete plyr en cambio se descargó para utilizar la función rbind.fill, la cual permite agregar las columnas con las variables a la base de datos, a pesar de no contar con todos los registros en la columna. Además de estas funciones de las librerías mencionadas, también se utilizará las funciones básicas que se encuentran en la consola de R, como as.character o as.numeric para transformar a las variables de la columna tanto a texto o como una variable numérica respectivamente.

Una vez consolidada la base de datos inicial, resulta imperativo realizar una fase de inspección visual y auditoría de calidad para garantizar la integridad de la información. Este proceso de exploración permite detectar posibles inconsistencias originadas durante el ingreso manual de los datos por parte de los usuarios en la plataforma. La identificación de valores atípicos o errores de digitación es un paso crítico, ya que la presencia de registros con precios desproporcionados respecto a las características técnicas del vehículo podría sesgar significativamente los promedios y distorsionar los resultados de los modelos predictivos.

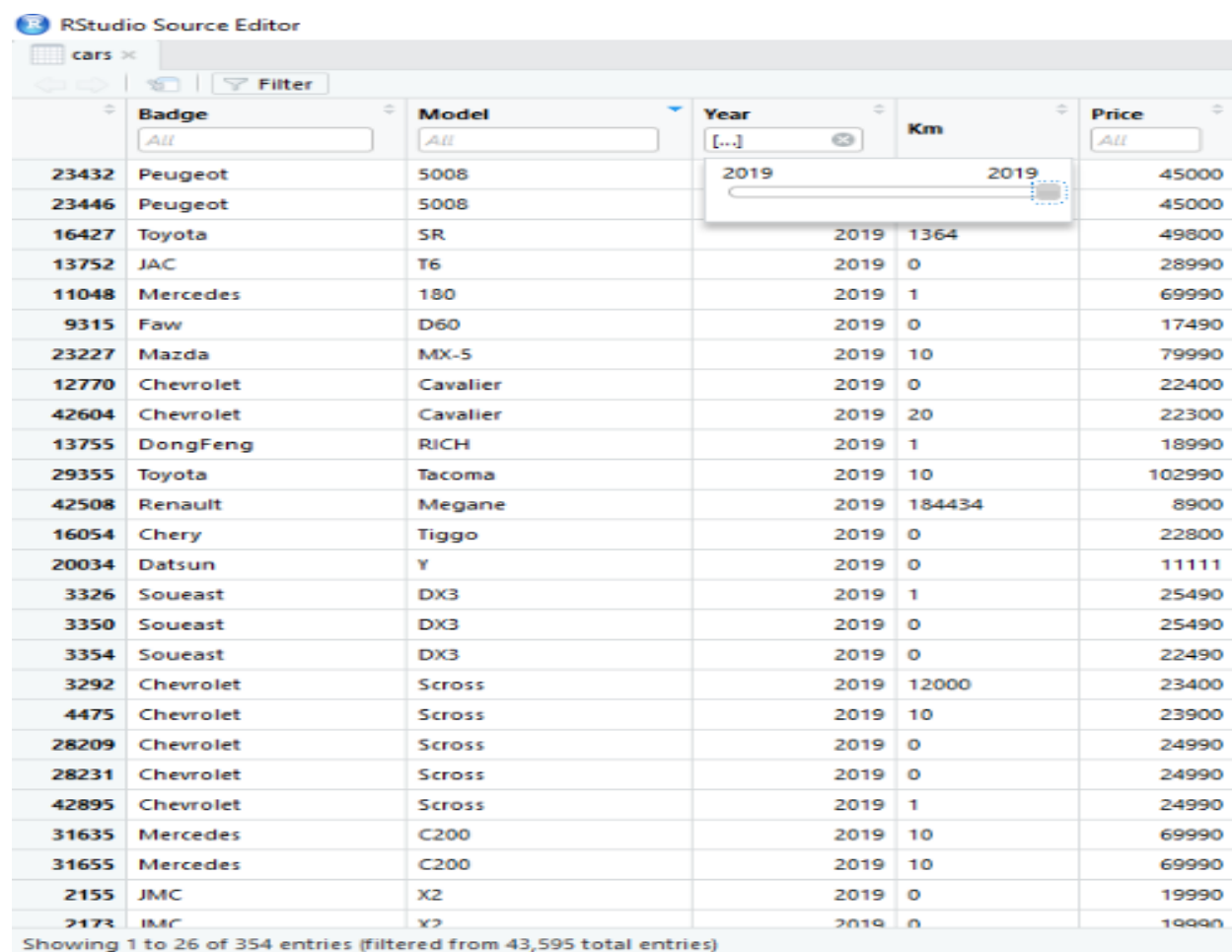
En este contexto, la organización de los registros bajo criterios específicos permite aislar aquellos casos donde la relación entre marca, modelo y valor de mercado no guarda una coherencia lógica. La detección de estos puntos anómalos, que a menudo superan los quinientos mil dólares en categorías de vehículos convencionales, evidencia la necesidad de aplicar filtros de limpieza antes de proceder con el análisis estadístico descriptivo. La siguiente representación visual ilustra este procedimiento de clasificación, exponiendo los registros que han sido marcados para su revisión o eliminación definitiva del conjunto de datos.

A continuación de tener la base se extenderá el análisis de las variables que componen la base para depurar. En la figura 18 se puede observar los outliers o valores atípicos que se debe retirar ya que se analiza el precio en el eje de las ordenadas y el kilometraje en el eje de las abscisas y el año en el que se fabricó, lo que se obtiene como resultado es que existen vehículos con un precio elevado a pesar de tener un año antiguo por lo que puede ser por digitación errónea del usuario. De igual manera se retiraron los

automóviles del año 2019 ya que el análisis se centra en vehículos de segunda mano, como muestra en la figura.

Figura 18

Identificación de datos



	Badge	Model	Year	Km	Price
23432	Peugeot	5008	2019		45000
23446	Peugeot	5008	2019		45000
16427	Toyota	SR	2019	1364	49800
13752	JAC	T6	2019	0	28990
11048	Mercedes	180	2019	1	69990
9315	Faw	D60	2019	0	17490
23227	Mazda	MX-5	2019	10	79990
12770	Chevrolet	Cavalier	2019	0	22400
42604	Chevrolet	Cavalier	2019	20	22300
13755	DongFeng	RICH	2019	1	18990
29355	Toyota	Tacoma	2019	10	102990
42508	Renault	Megane	2019	184434	8900
16054	Chery	Tiggo	2019	0	22800
20034	Datsun	Y	2019	0	11111
3326	Soueast	DX3	2019	1	25490
3350	Soueast	DX3	2019	0	25490
3354	Soueast	DX3	2019	0	22490
3292	Chevrolet	Scross	2019	12000	23400
4475	Chevrolet	Scross	2019	10	23900
28209	Chevrolet	Scross	2019	0	24990
28231	Chevrolet	Scross	2019	0	24990
42895	Chevrolet	Scross	2019	1	24990
31635	Mercedes	C200	2019	10	69990
31655	Mercedes	C200	2019	10	69990
2155	JMC	X2	2019	0	19990
2173	JMC	X2	2019	0	19990

Showing 1 to 26 of 354 entries (filtered from 43,595 total entries)

Nota. Base descarga

Para revisar los registros que tienen un precio extremadamente elevado se procedió a ordenar el precio en orden descendente para verificar si existen marcas y modelos que pueden tener precios elevados de comercialización y que el valor ha sido ingresado de manera errónea por los comerciantes y que para el análisis hay que eliminarlos, como se pudo observar en la figura hay vehículos con sus precios que salta a primera vista, como por ejemplo que tienen un precio superior a los 500.000 y que la marca y el modelo no están acorde.

Figura 19

Orden descendente de la variable precio.

Filter				
Badge	Model	Year	Km	Price
Volkswagen	Gol	2005	298000	978570
Changan	Pickup	2003	15000	900000
Kawasaki	600	2006	40000	850000
Chevrolet	Cruze	1981	6000000	800000
Mitsubishi	L200	1999	1	700000
Mini	One	1965	1000	500000
Maserati	S	2018	4602	260000
Mercedes	AMG	2015	20300	255000
Asia	Topic	2015	325345	250000
Maserati	S	2018	1	249000
Asia	Topic	2014	333	222000
Land	AutoBiography	2014	24500	220000
Mercedes	43	2018	1	208000
Lotus	400	2017	0	189900
Lotus	Sport-410	2017	0	189900
Maserati	Levante	2018	8400	189000
Mercedes	400	2018	1	187000
Maserati	Levante	2018	15	185000
Maserati	GHIBLI	2018	100	185000
Mercedes	AMG	2014	26000	185000
Porsche	911-Turbo	2007	20000	180000
Land	Velar	2018	13	179900
Chevrolet	Sail	2018	30000	169000
Lotus	380	2017	3800	169000
Mercedes	400	2017	4000	165000
BMW	X6	2018	23000	160000
Land	Velar	2018	1	155000

Nota. Base descarga

A continuación, se hizo un análisis de la variable kilometraje, ordenando de manera descendente y se encontró que, a pesar de no tener varios años el vehículo, el kilometraje es extremadamente alto llegando a tener 100000.000 de kilómetros de recorrido, lo cual también se concluye que es un error de ingreso de la variable en la página web por parte de la persona que comercializa.

Figura 23

Orden descendente de la variable kilometraje

Filter				
Badge	Model	Year	Km	Price
Chevrolet	Silverado	2000	10000000	18000
Chevrolet	Silverado	2000	10000000	18000
Ford	F150	2010	10000000	17000
Toyota	Cruiser	1980	10000000	13000
Volkswagen	Gol	2011	10000000	12800
Chery	Tiggo	2012	10000000	12000
Chevrolet	Aveo	2012	10000000	11800
Kia	Rio	2010	10000000	10200
Chevrolet	Malibu	1978	10000000	10000
Chevrolet	Malibu	1978	10000000	10000
Chevrolet	CS	1990	10000000	8000
Chevrolet	Corsa	2006	10000000	7700
Kia	R	2011	10000000	7300
Chevrolet	Corsa	2005	10000000	7200
Chevrolet	CD	1994	10000000	7200
Chevrolet	Corsa	2001	10000000	7100
Chevrolet	Spark	2005	10000000	6800
Chevrolet	2	1998	10000000	6700
Ford	CD	1993	10000000	6700
Ford	CD	1993	10000000	6700
Ford	CD	1993	10000000	6700
Chevrolet	1	2001	10000000	6500
Chevrolet	1	2001	10000000	6500
Chevrolet	3P	1992	10000000	6200
Chevrolet	Spark	2006	10000000	6000
Chevrolet	Spark	2006	10000000	6000
Nissan	Sentra	2003	10000000	6000

Nota. Base de datos.

En la siguiente figura se realiza la ordenación de manera ascendente de la variable años encontrando vehículos que se comercializan, que tienen una fabricación de hace más de 100 años, lo cual el mantener estos registros dificulta el análisis, por lo que

realizaremos de manera gráfica la identificación de los outliers, para tomar una decisión para eliminar algunos registros que no son relevantes para el análisis.

Figura 24

Orden ascendente de la variable años

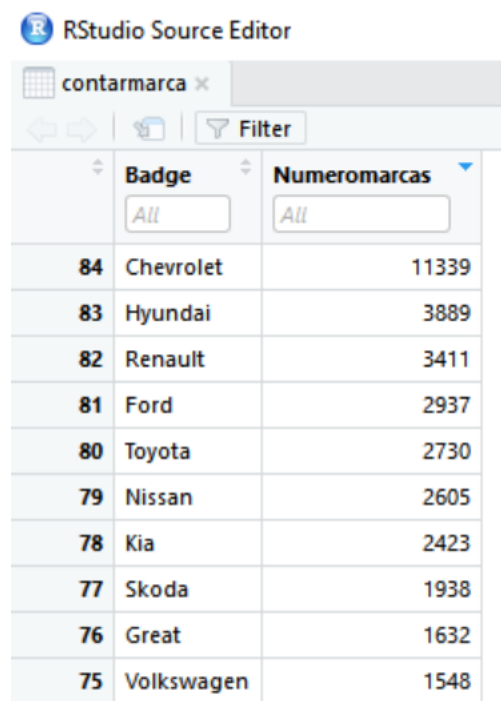
Filter				
Badge	Model	Year	Km	Price
All	All	All		All
Chevrolet	Spark	1900	0	0
Ford	31	1928	125000	56000
Jeep	Willys	1942	1000	15000
Jeep	Willys	1950	689000	12900
Chevrolet	Apache	1954	68000	25000
Chevrolet	Apache	1954	65000	30000
Ford	F100	1955	350593	31000
Chevrolet	Apache	1955	100	15000
Volkswagen	Escarabajo	1959	100000	14800
Jeep	Willys	1960	2000	28000
Volkswagen	Kombi	1962	1000	15000
Jeep	Willys	1964	259000	14000
Dodge	D100	1964	100000	15000
Oldsmobile	Cutlass	1964	31798	5990
Ford	Mustang	1965	111000	26500
Jeep	Willys	1965	90000	7000
Volkswagen	Escarabajo	1965	74632	15000
Mini	One	1965	1000	500000
Volkswagen	Escarabajo	1965	10	6900
Ford	Taurus	1965	1	17777
Jeep	Willys	1966	80000	18000
Mini	Van	1966	21999	12000
Volkswagen	Escarabajo	1967	270000	5900
Toyota	Stout	1967	111111	5500
Mercedes	250	1967	1	26990
Toyota	Crown	1968	160000	16900

Nota. Base descarga.

Posterior de retirar los vehículos del año y los datos atípicos (outliers), se empezó a realizar el análisis de la comercialización de vehículos de segunda mano. A continuación, se analizaron los datos, por ejemplo el análisis permitió conocer cuáles son las marcas que más se están comercializando en el país, por lo que se pudo realizar un *top ten* para identificar las marcas que más se comercializan con su respectiva cantidad, como se observa en la figura 34, donde se ubica en primer lugar la marca Chevrolet, en segundo lugar, se ubica muy por debajo la marca coreana Hyundai y muy de cerca en tercer puesto la marca Francesa Renault. Se puede concluir que estas son las marcas que más se comercializan en el Ecuador y que por lo tanto el adquirir ese vehículo va a ser más fácil para volver a venderlo, además es más probable que estas marcas no tengan escases de repuestos al momento de necesitarlos.

Figura 25

Número de automóviles agrupados por marca.



The image shows a screenshot of the RStudio Source Editor. At the top, there is a tab labeled 'contarmarca'. Below the tab, there is a toolbar with a 'Filter' button. The main area displays a table with the following data:

	Badge	Numeromarcas
84	Chevrolet	11339
83	Hyundai	3889
82	Renault	3411
81	Ford	2937
80	Toyota	2730
79	Nissan	2605
78	Kia	2423
77	Skoda	1938
76	Great	1632
75	Volkswagen	1548

Nota. Base de datos

También mediante el comando en RStudio se puede obtener diferentes medidas estadísticas de la base de datos como el máximo y mínimo de cada columna, de igual manera la media o promedio, la mediana y los cuartiles. Por ejemplo, se puede observar

en la figura siguiente, que en el precio va desde 1.000 dólares americanos como el valor mínimo que se comercializa un vehículo de segunda mano hasta los 145.000 dólares, la media y mediana de la columna Price 17.495 y 14.700 respectivamente, el primer cuartil de la variable Price es de 10.900 y el tercero cuartil es de 21.000 dólares.

Figura 26

Resumen estadístico de la base CARS

```
> summary(CARS)
      X      Badge      Model      Year
Min.   : 1  Length:42977  Length:42977  Min.   :1960
1st Qu.:10880 Class :character  Class :character  1st Qu.:2007
Median :21722 Mode  :character  Mode  :character  Median :2011
Mean   :21737
3rd Qu.:32556
Max.   :43595

      Km      Price      Numeromarcas
Min.   : 0  Min.   : 1000  Min.   : 10
1st Qu.: 62000 1st Qu.: 10900 1st Qu.: 1632
Median : 107000 Median : 14700 Median : 2730
Mean   : 120946 Mean   : 17495 Mean   : 4613
3rd Qu.: 165000 3rd Qu.: 21000 3rd Qu.: 11339
Max.   :1000000 Max.   :145000 Max.   :11339
> |
```

Nota. Base de datos

A continuación, se realizará un análisis de los vehículos para conocer cómo cambian su precio durante el transcurso de los años, cabe mencionar que este análisis se seguirá empleando el software R y se realizará con la librería ggplot2 para su presentación gráficamente.

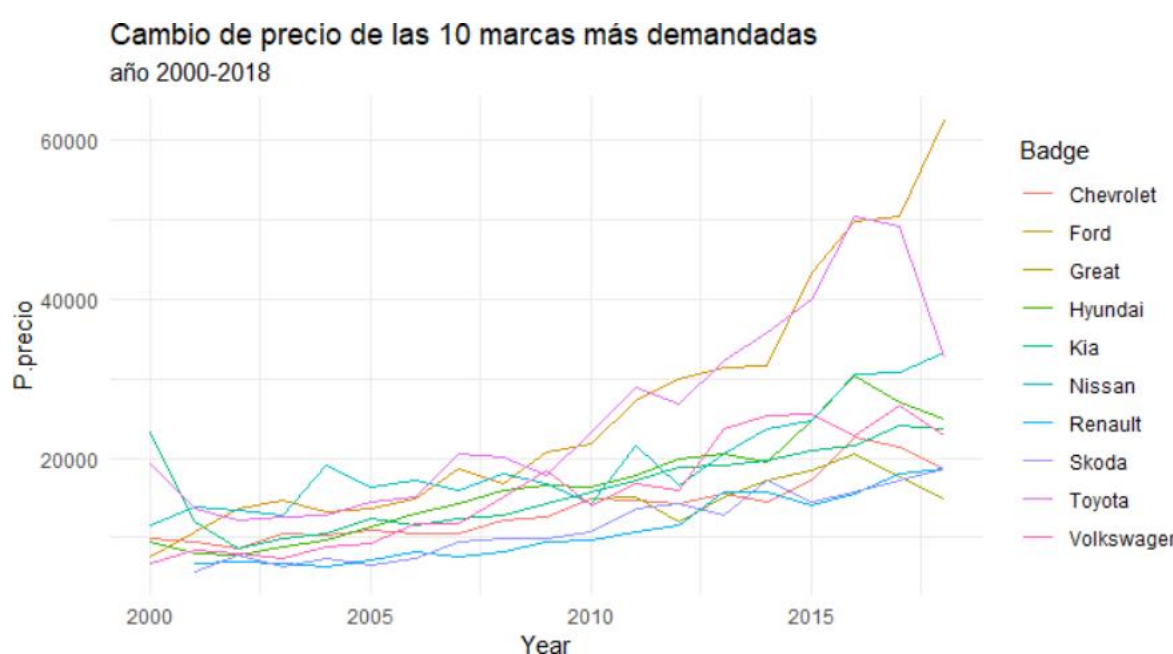
Para el análisis de las marcas que más varían el precio de un año a otro, se consideró primero las marcas más representativas y que más se comercializan en la base obtenida, también se consideró para el análisis, que se realice a partir de año 2000 en adelante principalmente por dos circunstancias, la primera es debido a que, en los años anteriores al seleccionado, no hay una gran cantidad de vehículos representativa en cada marca y también porque el año 2000, es un año importante para el Ecuador ya que cambio de moneda en ese entonces, paso de la moneda nacional sucre a la moneda norteamericana dólar, por lo que se cambió a la dolarización y el precio de vehículos y

su respectiva demanda también tuvo un cambio a partir de esta decisión política-económica.

En las figuras 27 y 28, se puede observar el precio promedio de las marcas más demandadas desde el año 2000, para conocer que vehículos tienen mayor impacto en el precio durante el tiempo. Como resultado se puede observar detalladamente y de manera individualizada de cada marca, en las mini gráficas de la figura 30, la marca que mayor cambia incrementando su precio es la marca Ford.

Figura 27

Variación de precio de las marcas más demandadas

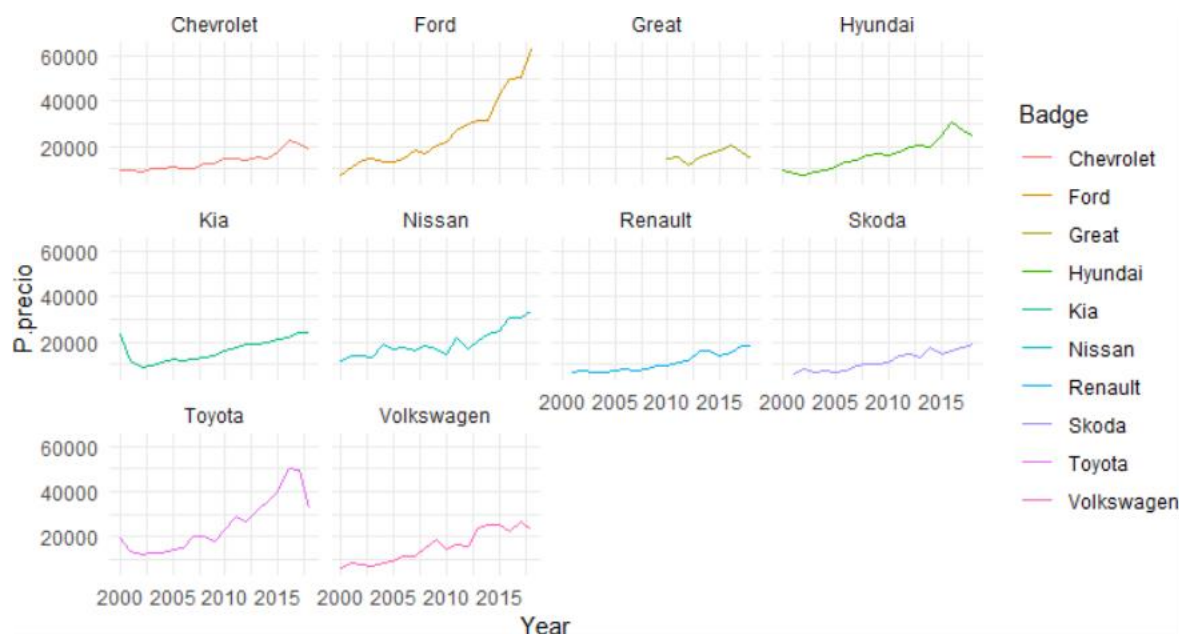


Nota. Cambio de precio

También llama la atención marcas de vehículos que su precio en lugar de ir en aumento en los últimos años va disminuyendo, a pesar que por su año se les considera menos antiguo, esto se debe a que en los últimos años Ecuador ha tenido diferentes políticas respecto a vehículos, por ejemplo, bajo los aranceles de vehículos provenientes de Europa y también a vehículos que cuiden el medio ambiente como los autos eléctricos e híbridos como lo tienen marcas Hyundai con sus modelos ionic y sonata, Toyota con el modelo Prius (el auto hibrido más antiguo en el Ecuador) y Kia con su modelo Soul y Niro.

Figura 27

Variación de precio de las marcas más demandadas



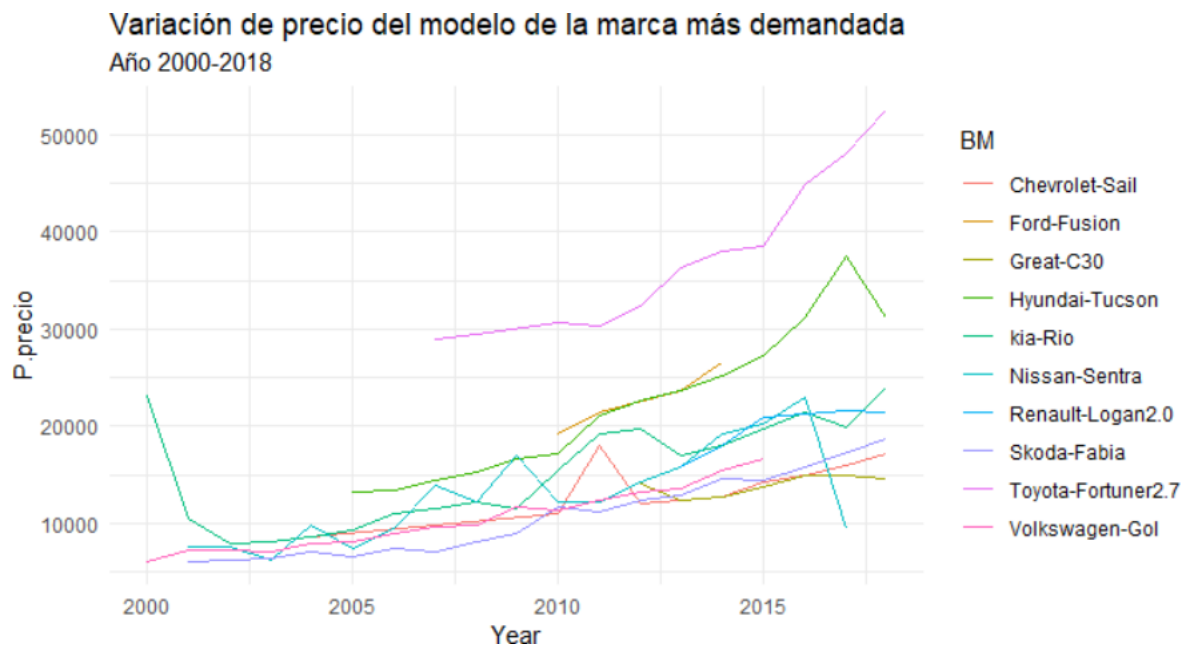
Nota. Marcas demandadas

También se efectuó un análisis seleccionando de estas 10 marcas (top ten) más demandas, el modelo más comercializado en la base dentro de estas, para identificar en primera instancia cual es el modelo que más se comercializa dentro del periodo 2000 y 2018, además si tiene el mismo comportamiento de su promedio de precio del modelo como su marca.

Como se observa, el comportamiento de algunos modelos es similar al comportamiento de la marca como Hyundai y su vehículo Tucson o la marca Great Wall y su modelo C30. Por el contrario, también hay modelos cuya tendencia de precio en el transcurso de los años es distinta a su marca sobre todo Nissan y su modelo Sentra o el modelo Fortuner y su marca Toyota, en la cual la tendencia es contraria entre marca y modelo.

Figura 28

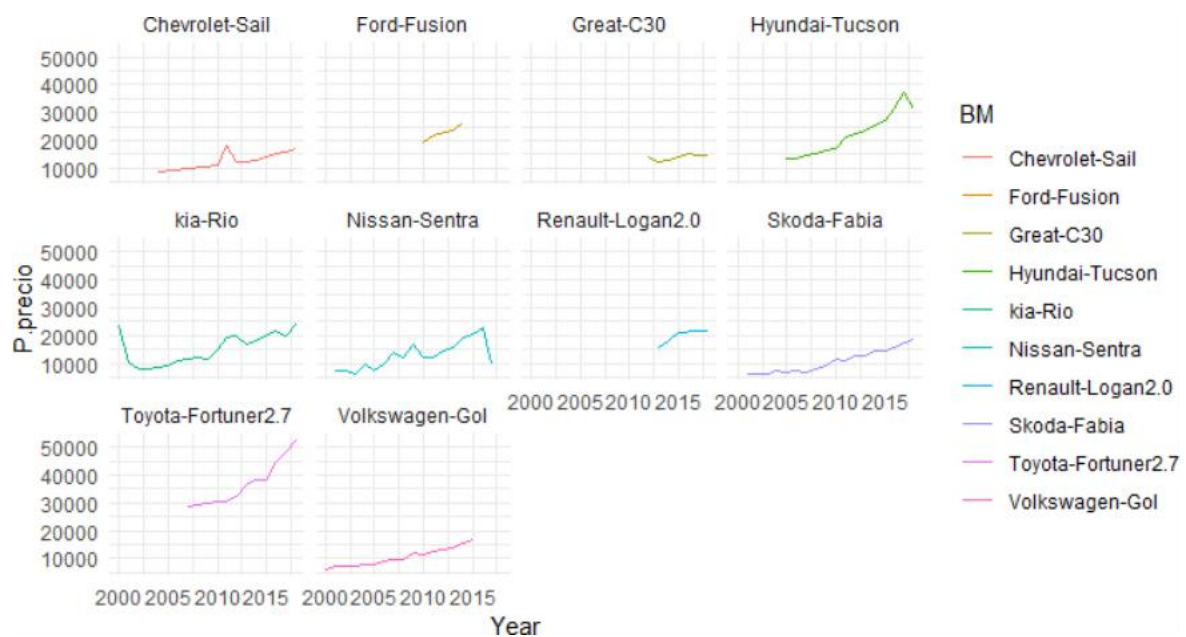
Variación de precio de las marcas más demandadas



Nota. Marcas demandadas

Figura 29

Mini gráficas de la variación del precio del modelo de la marca más demandada

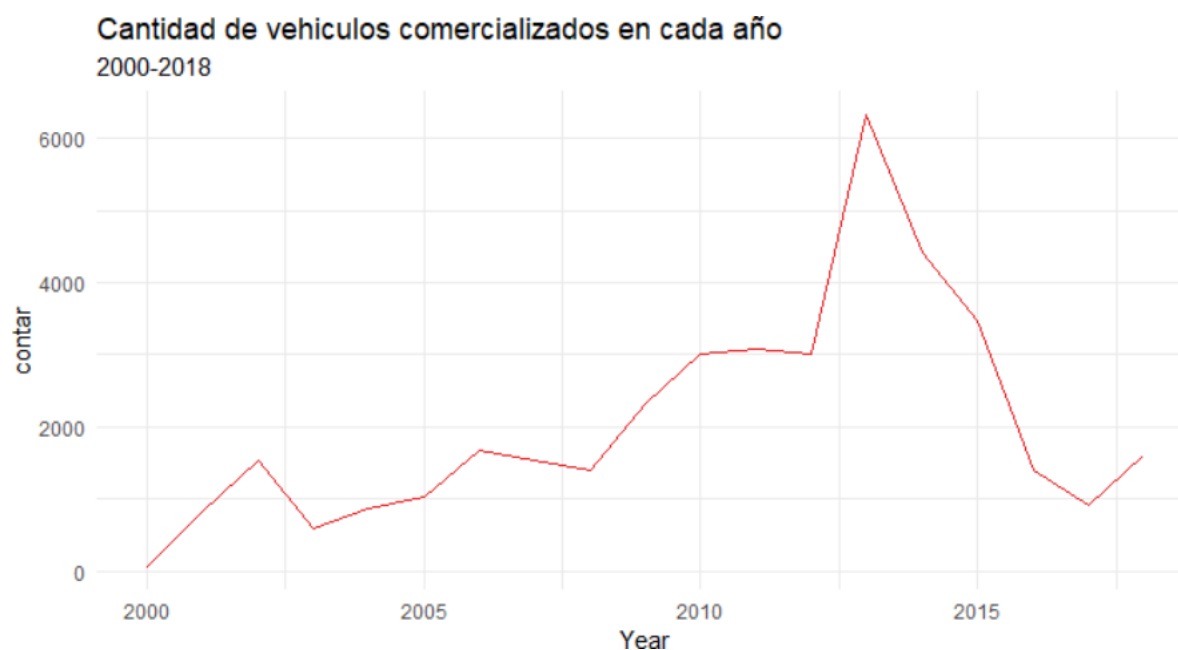


Nota. Marcas demandadas

Además, se comparó la cantidad de vehículos que pertenecen a cada año es decir los vehículos de segunda mano que se están comercializando, a que año corresponde más, resaltando de acuerdo con la figura 33 de manera visible el año 2013, vehículos que tiene 5 años de antigüedad. Este resultado no llama la atención, ya que es el año en el cual, el gobierno estableció cuotas a las importaciones de ciertos productos incluyendo el sector automotriz.

Figura 30

Cantidad de vehículos comercializados en cada año.

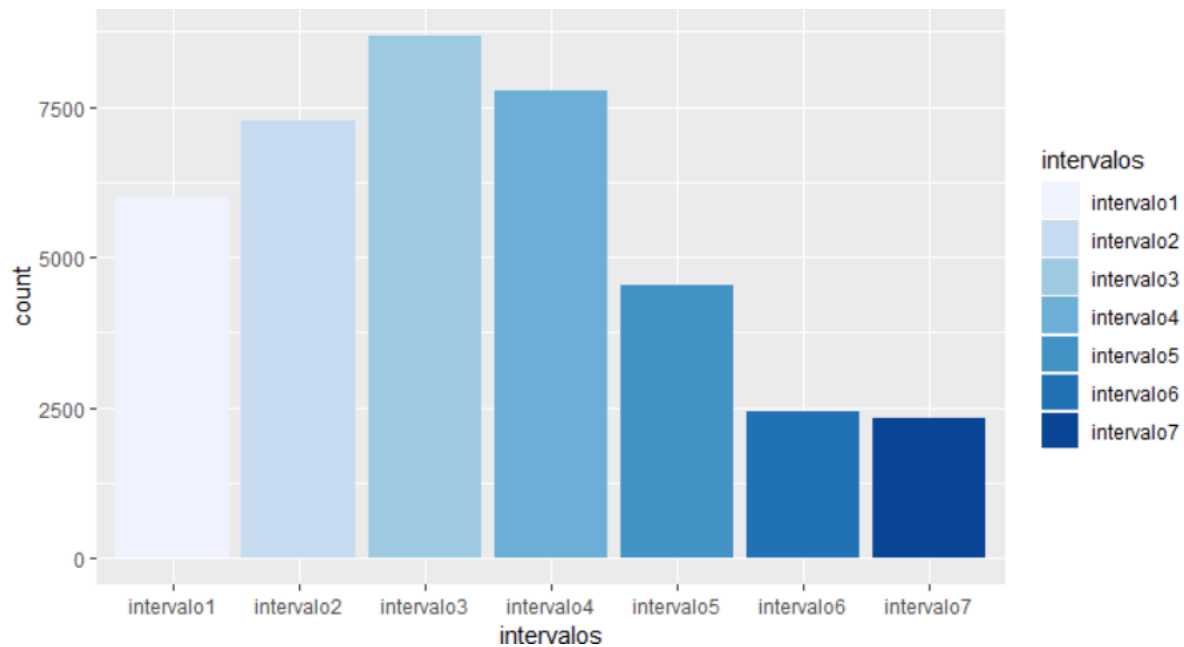


Nota. Vehículos comercializados

En el último estudio, se buscó analizar la cantidad de vehículos con su kilometraje, por lo cual se estableció rango de intervalos para el kilometraje, yendo en 40.000 kilómetros, teniendo siete intervalos. Como se puede observar el intervalo que mayor predomina es el 3 donde va de 80.000 a 120.000 kilómetros. Si comparáramos con el año 2013 y a la actualidad son 5 años, se podría concluir que cada año aproximadamente recorre un vehículo entre 16.000 y 24.000 kilómetros.

Figura 31

Vehículos por Kilometraje en intervalos

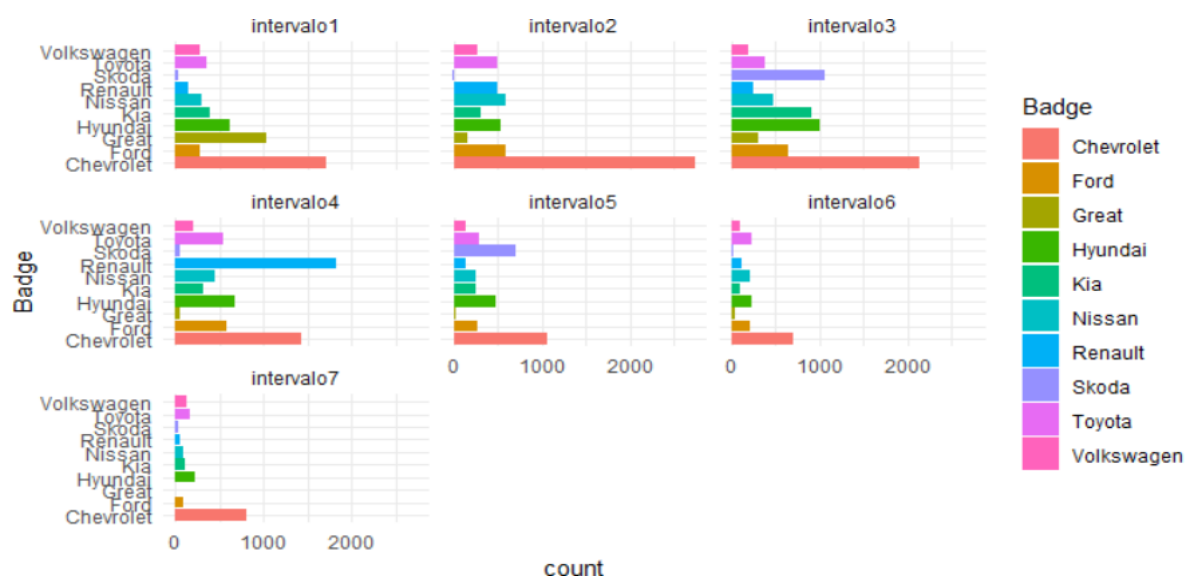


Nota. Vehículos comercializados

Por último en la siguiente figura se puede observar el top ten de marcas en cada intervalo predominando en casi todas Chevrolet, seguido de la marca Renault.

Figura 32

Marcas de vehículos por intervalo de Kilometraje



Nota. Vehículos comercializados

Después de realizar un resumen de información estadística de las variables que componen la base se va a aplicar regresión lineal con el objetivo de identificar las variables independientes relevantes que pueden predecir el precio en el cual se puede comercializar un automóvil de segunda mano con ciertas características.

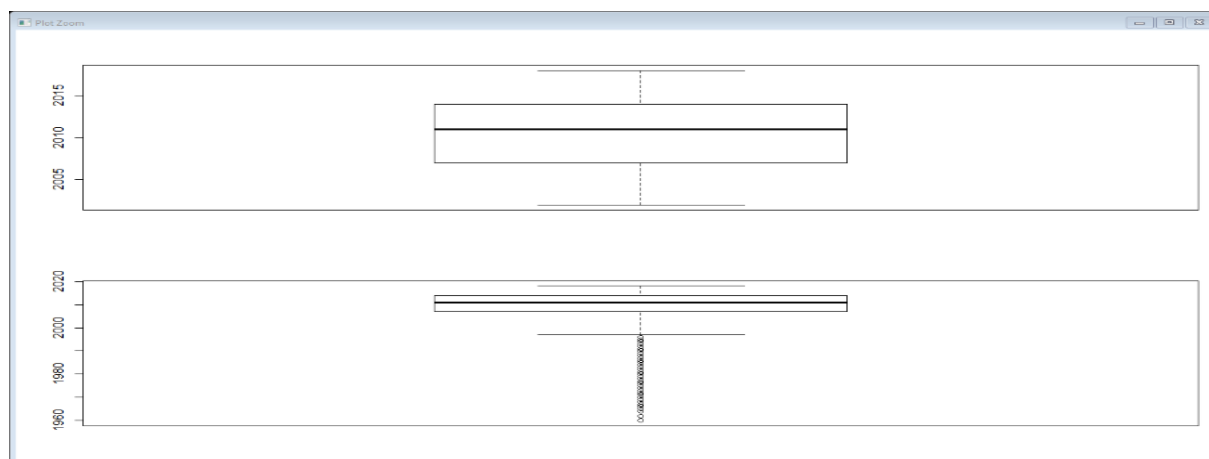
5.2 Regresiones

Para ejecutar la regresión se utilizó el software libre R y de igual manera el entorno de Rstudio. Se ha realizado 4 regresiones todas en función del precio, es decir que en todas las regresiones el precio es la variable dependiente y las demás serán consideradas variables independientes. Antes de continuar, el tipo de regresión que se empleara en el análisis es la regresión lineal, la cual es una técnica estadística para predecir el resultado de una variable numérica que en el presente trabajo es el precio del vehículo de segunda mano.

La primera regresión es el precio en función del kilometraje y año de comercialización del vehículo. La segunda regresión se efectuó con el precio en función de Kilometraje y años, pero agrupando los automóviles menores a 2002 como año 2002, por lo que se creó una columna adicional con esta particularidad. Se tomó como referencia este año, ya que de acuerdo con el siguiente grafico de cajas, no permite tener datos que se encuentren fuera del diagrama.

Figura 33

Diagrama de caja de los automóviles por su año



Nota. Vehículos comercializados

Para la tercera regresión, se tomó en cuenta tres variables independientes (Kilometraje, años y marca del vehículo que se comercializa) y como variable dependiente el precio, se debe de aclarar que la variable de años no es la agrupada. La última regresión empleada es utilizando tres variables independientes como en la anterior regresión, se seleccionaron, el kilometraje, años y el modelo del vehículo.

Los resultados obtenidos en cada una de las regresiones son los siguientes:

5.2.1 Regresión 1

Ejecutando la regresión 1, primero se puede observar unos valores residuales, los cuales constan los cuartiles, también los valores máximos y mínimos, además del error estándar, la estimación y el t valor.

Figura 34

Regresión lineal número 1

```
Call:
lm(formula = Price ~ Km + dumYear, data = CARS)

Residuals:
    Min       1Q   Median       3Q      Max
-22371  -5744  -1920   2721 120822

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.802e+04  9.997e+03   2.803  0.00506 **
Km          -1.227e-02  6.844e-04 -17.927 < 2e-16 ***
dumYear1962 -1.301e+04  1.414e+04  -0.920  0.35740
dumYear1964 -1.085e+04  1.414e+04  -0.767  0.44301
dumYear1965 -1.333e+04  1.118e+04  -1.192  0.23308
dumYear1966 -1.240e+04  1.224e+04  -1.013  0.31125
dumYear1967 -1.367e+04  1.154e+04  -1.184  0.23639
dumYear1968 -7.097e+03  1.118e+04  -0.635  0.52551
dumYear1969 -1.250e+04  1.155e+04  -1.082  0.27906
dumYear1970 -1.150e+04  1.044e+04  -1.101  0.27091
dumYear1971 -2.268e+04  1.080e+04  -2.100  0.03572 *
dumYear1972 -1.792e+04  1.029e+04  -1.742  0.08156 .
dumYear1973 -1.005e+04  1.006e+04  -0.999  0.31792
dumYear1974 -2.003e+04  1.010e+04  -1.984  0.04725 *
dumYear1975 -1.873e+04  1.016e+04  -1.843  0.06537 .
dumYear1976 -1.622e+04  1.095e+04  -1.481  0.13849
dumYear1977 -1.767e+04  1.020e+04  -1.733  0.08305 .
dumYear1978 -1.861e+04  1.012e+04  -1.840  0.06580 .
```

Nota. Fórmula de cálculo

Los resultados de la regresión que se muestran, el kilometraje es una variable relevante para el precio, con una significación del 99%, sin embargo, no todos los años son relevantes para explicar la variable dependiente, además r-cuadrado 0,1953, lo cual

nos indica que un 20% de las variables independientes explican la variable dependiente del modelo.

Figura 35

Resultados de la regresión 1

```
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

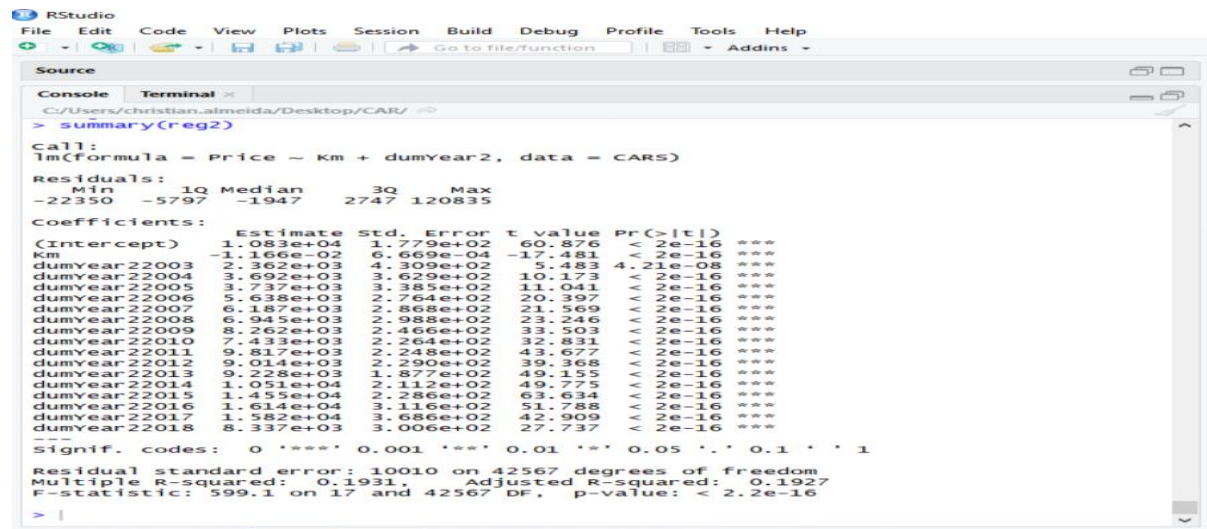
Residual standard error: 9997 on 42527 degrees of freedom
Multiple R-squared:  0.1964,    Adjusted R-squared:  0.1953
F-statistic: 182.3 on 57 and 42527 DF,  p-value: < 2.2e-16
```

Nota. Cálculo

5.2.2 Regresión 2

Figura 36

Regresión lineal número 2



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
C:/Users/christian.almeida/Desktop/CAR/
> summary(reg2)

Call:
lm(formula = Price ~ Km + dumYear2, data = CARS)

Residuals:
    Min       1Q   Median       3Q      Max
-22350  -5797  -1947    2747  120835

Coefficients:
(Intercept)      1.083e+04  1.779e+02  60.876 < 2e-16 ***
Km          -1.166e-02   6.669e-04 -17.481 < 2e-16 ***
dumyear22003  2.362e+03   4.309e+02  5.483  4.21e-08 ***
dumyear22004  3.692e+03   3.629e+02  10.173 < 2e-16 ***
dumyear22005  3.737e+03   3.385e+02  11.041 < 2e-16 ***
dumyear22006  5.638e+03   2.764e+02  20.397 < 2e-16 ***
dumyear22007  6.187e+03   2.868e+02  21.569 < 2e-16 ***
dumyear22008  6.945e+03   2.988e+02  23.246 < 2e-16 ***
dumyear22009  8.262e+03   2.466e+02  33.503 < 2e-16 ***
dumyear22010  7.433e+03   2.264e+02  32.831 < 2e-16 ***
dumyear22011  9.817e+03   2.248e+02  43.677 < 2e-16 ***
dumyear22012  9.014e+03   2.290e+02  39.368 < 2e-16 ***
dumyear22013  9.228e+03   1.877e+02  49.155 < 2e-16 ***
dumyear22014  1.051e+04   2.112e+02  49.775 < 2e-16 ***
dumyear22015  1.455e+04   2.286e+02  63.634 < 2e-16 ***
dumyear22016  1.614e+04   3.116e+02  51.788 < 2e-16 ***
dumyear22017  1.582e+04   3.686e+02  42.909 < 2e-16 ***
dumyear22018  8.337e+03   3.006e+02  27.737 < 2e-16 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10010 on 42567 degrees of freedom
Multiple R-squared:  0.1931,    Adjusted R-squared:  0.1927
F-statistic: 599.1 on 17 and 42567 DF,  p-value: < 2.2e-16
> |
```

Nota. Cálculo

En la regresión número 2, el kilometraje y los años agrupados si es menor a 2002 como 2002, es significativo con un nivel de confianza al 99%, sin embargo, el resultado

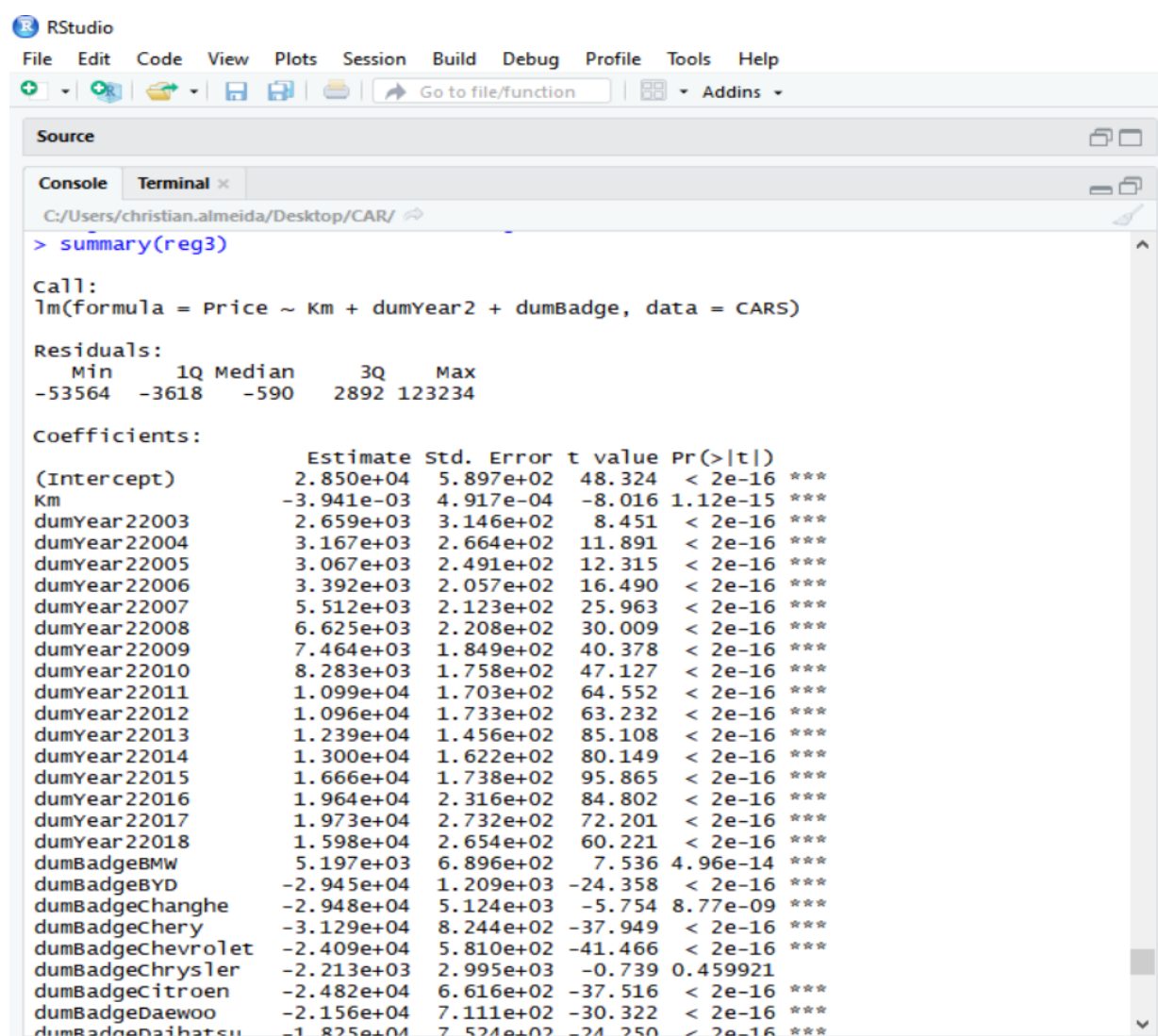
del r-cuadrado es de 0,1927, resultado similar al de la regresión número1, observando que las variables independientes pueden explicar el 19% de la variable dependiente.

5.2.3 Regresión 3

En la regresión número tres donde se consideran como variables independientes el kilometraje, la marca y los años de manera agrupada, como resultados básicos se obtiene el mínimo, el máximo y los cuartiles.

Figura 37

Regresión lineal número 3



```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source
Console Terminal x
C:/Users/christian.almeida/Desktop/CAR/
> summary(reg3)

Call:
lm(formula = Price ~ Km + dumYear2 + dumBadge, data = CARS)

Residuals:
    Min       1Q   Median       3Q      Max
-53564  -3618   -590    2892  123234

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.850e+04  5.897e+02  48.324 < 2e-16 ***
Km          -3.941e-03  4.917e-04  -8.016 1.12e-15 ***
dumYear22003  2.659e+03  3.146e+02   8.451 < 2e-16 ***
dumYear22004  3.167e+03  2.664e+02  11.891 < 2e-16 ***
dumYear22005  3.067e+03  2.491e+02  12.315 < 2e-16 ***
dumYear22006  3.392e+03  2.057e+02  16.490 < 2e-16 ***
dumYear22007  5.512e+03  2.123e+02  25.963 < 2e-16 ***
dumYear22008  6.625e+03  2.208e+02  30.009 < 2e-16 ***
dumYear22009  7.464e+03  1.849e+02  40.378 < 2e-16 ***
dumYear22010  8.283e+03  1.758e+02  47.127 < 2e-16 ***
dumYear22011  1.099e+04  1.703e+02  64.552 < 2e-16 ***
dumYear22012  1.096e+04  1.733e+02  63.232 < 2e-16 ***
dumYear22013  1.239e+04  1.456e+02  85.108 < 2e-16 ***
dumYear22014  1.300e+04  1.622e+02  80.149 < 2e-16 ***
dumYear22015  1.666e+04  1.738e+02  95.865 < 2e-16 ***
dumYear22016  1.964e+04  2.316e+02  84.802 < 2e-16 ***
dumYear22017  1.973e+04  2.732e+02  72.201 < 2e-16 ***
dumYear22018  1.598e+04  2.654e+02  60.221 < 2e-16 ***
dumBadgeBMW   5.197e+03  6.896e+02   7.536 4.96e-14 ***
dumBadgeBYD   -2.945e+04  1.209e+03  -24.358 < 2e-16 ***
dumBadgeChanghe -2.948e+04  5.124e+03  -5.754 8.77e-09 ***
dumBadgeChery -3.129e+04  8.244e+02  -37.949 < 2e-16 ***
dumBadgeChevrolet -2.409e+04  5.810e+02  -41.466 < 2e-16 ***
dumBadgeChrysler -2.213e+03  2.995e+03  -0.739 0.459921
dumBadgeCitroen -2.482e+04  6.616e+02  -37.516 < 2e-16 ***
dumBadgeDaewoo -2.156e+04  7.111e+02  -30.322 < 2e-16 ***
dumBadgeDaihatsu -1.825e+04  7.524e+02  -24.250 < 2e-16 ***

```

Nota. Cálculo

El resultado obtenido es que el kilometraje es significativo al igual que todos los años al 99%, la mayoría de marcas de vehículos son relevantes al 99% (por ejemplo, Skoda, Toyota y volvo), algunas son relevantes al 95%(como por ejemplo Lexus) y otras no son significativas (como por ejemplo Chriysler), el r-cuadrado a diferencia de las otras dos regresiones aumento al 58%, se concluye que el 58% de las variables explican el modelo, a pesar que es un buen resultado aplicaremos una última regresión considerando el modelo de los vehículos de segunda mano.

Figura 38

Continuación regresión lineal número 3

```
dumBadgeLada      -2.508e+04  1.582e+03 -15.857 < 2e-16 ***
dumBadgeLand      6.450e+02  8.655e+02  0.745 0.456159
dumBadgeLexus     -2.481e+03  1.166e+03 -2.127 0.033395 *
dumBadgeLifan     -3.055e+04  2.157e+03 -14.164 < 2e-16 ***
dumBadgeMazda     -1.866e+04  6.078e+02 -30.703 < 2e-16 ***
dumBadgeMercedes  -8.195e+03  7.130e+02  11.494 < 2e-16 ***
dumBadgeMG        -2.833e+04  1.792e+03 -15.806 < 2e-16 ***
dumBadgeMini      -1.066e+04  1.609e+03 -6.626 3.48e-11 ***
dumBadgeMitsubishi -1.240e+04  6.161e+02 -20.125 < 2e-16 ***
dumBadgeNissan     -1.873e+04  5.939e+02 -31.533 < 2e-16 ***
dumBadgeOldsmobile -2.180e+04  2.160e+03 -10.092 < 2e-16 ***
dumBadgePeugeot   -2.137e+04  6.516e+02 -32.795 < 2e-16 ***
dumBadgePolaris   -2.612e+04  2.468e+03 -10.581 < 2e-16 ***
dumBadgePorsche    5.015e+04  8.804e+02  56.961 < 2e-16 ***
dumBadgeRenault   -2.580e+04  5.907e+02 -43.684 < 2e-16 ***
dumBadgeSaic      -3.035e+04  1.530e+03 -19.827 < 2e-16 ***
dumBadgeSkoda     -2.485e+04  6.023e+02 -41.253 < 2e-16 ***
dumBadgeSoueast   -1.844e+04  5.127e+03 -3.597 0.000322 ***
dumBadgeSsangyong -2.241e+04  2.248e+03 -9.968 < 2e-16 ***
dumBadgeSubaru    -1.044e+04  2.995e+03 -3.486 0.000492 ***
dumBadgeSuzuki    -2.189e+04  6.113e+02 -35.811 < 2e-16 ***
dumBadgeToyota    -1.208e+04  5.936e+02 -20.356 < 2e-16 ***
dumBadgeVolkswagen -2.062e+04  6.060e+02 -34.020 < 2e-16 ***
dumBadgeVolvo     -9.443e+03  1.317e+03 -7.168 7.71e-13 ***
dumBadgeZotye     -2.675e+04  1.193e+03 -22.424 < 2e-16 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7198 on 42519 degrees of freedom
Multiple R-squared:  0.5835,    Adjusted R-squared:  0.5829
F-statistic: 916.5 on 65 and 42519 DF,  p-value: < 2.2e-16

> #REGRESION 3
#Precio en funcion de km, años y marca
dumBadge<- factor(CARS$Badge)
reg3<- lm(Price~Km+dumYear2+dumBadge, data=CARS)
summary(reg3)
```

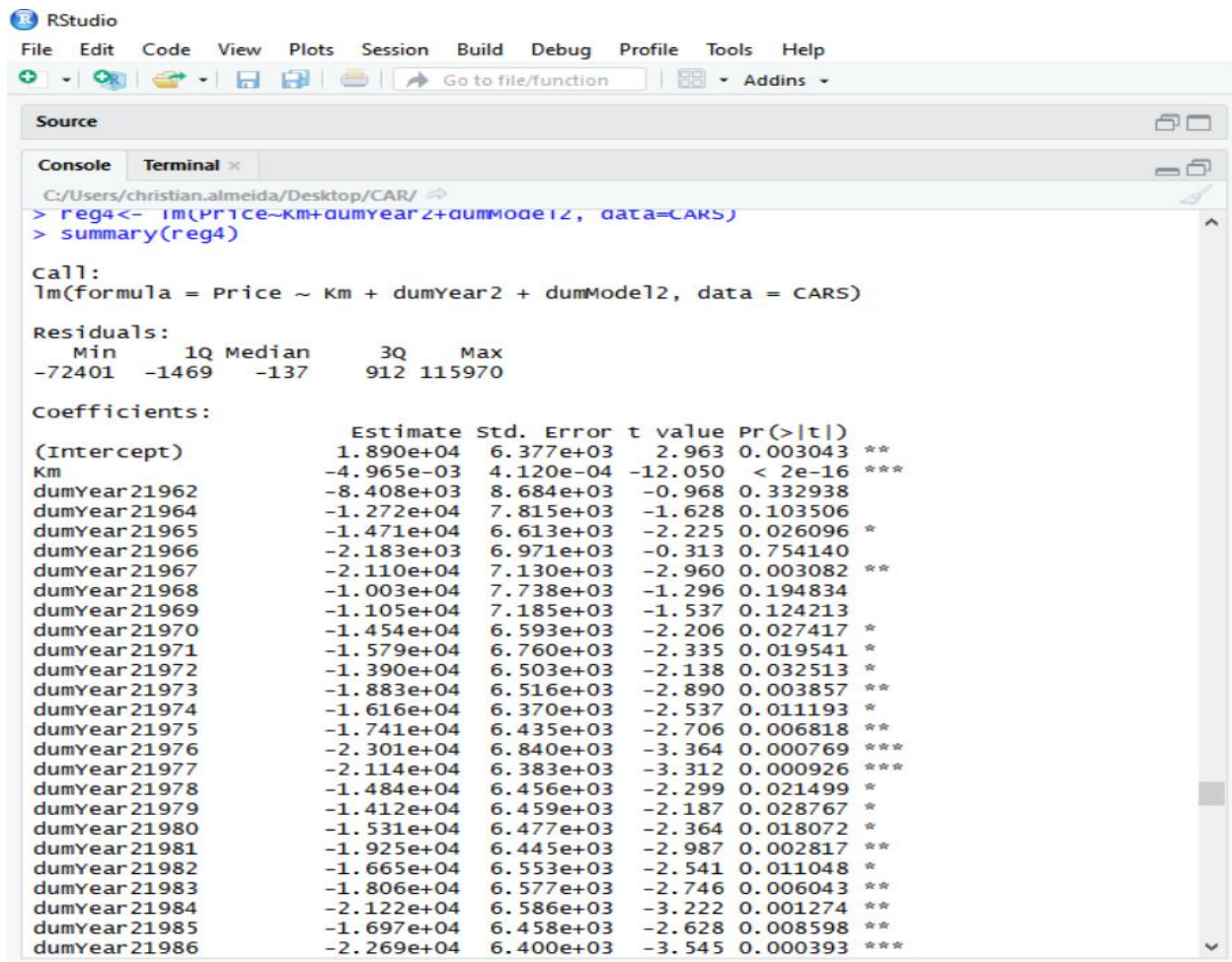
Nota. Cálculo

5.2.4 Regresión 4.

En la regresión 4, se ejecuta considerando las variables independientes al kilometraje, modelo de vehículo de segunda mano y año de fabricación sin ningún agrupamiento, como resultados básicos, se puede observar igual que las anteriores regresiones el máximo y mínimo al igual que los cuartiles.

Figura 39

Regresión lineal número 4



The screenshot shows the RStudio interface with the following content in the Console:

```
C:/Users/christian.almeida/Desktop/CAR/
> reg4 <- lm(Price~Km+dumYear2+dumModel2, data=CAKS)
> summary(reg4)
```

Call:
lm(formula = Price ~ Km + dumYear2 + dumModel2, data = CARS)

Residuals:

	Min	1Q	Median	3Q	Max
	-72401	-1469	-137	912	115970

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.890e+04	6.377e+03	2.963	0.003043	**
Km	-4.965e-03	4.120e-04	-12.050	< 2e-16	***
dumYear21962	-8.408e+03	8.684e+03	-0.968	0.332938	
dumYear21964	-1.272e+04	7.815e+03	-1.628	0.103506	
dumYear21965	-1.471e+04	6.613e+03	-2.225	0.026096	*
dumYear21966	-2.183e+03	6.971e+03	-0.313	0.754140	
dumYear21967	-2.110e+04	7.130e+03	-2.960	0.003082	**
dumYear21968	-1.003e+04	7.738e+03	-1.296	0.194834	
dumYear21969	-1.105e+04	7.185e+03	-1.537	0.124213	
dumYear21970	-1.454e+04	6.593e+03	-2.206	0.027417	*
dumYear21971	-1.579e+04	6.760e+03	-2.335	0.019541	*
dumYear21972	-1.390e+04	6.503e+03	-2.138	0.032513	*
dumYear21973	-1.883e+04	6.516e+03	-2.890	0.003857	**
dumYear21974	-1.616e+04	6.370e+03	-2.537	0.011193	*
dumYear21975	-1.741e+04	6.435e+03	-2.706	0.006818	**
dumYear21976	-2.301e+04	6.840e+03	-3.364	0.000769	***
dumYear21977	-2.114e+04	6.383e+03	-3.312	0.000926	***
dumYear21978	-1.484e+04	6.456e+03	-2.299	0.021499	*
dumYear21979	-1.412e+04	6.459e+03	-2.187	0.028767	*
dumYear21980	-1.531e+04	6.477e+03	-2.364	0.018072	*
dumYear21981	-1.925e+04	6.445e+03	-2.987	0.002817	**
dumYear21982	-1.665e+04	6.553e+03	-2.541	0.011048	*
dumYear21983	-1.806e+04	6.577e+03	-2.746	0.006043	**
dumYear21984	-2.122e+04	6.586e+03	-3.222	0.001274	**
dumYear21985	-1.697e+04	6.458e+03	-2.628	0.008598	**
dumYear21986	-2.269e+04	6.400e+03	-3.545	0.000393	***

Nota. Cálculo

Se obtiene como resultado que la variable kilometraje es significativa de igual manera ciertos años con un nivel de confianza del 95% (por ejemplo, los años 1978, 1979, 1980, 18982), en cambio un gran número de modelos es significativa a un nivel de confianza del 99% (como por ejemplo los años 1974, 1981, 1983, 1986) el r-cuadrado es del 0,7542, es decir que las variables independientes explican el modelo en un 75%, como observación es el r-cuadrado más alto de todas las regresiones aplicadas anteriormente.

Para la regresión cuatro, que es el modelo que más variables tiene y que a su vez son las variables que más explican a la variable dependiente por el r-cuadrado mayor

respecto de las otras regresiones, se le realizó un análisis de heterocedasticidad y de multicolinealidad. Para ello se ejecutará en R.

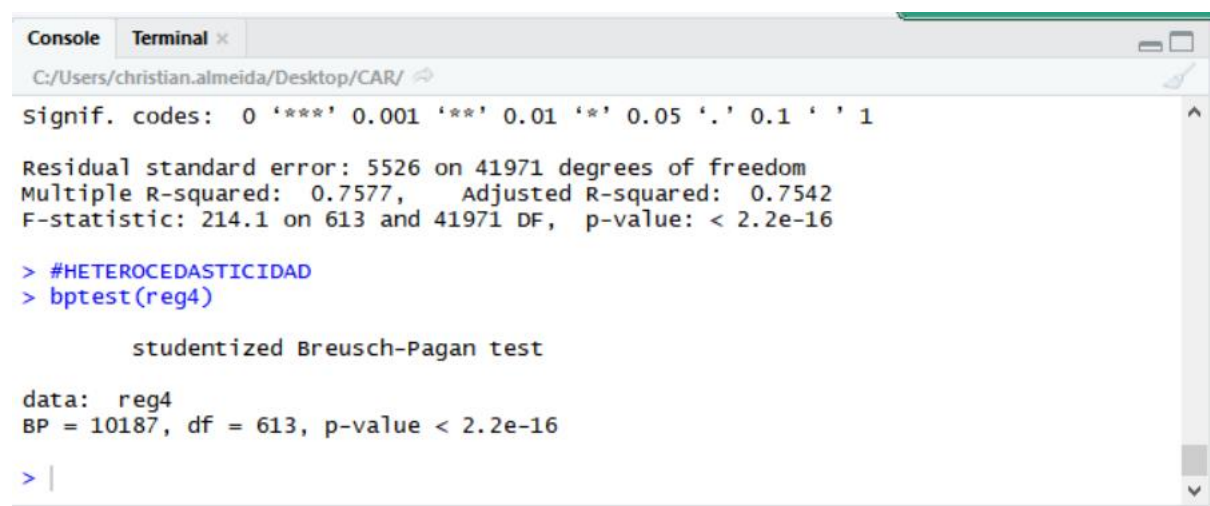
Antes de continuar con el análisis, explicaré que es la multicolinealidad y heterocedasticidad, los cuales son requisitos básicos que deben cumplir en las regresiones lineales. La heterocedasticidad compara los errores de la varianza y se dice que hay heterocedasticidad cuando los errores no son constantes a lo largo de la regresión, lo cual no cumple con un requisito de las regresiones lineales, que es que debe existir homocedasticidad (lo contrario a heterocedasticidad), por lo que para que exista homocedasticidad se debe corregir la heterocedasticidad en el modelo de la regresión lineal.

Ahora explicare la multicolinealidad, el cual también es un requisito básico que debe cumplir en la regresión lineal, la multicolinealidad permite identificar si existe correlación entre las variables independientes del modelo, es decir si están correlacionadas las variables predictoras de la regresión lineal, si existe multicolinealidad no es bueno para el modelo y se tendría que considerar diferentes acciones como la eliminación de las variables o transformar variables.

Para identificar la heterocedasticidad, se realiza mediante la función “bptest”, obteniendo los resultados que se muestra en la siguiente figura:

Figura 40

Análisis de heterocedasticidad



```
Console Terminal x
C:/Users/christian.almeida/Desktop/CAR/
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5526 on 41971 degrees of freedom
Multiple R-squared:  0.7577,    Adjusted R-squared:  0.7542 
F-statistic: 214.1 on 613 and 41971 DF,  p-value: < 2.2e-16

> #HETEROCEDASTICIDAD
> bptest(reg4)

      studentized Breusch-Pagan test

data:  reg4
BP = 10187, df = 613, p-value < 2.2e-16

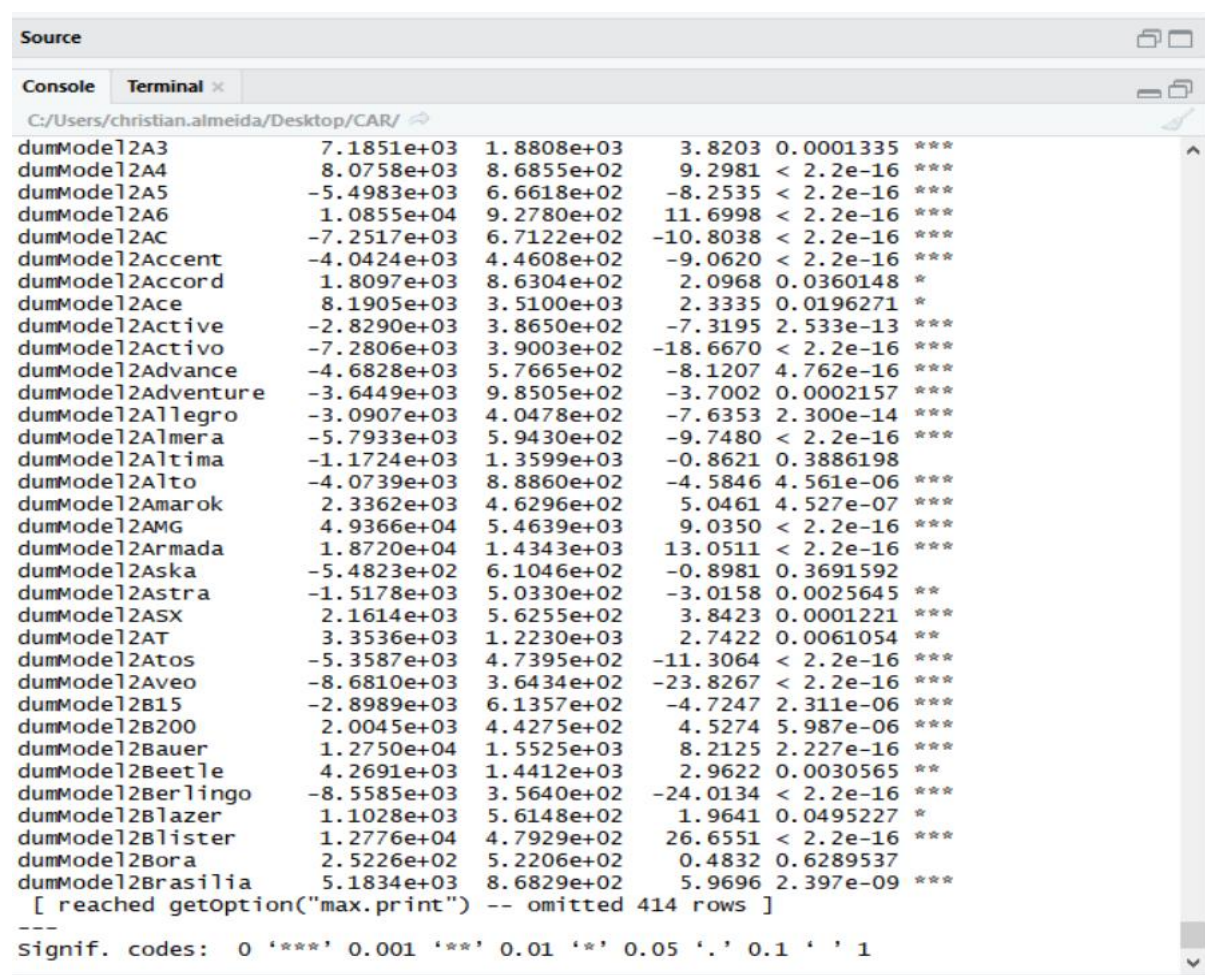
> |
```

Nota. Cálculo

Como se puede observar el resultado del p-value es inferior al 1% por lo que se puede concluir que, si existe heterocedasticidad, lo cual no es bueno para el modelo ya que la varianza del error no es constante y no permite que las estimaciones sean precisas, para corregir la heterocedasticidad en el modelo se emplea la función "vcovHC". Después de corregir la heterocedasticidad, volvemos a revisar los resultados obtenidos de la regresión número 4, en la siguiente figura.

Figura 41

Modelo con corrección de heterocedasticidad



Model	Coef	Std. Err.	t-stat	p-value	Signif.
dumModel2A3	7.1851e+03	1.8808e+03	3.8203	0.0001335	***
dumModel2A4	8.0758e+03	8.6855e+02	9.2981	< 2.2e-16	***
dumModel2A5	-5.4983e+03	6.6618e+02	-8.2535	< 2.2e-16	***
dumModel2A6	1.0855e+04	9.2780e+02	11.6998	< 2.2e-16	***
dumModel2AC	-7.2517e+03	6.7122e+02	-10.8038	< 2.2e-16	***
dumModel2Accent	-4.0424e+03	4.4608e+02	-9.0620	< 2.2e-16	***
dumModel2Accord	1.8097e+03	8.6304e+02	2.0968	0.0360148	*
dumModel2Ace	8.1905e+03	3.5100e+03	2.3335	0.0196271	*
dumModel2Active	-2.8290e+03	3.8650e+02	-7.3195	2.533e-13	***
dumModel2Activo	-7.2806e+03	3.9003e+02	-18.6670	< 2.2e-16	***
dumModel2Advance	-4.6828e+03	5.7665e+02	-8.1207	4.762e-16	***
dumModel2Adventure	-3.6449e+03	9.8505e+02	-3.7002	0.0002157	***
dumModel2Allegro	-3.0907e+03	4.0478e+02	-7.6353	2.300e-14	***
dumModel2Almera	-5.7933e+03	5.9430e+02	-9.7480	< 2.2e-16	***
dumModel2Altim	-1.1724e+03	1.3599e+03	-0.8621	0.3886198	
dumModel2Alto	-4.0739e+03	8.8860e+02	-4.5846	4.561e-06	***
dumModel2Amarok	2.3362e+03	4.6296e+02	5.0461	4.527e-07	***
dumModel2AMG	4.9366e+04	5.4639e+03	9.0350	< 2.2e-16	***
dumModel2Armada	1.8720e+04	1.4343e+03	13.0511	< 2.2e-16	***
dumModel2Aska	-5.4823e+02	6.1046e+02	-0.8981	0.3691592	
dumModel2Astra	-1.5178e+03	5.0330e+02	-3.0158	0.0025645	**
dumModel2ASX	2.1614e+03	5.6255e+02	3.8423	0.0001221	***
dumModel2AT	3.3536e+03	1.2230e+03	2.7422	0.0061054	**
dumModel2Ato	-5.3587e+03	4.7395e+02	-11.3064	< 2.2e-16	***
dumModel2Aveo	-8.6810e+03	3.6434e+02	-23.8267	< 2.2e-16	***
dumModel2B15	-2.8989e+03	6.1357e+02	-4.7247	2.311e-06	***
dumModel2B200	2.0045e+03	4.4275e+02	4.5274	5.987e-06	***
dumModel2Bauer	1.2750e+04	1.5525e+03	8.2125	2.227e-16	***
dumModel2Beetle	4.2691e+03	1.4412e+03	2.9622	0.0030565	**
dumModel2Berlingo	-8.5585e+03	3.5640e+02	-24.0134	< 2.2e-16	***
dumModel2Blazer	1.1028e+03	5.6148e+02	1.9641	0.0495227	*
dumModel2Blister	1.2776e+04	4.7929e+02	26.6551	< 2.2e-16	***
dumModel2Bora	2.5226e+02	5.2206e+02	0.4832	0.6289537	
dumModel2Brasilia	5.1834e+03	8.6829e+02	5.9696	2.397e-09	***

[reached getoption("max.print") -- omitted 414 rows]

 signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Nota. Cálculo

Como se observa los resultados obtenidos con la corrección de la heterocedasticidad, son distintos a los aplicados anteriormente, por ejemplo, los modelos Accord, Adventure, Almera, Aska y Astra que antes no eran significativos, con la

corrección son significativos para el modelo con un nivel de confianza entre el 95% y 99%.

También se realizó análisis de multicolinealidad entre las variables, mediante la función “imcdiag” en R, por sus siglas que significan Individual Multicollinearity Diagnostic Measures para conocer si existe o no multicolinealidad. Como se puede observar en el gráfico, no existe multicolinealidad, lo cual es bueno para el modelo, ya que no están correlacionados entre las variables independientes, como aclaración, para que exista multicolinealidad el valor resultante debe ser mayor a 5, en algunos libros indica que debe ser mayor a 10 (Gujarati,2010), el resultado obtenido es de 1,3862 menor a lo mencionado, gráficamente se lo puede identificar en la figura 41.

5.3 Predicciones

A continuación, mediante R se probaron las regresiones realizadas mediante la función predict.lm la cual nos permitió, de acuerdo con la regresión que se ejecutó, poder determinar el precio en función de las variables que se ingresó en el modelo.

La primera prueba que se realizó fue con la regresión de las variables independientes (año y marca) para conocer qué precio tendría como resultado el costo del vehículo de segunda mano. Como se observa en la figura se tomó la marca “Hyundai” y el año 2015 para conocer el precio. Dando como resultado \$21.917 dólares americanos. Como la marca no es relevante ya que depende del modelo porque existen autos de alta gama y baja gama en las diferentes marcas los precio de estos también varían.

Figura 42.

Predicción 1 considerando año y marca

```
131
132
133 library(leaps)
134 # combinatorio
135 #AYUDA A ELEGIR EL MEJOR MODELO
136 combinatorio<- regsubsets(Price~MARCA+MODELO+Km+Year1, data = CARS,method = "AIC")
137 plot(combinatorio)
138
139
140
141 #PREDICCIONES
142 #Predicción del precio de un auto del año 2015 de marca Hyundai
143 reg<- lm(Price~Year+Badge, data = CARS)
144 newdata<- data.frame(Year=2015, Badge="Hyundai")
145 predict<- predict.lm(reg, newdata,level = 0.99)
146 predict
147
148
149
```

Nota. Cálculo

Como la marca, no es un factor muy relevante, ya que dentro de las marcas hay modelos de alta gama y baja gama, los precios de estos también varían. Por lo que a continuación se realizara con la regresión que toma en cuenta el precio del vehículo en función del año y modelo de los automóviles como variables independientes.

Figura 43

Predicción considerando año y modelo del vehículo

```
predict<- predict.lm(reg, newdata, level = 0.99)
predict

#Predicción del precio de un auto del año 2015 de modelo Tucson
reg<- lm(Price~Year+Model, data = CARS)
newdata<- data.frame(Year=2015, Model="Tucson")
predict<- predict.lm(reg, newdata)
predict
```

Nota. Cálculo

Para la predicción se utilizó como datos el mismo año 2015 pero el modelo del vehículo se estableció Tucson, un modelo muy comercializado de la marca Hyundai. Como resultado se obtiene que el precio de acuerdo con el modelo seleccionado es de \$23.779 dólares.

La siguiente predicción se efectuará tomando en cuenta a más de las dos variables independientes anteriores, el kilometraje para conocer el precio que tendría el vehículo con la presencia del recorrido que le dio el dueño del automóvil, por lo que la regresión que se debe tomar en cuenta es la regresión número 4.

Figura 44

Predicción considerando año, modelo y kilometraje del vehículo

```
#####Prediccion del precio con año 2015, modelo Tucson y recorrido 54000 km  
  
reg<- lm(Price~Year+Model+Km, data = CARS)  
newdata<- data.frame(Year=2015, Model="Tucson", Km=54000)  
predict<- predict.lm(reg, newdata)  
predict  
  
#####Prediccion del precio con año 2015, modelo Tucson y recorrido 180000 km  
  
reg<- lm(Price~Year+Model+Km, data = CARS)  
newdata<- data.frame(Year=2015, Model="Tucson", Km=180000)  
predict<- predict.lm(reg, newdata)  
predict
```

Nota. Cálculo

Como se puede observar en la figura 51, se realizó dos predicciones considerando el año de fabricación del vehículo como 2015, el mismo modelo en los dos casos, pero con kilometraje distinto en la primera predicción se consideró 54.000 kilómetros y en la segunda predicción el vehículo tendría 180.000 kilómetros de recorrido. Como resultado se obtiene en la primera predicción de acuerdo con los datos ingresados, el precio sería de \$24.327 dólares, y en la segunda predicción en la cual tiene un mayor kilometraje, el precio del automóvil, sería de \$22.902 dólares. Se puede concluir que el precio si tiene injerencia al momento de establecer el precio de comercialización de un vehículo de segunda mano.

Para tener una comparación, se realizaron las mismas predicciones para un modelo de vehículo de gama alta, de gama intermedia y otro de gama baja, para el vehículo de gama alta se seleccionó la marca Mercedes Benz con el modelo AMG, para la gama intermedia se seleccionó la marca Hyundai con el modelo Tucson y para la gama baja se eligió la marca Chevrolet con el modelo Aveo.

Obteniendo la siguiente tabla de comparación:

Tabla 4*Resumen de predicción por modelo*

Marca	Modelo	2015			2016		
		2015	54000 km	180000 km	2016	54000 km	180000 km
Chevrolet	Aveo	15.42 0	16.00 2	14.57 7	16.41 8	16.92 4	15.50 0
Hyundai	Tucson	23.77 9	24.32 7	22.90 2	24.77 6	25.24 9	23.824
Mercedes Benz	AMG	72.42 0	72.29 4	70.86 9	73.41 8	73.21 6	71.79 2

Nota. Fuente: realizado por el autor.

Como conclusiones se obtiene que en el modelo Aveo, es más sensible al precio tanto si se compara año a año o si cambia el kilometraje, en cambio en el modelo de alta gama el precio no varía significativamente si el kilometraje aumenta o si varía de año a año.

También a continuación se presenta una tabla comparando la marca y el año:

Tabla 5*Resumen de predicción por marca*

Marca	2015	2016
Chevrolet	17.945	18.754
Tucson	21.927	22.736
Mercedes benz	50.900	51.709

Nota. Fuente: realizado por el autor.

Se puede observar y tener una idea de si el vehículo pertenece a cierta marca va a tener un precio de referencia a pesar de que va a depender de los modelos que tenga la marca.

5.4 Web scraping en PatioTuerca: cómo se hizo

El proceso de recolección de datos se llevó a cabo mediante técnicas de web scraping, enfocadas en el portal de vehículos usados PatioTuerca.com.ec, una de las plataformas más utilizadas en Ecuador. Utilizando herramientas como Python,

BeautifulSoup y Selenium, se desarrolló un script automatizado que extrajo información clave de cada anuncio publicado, incluyendo:

- Marca y modelo del vehículo
- Año de fabricación
- Kilometraje
- Precio publicado
- Ubicación geográfica
- Tipo de combustible
- Tipo de transmisión
- Número de puertas
- Motor y cilindraje
- Fecha de publicación del anuncio

Se tomaron precauciones para evitar sobrecargar los servidores del sitio, respetando intervalos de tiempo entre solicitudes y empleando técnicas que simulan el comportamiento humano en la navegación.

5.4.1 Estructura de los datos obtenidos

El resultado del scraping fue almacenado en una base de datos estructurada en formato CSV y luego importado a pandas para su manipulación. El conjunto de datos contenía más de 30.000 registros únicos. Cada fila representaba un vehículo, y cada columna correspondía a una característica.

Se identificaron inconsistencias iniciales como campos vacíos, formatos no uniformes (por ejemplo, el kilometraje reportado en diferentes unidades) y datos atípicos (como vehículos de años improbables o precios excesivos).

5.4.2 Limpieza y transformación de los datos

Para garantizar la calidad del análisis, se aplicaron procesos de limpieza que incluyeron:

- Conversión de precios y kilometrajes a valores numéricos homogéneos.
- Eliminación de registros con información insuficiente o incoherente.
- Normalización de nombres de marcas y modelos.
- Conversión de fechas a formato estándar.
- Codificación de variables categóricas para análisis posterior.

Este preprocesamiento fue fundamental para permitir el análisis estadístico y la construcción de modelos predictivos confiables.

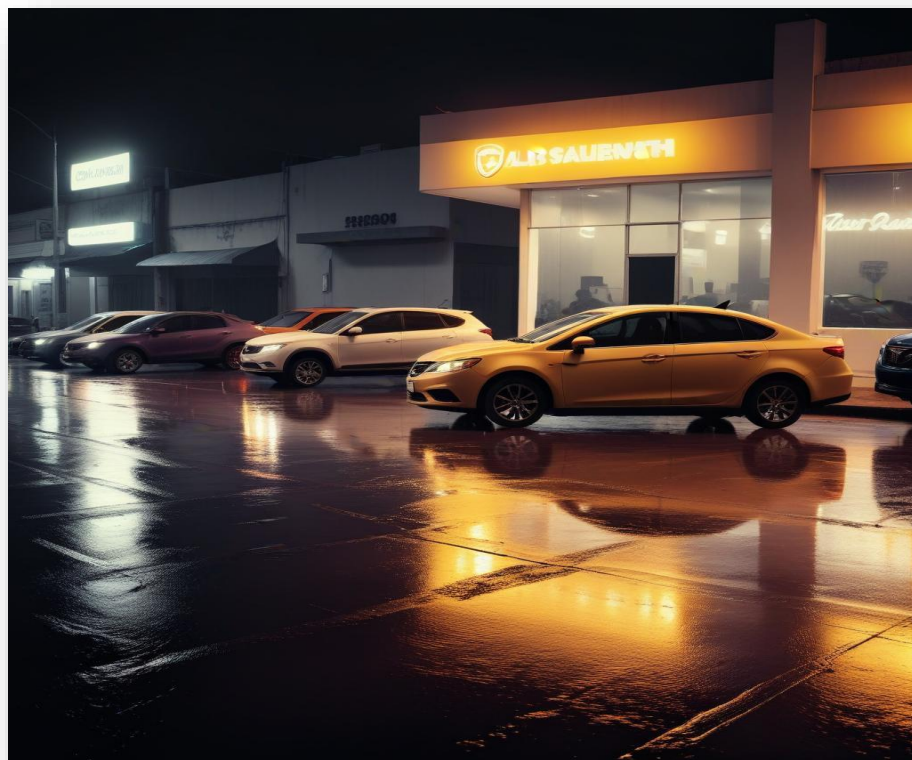
La investigación desarrollada demuestra que la integración de técnicas de ciencia de datos con el mercado automotor ecuatoriano permite transformar la incertidumbre comercial en un sistema de valoración objetivo y transparente. A través de un procedimiento riguroso de web scraping en R, se logró consolidar una base de datos masiva de más de 43.000 registros, superando las limitaciones de la recolección manual y eliminando los sesgos propios de la información fragmentada. Este proceso técnico evidenció que, aunque el mercado digital es una fuente rica en información, requiere una fase crítica de depuración para corregir errores de digitación en precios y kilometrajes, garantizando así que los análisis posteriores se fundamenten en realidades comerciales y no en valores atípicos.

Los resultados estadísticos y el análisis exploratorio revelaron patrones claros de dominio y comportamiento en el parque automotor nacional, donde marcas como Chevrolet, Hyundai y Renault no solo lideran en volumen de oferta, sino que también definen los estándares de liquidez en la reventa. Se pudo constatar que el precio de los vehículos no responde únicamente a la antigüedad, sino que está fuertemente influenciado por factores económicos históricos como la dolarización o las cuotas de importación y por la eficiencia tecnológica de modelos híbridos y europeos. La visualización de estos datos permitió identificar que la depreciación no es uniforme, variando significativamente entre modelos de gama alta, donde el kilometraje tiene un impacto menor, y modelos de gama baja, que muestran una alta sensibilidad al recorrido acumulado.

En la aplicabilidad, el éxito de la regresión lineal múltiple, que alcanzó un coeficiente de determinación del 75%, valida la capacidad de los modelos estadísticos para predecir precios de mercado con alta precisión. Al corregir problemas técnicos como la heterocedasticidad, el modelo final se convirtió en una herramienta robusta capaz de estimar el valor comercial de un vehículo en función de su modelo, año y kilometraje. En conclusión, este estudio entrega una metodología aplicada que reduce la asimetría de información, protege al consumidor de la especulación y proporciona a los tomadores de decisiones una base científica para entender las dinámicas de movilidad en Ecuador.

Capítulo 6

Explorando el comportamiento del mercado automotor ecuatoriano



Fuente: [Free text-to-image generator](#) | [Freepik AI](#)

El capítulo de análisis estadístico y visualización constituye la fase donde la información recolectada se transforma en conocimiento estratégico a través de la descripción detallada del mercado automotor ecuatoriano. Mediante el uso de estadísticas descriptivas, se establece un perfil general del vehículo promedio en el país, el cual ronda los trece mil quinientos dólares con un recorrido de ciento diez mil kilómetros, predominando los motores a gasolina y la transmisión manual. Estos indicadores iniciales permiten dimensionar la magnitud del sector y observar las marcadas diferencias geográficas, especialmente en centros urbanos como Quito y Guayaquil, donde la concentración de la oferta influye directamente en los niveles de precios y en la dinámica de las transacciones.

El estudio profundiza en un análisis comparativo que identifica a marcas como Chevrolet, Kia, Hyundai y Toyota como los pilares del mercado de segunda mano debido a su alta disponibilidad de repuestos y servicios postventa. Un hallazgo fundamental es la variabilidad en la retención del valor de reventa, donde ciertos modelos como la Toyota Hilux o la Kia Sportage demuestran una depreciación anual mucho más lenta en comparación con los vehículos compactos urbanos. Esta sección permite comprender cómo la reputación de durabilidad y la utilidad funcional de cada modelo actúan como factores determinantes que protegen la inversión de los propietarios frente al paso del tiempo y la competencia del mercado.

Para facilitar la interpretación de estas tendencias complejas, el capítulo concluye con la implementación de visualizaciones interactivas desarrolladas en plataformas como Plotly y Tableau. Estas herramientas permiten a los usuarios explorar mapas de calor provinciales, diagramas de dispersión y tablas dinámicas que segmentan el mercado por tipo de transmisión o combustible. Esta aproximación visual no solo democratiza el acceso a la información técnica, sino que se convierte en un recurso práctico para que compradores y vendedores identifiquen con precisión los rangos de precios competitivos y las oportunidades de inversión basadas en datos reales y actualizados.

6.1 Análisis Estadístico y Visualización

6.1.1 Estadísticas descriptivas de los vehículos

Una vez limpios los datos, se realizó un análisis estadístico descriptivo para conocer el comportamiento general del mercado. Algunos hallazgos clave incluyen:

- Precio promedio de los vehículos: USD 13.500, con una mediana de USD 11.800.
- Año promedio de fabricación: 2014.
- Kilometraje promedio: 110.000 km.
- Distribución por tipo de combustible: 85% gasolina, 12% diésel, 3% híbridos o eléctricos.
- Transmisión: 76% manual, 24% automática.

Se observaron diferencias importantes por ciudad, siendo Quito y Guayaquil las regiones con mayor volumen de publicaciones, y donde los precios tienden a ser ligeramente más altos.

6.1.2 Análisis comparativo entre marcas, modelos y años

El análisis comparativo se centró en identificar patrones de comportamiento en el mercado automotriz, considerando variables como marca, modelo, año de fabricación y su influencia en el valor de reventa.

6.1.2.1 Marcas más populares. Los datos revelan que Chevrolet, Kia, Hyundai y Toyota lideran las preferencias del consumidor, concentrando un alto porcentaje de las transacciones en el segmento de vehículos usados. Este predominio se asocia no solo a la disponibilidad y cobertura de servicios postventa, sino también a la percepción de confiabilidad y el acceso a repuestos.

6.1.2.2 Modelos con mejor retención de valor. Dentro de este grupo, destacan modelos como la Toyota Hilux y la Kia Sportage, que exhiben una capacidad notable para mantener un precio competitivo incluso con varios años de uso. Este fenómeno se explica por factores como su reputación de durabilidad, la alta demanda en sectores

productivos (como transporte y construcción en el caso de la Hilux) y un posicionamiento sólido en la mente del consumidor.

6.1.2.3 Depreciación anual. El estudio identificó una pérdida promedio de valor del 8% anual, aunque esta tasa no es uniforme:

- Vehículos utilitarios y camionetas de trabajo presentan una depreciación más lenta debido a su utilidad funcional.
- Modelos urbanos o compactos tienden a depreciarse más rápido, en parte por la rápida renovación de diseños y la fuerte competencia de precios en este segmento.

En conjunto, este análisis permite comprender la dinámica competitiva y las tendencias de oferta-demanda en distintos nichos del mercado automotor. La información obtenida es clave para orientar decisiones estratégicas en compra, venta o inversión, anticipando el comportamiento del valor residual según la marca, el modelo y el año de fabricación.

6.1.3 Visualización interactiva de los datos

Se desarrollaron y publicaron gráficos interactivos empleando Plotly y Tableau Public, los cuales facilitaron la exploración visual de los datos desde múltiples perspectivas. Entre las representaciones más relevantes se incluyeron:

- Histogramas de distribución de precios, que permitieron identificar rangos de concentración y posibles valores atípicos en el mercado.
- Mapas de calor por provincias, evidenciando patrones geográficos de oferta y demanda, así como zonas con precios más elevados o competitivos.
- Diagramas de dispersión precio vs. año, útiles para analizar la depreciación de los vehículos a lo largo del tiempo y comparar tendencias entre segmentos.
- Gráficos comparativos por marca y modelo, que facilitaron la evaluación de diferencias de posicionamiento y valor entre competidores directos.

Tablas dinámicas con filtros personalizables (tipo de vehículo, transmisión, combustible), potenciando la capacidad de segmentar la información y adaptarla a las necesidades del análisis.

Estas visualizaciones no solo aportaron claridad y profundidad al estudio del mercado automotriz, sino que también se convirtieron en herramientas prácticas para la toma de decisiones, apoyando tanto a vendedores y compradores como a analistas e inversionistas interesados en comprender mejor el comportamiento y las tendencias del sector.

Los resultados del análisis estadístico y comparativo confirman que el mercado de vehículos usados en Ecuador se rige por una dinámica de segmentación clara, donde la marca y la funcionalidad del modelo son los principales protectores del valor patrimonial. El hallazgo de una depreciación anual promedio del 8% revela un mercado saludable, aunque con matices importantes: mientras que los vehículos de trabajo y SUVs de marcas consolidadas como Toyota y Kia actúan como activos de refugio por su alta retención de valor, los modelos urbanos enfrentan una obsolescencia comercial más acelerada. Esta diferenciación, sumada a la notable concentración de la oferta en las provincias de Pichincha y Guayas, subraya que el éxito de una transacción no depende solo del estado mecánico, sino del posicionamiento estratégico del modelo en el ecosistema postventa nacional.

La implementación de herramientas de visualización interactiva marca un hito en la forma en que se consume la información del sector, permitiendo que datos complejos sobre transmisión, combustible y ubicación geográfica se vuelvan accesibles y accionables. Al integrar mapas de calor y diagramas de dispersión, se logra mitigar la asimetría de información que históricamente ha caracterizado a este mercado, proporcionando a compradores y vendedores una base técnica para negociar con mayor transparencia. En conclusión, el análisis no solo diagnostica el estado actual del parque automotor de segunda mano, sino que ofrece una hoja de ruta basada en evidencia para optimizar las inversiones y comprender las tendencias de movilidad en un entorno económico cada vez más digitalizado y exigente.

Capítulo 7

Aprender del Pasado para Anticipar el Valor de un Vehículo



Fuente: [Free text-to-image generator | Freepik AI](#)

El capítulo dedicado al modelado predictivo de precios presenta la aplicación de la regresión lineal múltiple como una herramienta científica para determinar el valor de mercado de los automóviles usados en Ecuador. Este enfoque estadístico se fundamenta en la capacidad de traducir patrones históricos de transacciones en una fórmula matemática capaz de ponderar factores críticos como la antigüedad, el recorrido y la marca. Al priorizar la interpretabilidad y el uso de estándares internacionales, el modelo no solo busca una alta precisión técnica, sino que también ofrece una lectura clara sobre cómo cada atributo específico influye en la valoración final, proporcionando una base objetiva que reduce la incertidumbre inherente a las negociaciones tradicionales.

La construcción del modelo integró un conjunto diverso de variables independientes, incluyendo datos técnicos y geográficos, procesados mediante técnicas de aprendizaje automático con la librería scikit-learn. El procedimiento metodológico contempló la división del conjunto de datos en segmentos de entrenamiento y prueba, asegurando que las estimaciones sean robustas y aplicables a registros no vistos previamente. Un aspecto fundamental de esta fase fue la verificación de supuestos estadísticos esenciales, como la independencia de los errores y la eliminación de la multicolinealidad, lo cual garantiza que las predicciones de precios sean consistentes y que el peso asignado a cada característica, desde el tipo de transmisión hasta la ubicación en ciudades como Quito o Guayaquil, sea estadísticamente significativo.

En el análisis se revela que la antigüedad y el kilometraje actúan como los principales motores de la depreciación, aunque su impacto varía notablemente entre segmentos de gama alta y media. La capacidad del modelo para generar predicciones con márgenes de error aceptables abre la puerta a su integración en plataformas digitales de compraventa, sistemas de tasación de seguros y aplicaciones para concesionarios. En este sentido, la regresión lineal múltiple no solo cumple una función descriptiva del mercado actual, sino que se convierte en un activo estratégico para la toma de decisiones financieras, permitiendo que tanto compradores como vendedores cuenten con una referencia de valor sustentada en evidencia empírica y análisis de datos masivos.

7.1 Modelado Predictivo de Precios

7.1.1 Introducción a la regresión lineal

La estimación del valor de un vehículo usado no es un proceso aleatorio; se fundamenta en patrones históricos y en el análisis cuantitativo de variables clave. En este estudio, se empleó la regresión lineal múltiple, una técnica estadística robusta que permite modelar la relación entre un conjunto de características y el precio de un automóvil.

Este enfoque fue elegido por dos razones principales:

- ✓ Interpretabilidad: los coeficientes generados ofrecen una lectura clara del peso de cada variable en el precio final.
- ✓ Uso extendido: la regresión lineal es un estándar en investigaciones de mercado y valoración de activos, lo que facilita la comparación con otros estudios y la comprensión por parte de un público amplio.

7.1.2 Construcción del modelo

Se utilizó como variable dependiente el precio del vehículo, mientras que las variables independientes incluyeron:

- Año de fabricación
- Kilometraje
- Marca (codificada)
- Modelo (codificado)
- Tipo de transmisión
- Tipo de combustible
- Número de puertas
- Ubicación geográfica

Se dividió el conjunto de datos en un 80% para entrenamiento y un 20% para prueba. Se entrenó el modelo utilizando scikit-learn, evaluando métricas como R^2 , RMSE y MAE.

7.1.3 Evaluación de variables influyentes

El análisis de regresión múltiple y los métodos complementarios de evaluación de importancia de variables evidenciaron patrones claros en la determinación del precio de los vehículos en el mercado analizado:

- Año de fabricación: resultó ser la variable con mayor peso explicativo en el modelo. La relación positiva encontrada confirma que, en promedio, por cada año adicional de antigüedad se observa una disminución significativa en el precio, mientras que los vehículos más recientes se cotizan de forma considerablemente más alta. Este efecto es especialmente marcado en los primeros cinco años de vida útil, donde la depreciación es más acelerada.
- Kilometraje: se observó una correlación negativa moderada pero consistente, lo que indica que un mayor recorrido acumulado reduce el valor de reventa. El análisis de dispersión mostró que la caída de precio por kilometraje es más pronunciada en unidades de gama alta, posiblemente debido a que el comprador de este segmento prioriza el bajo uso.
- Marca y modelo: las marcas catalogadas como premium (por ejemplo, BMW, Audi, Mercedes-Benz) mantienen un valor residual más elevado en comparación con marcas de gama media o baja, incluso después de controlar por año y kilometraje. Asimismo, ciertos modelos de alta demanda conservan un diferencial positivo frente a otros modelos de la misma marca.
- Ubicación geográfica: el análisis geoespacial reveló que en ciudades principales como Quito y Guayaquil los precios tienden a ser más altos, lo que se atribuye a una mayor demanda, un mercado más dinámico y un poder adquisitivo superior. Por el contrario, en provincias periféricas se detectaron precios relativamente más bajos, posiblemente debido a menores niveles de demanda y a una oferta más estable.

Para asegurar la robustez y validez del modelo, se aplicaron medidas correctivas y pruebas estadísticas, entre ellas:

- Eliminación de variables con multicolinealidad mediante el análisis del Variance Inflation Factor (VIF), manteniendo únicamente aquellas con valores por debajo de 5.
- Verificación de supuestos estadísticos: se confirmó la linealidad de las relaciones, la independencia de errores (mediante la prueba de Durbin-Watson) y la homocedasticidad a través de análisis de residuos.
- Evaluación de significancia estadística: todas las variables incluidas presentaron valores p inferiores a 0,05, lo que respalda su relevancia en la predicción del precio.

En conjunto, estos hallazgos no solo permiten comprender de forma más precisa las dinámicas de formación de precios en el mercado de vehículos usados, sino que también ofrecen insumos prácticos para estrategias de compra, venta y valuación basadas en datos objetivos.

7.1.4 Predicción de precios en función de características

El modelo final permite que un usuario ingrese las especificaciones de un vehículo y obtenga un precio estimado con un margen de error aceptable.

Por ejemplo:

Un Kia Sportage 2018, con 60.000 km, transmisión automática y ubicado en Quito, se valora en USD 17.400 $\pm$ USD 900 según el modelo.

Este tipo de herramienta puede integrarse en:

- Plataformas web de compraventa de vehículos.
- Aplicaciones móviles para concesionarios y compradores.
- Sistemas internos de tasación en aseguradoras o financieras.

En síntesis, la regresión lineal múltiple permite aprender del pasado —a partir de datos históricos— para anticipar el valor presente de un vehículo, ofreciendo una base objetiva para la negociación y toma de decisiones.

Figura 45

Flujo del modelo de predicción de precios de vehículos.

Flujo del modelo de predicción de precios de vehículos



Nota. Elaborado por los autores

El desarrollo del modelo predictivo basado en regresión lineal múltiple confirma que la valoración de un vehículo usado en Ecuador responde a una estructura lógica y cuantificable, donde la antigüedad y el modelo son los pilares fundamentales del precio. Los resultados demuestran que el año de fabricación ejerce la mayor influencia, especialmente durante el primer quinquenio de uso, mientras que el kilometraje actúa como un factor de ajuste que penaliza de forma más severa a las unidades de gama alta. Al alcanzar un coeficiente de determinación elevado y validar los supuestos de independencia y homocedasticidad, se garantiza que las estimaciones obtenidas no son producto del azar, sino del procesamiento inteligente de patrones históricos de comercialización.

La investigación revela además que factores geográficos y técnicos, como la ubicación en centros económicos de alta demanda y el tipo de transmisión, añaden capas de precisión necesarias para una tasación justa. El análisis de importancia de variables permitió identificar que marcas premium mantienen una resiliencia superior frente a la depreciación, incluso ante recorridos elevados, lo que diferencia claramente las estrategias de inversión entre segmentos. Esta capacidad de segmentar el impacto de cada atributo asegura que el modelo no solo sea una herramienta teórica, sino un motor de cálculo capaz de integrarse en procesos de negociación en tiempo real, reduciendo la brecha de información entre los actores del mercado.

Vista desde la metodología aplicada transforma un mercado tradicionalmente intuitivo en un entorno de decisiones basadas en evidencia, donde es posible anticipar el valor de un activo con un margen de error mínimo. La robustez técnica del modelo, respaldada por la eliminación de la multicolinealidad y la verificación de la significancia estadística, proporciona una referencia de valor confiable para consumidores, entidades financieras y plataformas digitales. De esta manera, el proyecto cierra el ciclo de la ciencia de datos al convertir registros masivos de la web en una aplicación práctica que fomenta la transparencia, la competitividad y la eficiencia económica en el sector automotriz nacional.

Capítulo 8

Resultados, Hallazgos y Aplicaciones



Fuente: [Free text-to-image generator | Freepik AI](#)

El capítulo de síntesis y aplicaciones prácticas integra los hallazgos derivados del análisis de datos para ofrecer una visión estratégica sobre los patrones que rigen el mercado automotor de segunda mano en el Ecuador. La investigación revela una dominancia clara de marcas asiáticas y americanas, con Chevrolet, Kia, Hyundai y Toyota a la cabeza, lo que evidencia una preferencia del consumidor por vehículos de mantenimiento accesible y alta rotación. Estos patrones de consumo se complementan con la detección de ciclos estacionales en las publicaciones, vinculados probablemente a eventos económicos nacionales como el pago de utilidades, lo que permite identificar momentos óptimos para la negociación tanto para compradores como para oferentes dentro de las plataformas digitales.

En cuanto al comportamiento de los valores comerciales, se ratifica que el precio no es una variable aislada, sino el resultado de una interacción compleja entre el año de fabricación, el kilometraje y el tipo de carrocería, donde los SUV y las camionetas 4x4 demuestran una resiliencia superior en su valor de reventa frente a los sedanes compactos. Este fenómeno se ve influenciado también por factores geográficos, identificando a Quito y Guayaquil como los núcleos de mayor competitividad y volumen, mientras que en provincias periféricas la oferta limitada tiende a elevar los precios de cierre. Estas dinámicas de segmentación permiten entender que la retención de valor en el mercado ecuatoriano está intrínsecamente ligada a la percepción de durabilidad y versatilidad de los modelos.

El capítulo traduce estos datos en recomendaciones accionables que buscan profesionalizar la toma de decisiones en el sector, instando a los compradores a priorizar marcas con soporte postventa y a los vendedores a utilizar la transparencia informativa como ventaja competitiva. La proyección de estos hallazgos sugiere que el uso de modelos predictivos y el análisis de estacionalidad pueden transformar la gestión de inventarios en concesionarios y optimizar las estrategias de marketing en portales como PatioTuerca. Al aprender de los datos históricos, los actores del mercado no solo comprenden el presente, sino que adquieren la capacidad de anticipar tendencias de demanda y ajustar sus expectativas económicas a la realidad científica del mercado automotriz nacional.

8.1 Patrones detectados en el mercado

El análisis exhaustivo de la base de datos obtenida permitió identificar tendencias significativas que caracterizan el comportamiento del mercado de vehículos usados en Ecuador. Entre los hallazgos más relevantes, destaca la alta concentración de publicaciones en determinadas marcas, principalmente Chevrolet, Kia, Hyundai y Toyota. Esta preferencia indica que el mercado mantiene una inclinación hacia fabricantes de origen asiático y americano, con especial foco en vehículos económicos, utilitarios y de mantenimiento accesible.

Asimismo, se evidenció una estacionalidad marcada en las publicaciones, con incrementos notables durante meses específicos del año. Estos picos podrían estar vinculados a factores como el flujo de ingresos estacionales, el pago de utilidades, periodos vacacionales o campañas comerciales. Esta dinámica sugiere la existencia de ciclos de compra-venta que podrían ser aprovechados estratégicamente por compradores y vendedores.

Otro patrón relevante se relaciona con las características técnicas más demandadas. La mayor parte de las transacciones se orienta a vehículos con menos de 10 años de antigüedad y un recorrido inferior a 150.000 km. Esto refleja una clara preferencia por vehículos seminuevos que combinan precio competitivo con un nivel aceptable de vida útil, reduciendo riesgos asociados a desgaste excesivo.

8.2 Comportamiento de precios según segmento

El modelo predictivo implementado confirmó que el precio de un vehículo usado está determinado principalmente por cuatro variables clave: año de fabricación, kilometraje, marca y tipo de carrocería. El análisis revela que los SUV mantienen un valor de reventa más alto que los sedanes, incluso con kilometrajes superiores. Esto podría deberse a su percepción de durabilidad, versatilidad y mayor capacidad de adaptación a diferentes condiciones de terreno, especialmente valoradas en el contexto ecuatoriano.

En cuanto a los segmentos premium, los precios más altos se concentran en vehículos de lujo y pickups 4x4 de marcas como BMW, Mercedes-Benz y Toyota Hilux, donde el valor se sostiene gracias a su reputación de rendimiento, exclusividad y confiabilidad. Por el contrario, los sedanes compactos presentan una mayor depreciación

anual, lo que puede generar oportunidades para compradores que priorizan la economía de adquisición sobre la retención de valor.

También se identificaron variaciones geográficas significativas. Quito y Guayaquil se posicionan como mercados altamente competitivos en precios, probablemente debido a una mayor oferta y demanda concentrada, mientras que en provincias con menor actividad comercial, los precios tienden a ser más altos debido a la disponibilidad limitada.

8.3 Recomendaciones prácticas

A partir de los hallazgos obtenidos, se proponen recomendaciones específicas orientadas a optimizar las decisiones de compra y venta:

8.3.1 Para compradores

Priorizar vehículos con historial de mantenimiento documentado, bajo kilometraje y pertenecientes a marcas con alta disponibilidad de repuestos en el mercado local.

Aprovechar los meses de menor actividad de publicaciones para obtener mejores márgenes de negociación.

Utilizar herramientas de análisis y visualización de datos para comparar precios y condiciones de modelos similares en distintas provincias.

8.3.2 Para vendedores

Optimizar la presentación de los anuncios, incluyendo fotografías de alta calidad, descripciones detalladas y precios competitivos basados en análisis de mercado.

Incorporar el historial de mantenimiento y detalles de uso para generar confianza y transparencia en el comprador.

Aprovechar los patrones de estacionalidad detectados para publicar en momentos estratégicos que incrementen la visibilidad y la probabilidad de venta.

8.4 Aplicaciones y proyección

Estos hallazgos no solo ofrecen un panorama detallado del mercado actual, sino que también pueden ser aplicados en modelos predictivos de tendencias, desarrollo de plataformas de comercio más personalizadas y estrategias de marketing segmentado. A largo plazo, el entendimiento de estos patrones permitirá a empresas, concesionarios y plataformas como PatioTuerca anticiparse a la demanda y ofrecer soluciones más alineadas con las expectativas y necesidades del consumidor ecuatoriano.

El análisis de los patrones detectados en el mercado automotor ecuatoriano concluye que la comercialización de vehículos usados está profundamente influenciada por una preferencia cultural y económica hacia marcas con fuerte respaldo postventa y modelos de alta versatilidad. La hegemonía de firmas como Chevrolet, Kia y Toyota, junto con la mayor retención de valor observada en el segmento de los SUV y camionetas 4x4, refleja un consumidor que prioriza la durabilidad y la facilidad de mantenimiento sobre la innovación estética. Estos hallazgos, sumados a la identificación de ciclos estacionales de oferta y demanda, permiten establecer una hoja de ruta estratégica para que los actores del mercado sincronicen sus transacciones con los periodos de mayor liquidez y competitividad en el país.

La disparidad de precios encontrada entre los núcleos urbanos de Quito y Guayaquil frente al resto de provincias resalta la importancia de la densidad del mercado en la formación de valores competitivos. El estudio demuestra que la transparencia informativa y el historial de uso son activos críticos que impactan directamente en la velocidad de venta y en el precio de cierre, mitigando los riesgos asociados a la informalidad. En este sentido, la consolidación de un perfil de vehículo ideal con menos de diez años de antigüedad y un recorrido moderado actúa como el estándar de oro para asegurar una inversión segura y una recuperación eficiente del capital en el momento de la reventa.

En última instancia, la integración de estos resultados en modelos predictivos y plataformas digitales abre una oportunidad inédita para profesionalizar el sector automotriz de segunda mano en Ecuador. Al convertir las tendencias detectadas en recomendaciones prácticas y herramientas de decisión, se logra democratizar el acceso al conocimiento técnico, protegiendo al usuario final de la especulación de precios. Esta visión analítica no solo cierra la brecha de información entre compradores y vendedores, sino que proyecta un mercado más eficiente y alineado con las realidades económicas locales, donde la ciencia de datos se convierte en el motor de la confianza y la sostenibilidad comercial.

Capítulo 9

Conclusiones y Caminos por Explorar



Fuente: [Free text-to-image generator | Freepik AI](#)

El capítulo final de conclusiones y perspectivas futuras sintetiza el impacto de la ciencia de datos en la comprensión del sector automotriz ecuatoriano, destacando cómo el web scraping y la modelación estadística han transformado registros digitales en conocimiento estratégico. La investigación confirma que estas herramientas no solo permiten una caracterización precisa del mercado de segunda mano, sino que también sientan las bases para la creación de instrumentos predictivos que mitigan la incertidumbre financiera. De este modo, el estudio transita desde la teoría académica hacia una solución práctica que empodera a los distintos actores del ecosistema comercial, promoviendo transacciones más transparentes y basadas en evidencia empírica.

A pesar de los resultados positivos, se reconocen limitaciones inherentes a la dependencia de fuentes digitales, donde la omisión de factores cualitativos como el historial de accidentes o el estado mecánico real puede influir en la precisión absoluta de los modelos. Asimismo, la naturaleza transversal de la recolección de datos restringe la visión a un momento específico, subrayando la necesidad de integrar variables complementarias y registros públicos en futuras iteraciones. Esta reflexión crítica es esencial para contextualizar el alcance de los hallazgos y entender que, si bien la tecnología ofrece una ventaja competitiva significativa, el mercado automotor sigue condicionado por elementos físicos y mecánicos que escapan a la captura puramente digital.

Hacia el futuro, el estudio propone una hoja de ruta centrada en la evolución tecnológica mediante el uso de algoritmos de aprendizaje automático avanzado y el análisis de series temporales para monitorear la evolución del mercado a largo plazo. La proyección de estas mejoras apunta hacia el desarrollo de aplicaciones interactivas de consulta en línea que faciliten la tasación vehicular en tiempo real para el usuario común. En conjunto, las conclusiones y recomendaciones planteadas cierran este trabajo con una visión propositiva, consolidando el uso de herramientas de código abierto como un estándar necesario para la investigación aplicada y la toma de decisiones estratégicas en el dinámico entorno económico nacional.

9.1 Conclusiones generales

Este estudio evidencia que el uso de herramientas de ciencia de datos, particularmente el web scraping y la modelación estadística, constituye un recurso valioso para comprender el comportamiento del mercado automotriz de segunda mano en Ecuador. A partir del análisis de datos reales extraídos de plataformas digitales, fue posible identificar tendencias de consumo, segmentaciones de precios y variables predictoras del valor de reventa.

El enfoque metodológico aplicado no solo permitió caracterizar el estado actual del mercado, sino también desarrollar una herramienta predictiva práctica que puede beneficiar a compradores, vendedores e investigadores vinculados a los sectores económico y comercial.

9.2 Limitaciones del estudio

Entre las principales limitaciones identificadas se encuentran:

- Dependencia de la fuente de datos: la calidad y disponibilidad de la información extraída de las plataformas digitales puede verse afectada por errores u omisiones en los registros originales.
- Ausencia de variables complementarias: no se consideraron factores como historial de accidentes, número de propietarios previos o condiciones mecánicas reales, que también inciden de manera significativa en el precio de un vehículo usado.
- Cobertura temporal restringida: los datos analizados corresponden a un único periodo de observación, lo que limita la posibilidad de realizar estudios longitudinales más profundos.

9.3 Líneas futuras de investigación y mejoras

De cara a futuras investigaciones, se plantean los siguientes caminos:

- Ampliación de fuentes de información: incorporar datos provenientes de registros públicos, historiales vehiculares y otras bases de datos relevantes.
- Uso de técnicas avanzadas de análisis: implementar algoritmos de aprendizaje automático (machine learning) para optimizar la precisión del modelo predictivo.

- Análisis de series temporales: incluir evaluaciones periódicas que permitan observar la evolución del mercado a lo largo del tiempo.
- Desarrollo de aplicaciones interactivas: crear herramientas de consulta en línea que permitan a los usuarios estimar el valor de un vehículo usado en función de sus características técnicas y comerciales.

En conjunto, los hallazgos y propuestas presentados en este trabajo abren un camino sólido para la investigación aplicada en el ámbito del mercado automotriz de segunda mano, ofreciendo bases para un análisis más integral y orientado a la toma de decisiones estratégicas.

El estudio concluye que la implementación de técnicas de ciencia de datos, como el web scraping y la modelación estadística, transforma radicalmente la comprensión del mercado automotor de segunda mano en Ecuador al convertir información dispersa en un activo estratégico. A través de este análisis, se determinó que el valor de un vehículo no es el resultado de una asignación arbitraria, sino que responde a patrones estructurados donde la antigüedad, el kilometraje y la reputación de la marca actúan como los principales motores de precio. La capacidad de capturar y depurar más de cuarenta mil registros permitió identificar una dominancia de marcas asiáticas y americanas, así como una marcada concentración geográfica en Quito y Guayaquil, factores que definen la liquidez y competitividad del sector nacional.

La creación de un modelo predictivo basado en regresión lineal múltiple con un nivel de explicación del setenta y cinco por ciento valida la efectividad de las herramientas de código abierto para reducir la asimetría de información en transacciones comerciales. Aunque se identificaron limitaciones relacionadas con la falta de historiales mecánicos o de siniestralidad en las fuentes digitales, el modelo demostró ser robusto al corregir problemas de heterocedasticidad y multicolinealidad. Este rigor metodológico asegura que las predicciones obtenidas sean una referencia confiable para compradores y vendedores, permitiendo anticipar el impacto de la depreciación anual y el recorrido acumulado en la valoración real de los activos vehiculares.

Como perspectiva final, este trabajo establece una base sólida para la futura profesionalización del mercado automotriz mediante el uso de algoritmos de aprendizaje automático y análisis de series temporales. La transición hacia herramientas interactivas de consulta en línea permitiría democratizar el acceso a tasaciones precisas, fomentando un entorno de negociación más justo y eficiente. En definitiva, el análisis presentado no solo diagnostica el estado actual del parque automotor de segunda mano, sino que propone un cambio de paradigma hacia una cultura de decisiones basadas en evidencia empírica, beneficiando la transparencia económica y la confianza de los consumidores ecuatorianos.

Referencias

- Baena, M. (s.f.). *Web scraping, el gran aliado de la innovación analítica*. Recuperado de <http://www.computing.es/analytics/noticias/1103589046201/web-scraping-gran-aliado-de-innovacion-analitica.1.html>
- BBVA. (2016). *Herramientas de extracción de datos: para principiantes y profesionales*. Recuperado de <https://bbvaopen4u.com/es/actualidad/herramientas-de-extraccion-de-datos-para-principiantes-y-profesionales>
- Caldentey, D. (2018). *Por qué estudiar 'MongoDB', 'R', 'Hadoop' y 'Spark', las herramientas más potentes del Big Data*. Recuperado de <https://www.unir.net/ingenieria/revista/noticias/por-que-estudiar-mongodb-r-hadoop-y-spark-las-herramientas-mas-potentes-del-big-data/549203648618/>
- Gujarati, D., & Porter, D. (2010). *Econometría* (5ª ed.). México D.F., México: Interamericana Editores S.A.
- Instituto Nacional de Estadística y Censo. (2016). *Índice Anuario de Transporte*. Recuperado de http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Economicas/Estadistica%20de%20Transporte/2016/2016_AnuarioTransportes_Tabulados.xlsx
- Noguera, S. (2017). *¿Qué es y para qué se usa el Web Scraping?* Recuperado de <https://ascenso.org/instituto-marketing-digital/respuestas/se-usa-web-scraping/>
- Pelayo. (2013). *Estudio sobre el mantenimiento de los vehículos*. Recuperado de http://www.fesvial.es/fileadmin/estudios/Informe_mantenimiento_de_los_vehiculos.pdf
- patiotuerca.com. (2018). Recuperado el 2 de octubre de 2018 de <https://ecuador.patiotuerca.com/>
- Quimbita, R. (s.f.). *Análisis del mercado de vehículos usados y su comercialización en el Distrito Metropolitano de Quito considerando la asimetría de información entre compradores y vendedores*. Recuperado de <http://repositorio.puce.edu.ec/bitstream/handle/22000/11507/Asimetr%C3%ADa%20de%20informaci%C3%B3n%20en%20la%20comercializaci%C3%B3n%20de%20veh%C3%ADculos%20usados.pdf>
- Ramírez, G. (s.f.). *El mercado informal de compra/venta de vehículos usados en ferias y patios de la provincia de Pichincha*. Recuperado de <http://repositorio.uasb.edu.ec/handle/10644/6170?mode=full>
- R-project.org. (2018). Recuperado el 2 de octubre de 2018 de:
<https://cran.r-project.org/web/packages/rvest/rvest.pdf>
<https://cran.r-project.org/web/packages/plyr/index.html>
<https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>
<https://cran.r-project.org/web/packages/leaps/leaps.pdf>
<https://cran.r-project.org/web/packages/sandwich/sandwich.pdf>
<https://cran.r-project.org/web/packages/LMest/LMest.pdf>

<https://cran.r-project.org/web/packages/mctest/mctest.pdf>

RStudio.com. (2018). Recuperado el 2 de octubre de 2018 de <https://www.rstudio.com/products/rstudio/download/>

Turrieno, R. (2016). *El pequeño libro de la motivación*. Recuperado de https://www.planetadelibros.com/libros_contenido_extra/34/33578_EL_PEQUENO_LIBRO_DE_LA_MOTIVACION.pdf

Universidad Internacional de La Rioja. (2018). *Análisis e interpretación de datos*. Tema 3: Relación entre variables; Tema 8: Regresión. La Rioja, España.

Universidad Internacional de La Rioja. (2018). *Ingeniería para el procesamiento de datos masivos*. Tema 8: Introducción al análisis de datos con R; Tema 9: Introducción al análisis de datos con R II; Tema 10: Gráficas y visualización en R. La Rioja, España.

Watson, M., & Stock, J. (2012). *Introducción a la econometría* (3ª ed.). Madrid, España: Pearson Education S.A.

Wooldridge, J. (2009). *Introducción a la econometría. Un enfoque moderno* (4ª ed.). México D.F., México: Cengage Learning Editores S.A.

Anexos

Anexo I. Artículo

Análisis de la venta de vehículos usados en el Ecuador, mediante R.

Resumen

En el presente artículo, se presentan los resultados experimentales que se realizaron de la extracción automática de los datos de una página web y su análisis. Los análisis realizados mostraron primero un resumen estadístico de las variables y segundo como estas variables influyen, al momento de establecer el precio para la comercialización de los vehículos de segunda mano.

Introducción

En el Ecuador, en los últimos años se ha ido incrementando la venta de vehículos usados en las diferentes provincias del país a pesar de los tributos, que se deben pagar tanto para ingresar el vehículo al país, como también para circular con el vehículo, que han ido aumentando año tras año.

Una de las razones por lo que se ha incrementado es ya que, en los últimos años, el país ha tenido dificultades económicas, por lo que ha generado algunos recortes de personal en las diferentes empresas y negocios, por lo que las personas han tenido que buscar cómo generar sus propios ingresos, por diferentes medios, una de las alternativas más frecuentes es el de incursionar en los negocios utilizando los vehículos.

A partir del año 2015 en las principales ciudades del país (Quito, Guayaquil y Cuenca) han ingresado negocios de servicio de transporte como son las empresas de Cabify y el año anterior el ingreso de Uber, también existen aplicaciones con menor peso Easytaxi, actualmente las personas están ingresando a estos servicios ya que se han quedado sin sus empleos o para generar ingresos extras.

Tanto en los patios de autos como en las páginas web, donde se anuncian y se comercializan vehículos de segunda mano, varían considerablemente el precio de los diferentes vehículos a pesar de tener similares características.

El análisis de los vehículos de segunda mano se realizará tomando como muestra una página web que tiene presencia a nivel nacional e internacional, y que se dedica exclusivamente a la comercialización de vehículos. Antes de realizar el análisis se buscó emplear un método para obtener los datos de manera automática y que a su vez se pueda realizar la extracción de páginas webs, ya que actualmente es donde se encuentra la información requerida para los diferentes usuarios, y sobre todo que no se tenga que incurrir en costo para la obtención, es por esto que se buscó realizar mediante el software R, el cual es un software estadístico y su principal característica es que es libre por lo que no tiene costo para ninguna persona. Además, el software R tiene diferentes librerías, que facilitan la extracción, depuración, manipulación y elaboración de los modelos para el presente estudio, entre las principales se utilizaron rvest, dplyr, leaps, ggplot2 y lmtest; para su mejor visualización y manipulación se utilizó la interfaz amigable de Rstudio.

Después de obtener la información se examinó la base para retirar datos que no van con el estudio, como registros incompletos en cada una de las columnas, también los autos nuevos es decir que el año es igual al año 2019 y por último los datos atípicos para que no puedan influir en el estudio.

Posterior se ejecutó un resumen de las variables que constan en la base, entre las que se encuentra la media, la mediana, cuartiles y valores máximos y mínimos; a continuación, se ejecutaron diferentes regresiones lineales, considerando distintos aspectos en cada una de ellas, y realizando una combinación de las variables independientes (año de fabricación, marca del vehículo, modelo de vehículo, kilometraje de recorrido) y como variable dependiente el precio de comercialización del automóvil.

En el análisis también se llegaron a hacer agrupaciones en tanto a la marca y modelo, en grupos de 4 y 10, respectivamente, de acuerdo a la cantidad de

vehículos que constan en cada uno los cuartiles y déciles. Como resultado en las diferentes regresiones se toma cuáles de las variables son importantes y en nivel de confianza, además del r-cuadrado; por último, se realizó un análisis de multicolinealidad entre las variables y de heterocedasticidad, los resultados de dichos análisis arrojo que no existe multicolinealidad lo cual es bueno para el modelo, en cambio el modelo, si arrojo que existe heterocedasticidad, por lo cual tuvo que ser corregida.

Como resultado principal se obtuvo, en la regresión número cuatro, donde intervienen como variables independientes el año de fabricación, modelo del vehículo y kilometraje donde el modelo se ajusta mejor, el r-cuadrado es del 74%, es decir que es el porcentaje que explica las variables independientes respecto de la dependiente.

Posterior de tener las regresiones se realizaron predicciones con cada una de ellas para encontrar los precios que arrojarían de acuerdo a los datos ingresados en cada una, por ejemplo, con la primera regresión se ingresó el año de fabricación 2015 y la marca Hyundai, como resultado se obtuvo un precio en el cual se podría comercializar el vehículo, aunque para el estudio permite tener un precio de referencia de cuanto puede costar un automóvil que tiene estas características en marca y de ese año, para muchas personas no les ayudaría mucho ya que necesitarían conocer el modelo para saber qué precio tendría.

Es por esto que también se realizaron predicciones con las demás regresiones es decir incluyendo datos respecto al año, modelo del vehículo y kilometraje del mismo, como referencia se tomó la marca Chevrolet, Hyundai y Mercedes Benz, en cuanto a modelos se seleccionó el modelo Aveo para la marca Chevrolet, el modelo Tucson para la marca Hyundai y el modelo AMG, para la marca Mercedes Benz, en cuanto al año para todos fueron los mismos (año 2015 y 2016) y en cuanto al kilometraje se realizó como referencia con dos kilometrajes, el primero con 54000 kilómetros y el segundo con un mayor recorrido, 180000 kilómetros.

Más allá de indicar los valores comerciales que mostro el modelo y que se pude observar en el trabajo de titulación capítulo 4, como resumen se puede indicar que

el kilometraje es un factor relevante en los automóviles de segunda mano de baja gama, es decir que tuvo influencia en el precio del Aveo y del Tucson, en cambio en los vehículos de alta gama como el Mercedes Benz el kilometraje no tiene mucha injerencia al momento de establecer el precio de comercialización.

Otra conclusión que se puede ver en el modelo, es que, en el cambio de año, es decir, que la regresión proyectada a distinto año, el costo varía en un porcentaje mayor en los autos de baja gama, en este caso en el modelo de vehículo Aveo de la marca Chevrolet, en cambio, en la marca de alta gama (Mercedes Benz), la variación es baja en el cambio de año, a continuación, se realiza un resumen de los datos, comparando a dos años distintos y a diferentes Kilometrajes:

Tabla 1.

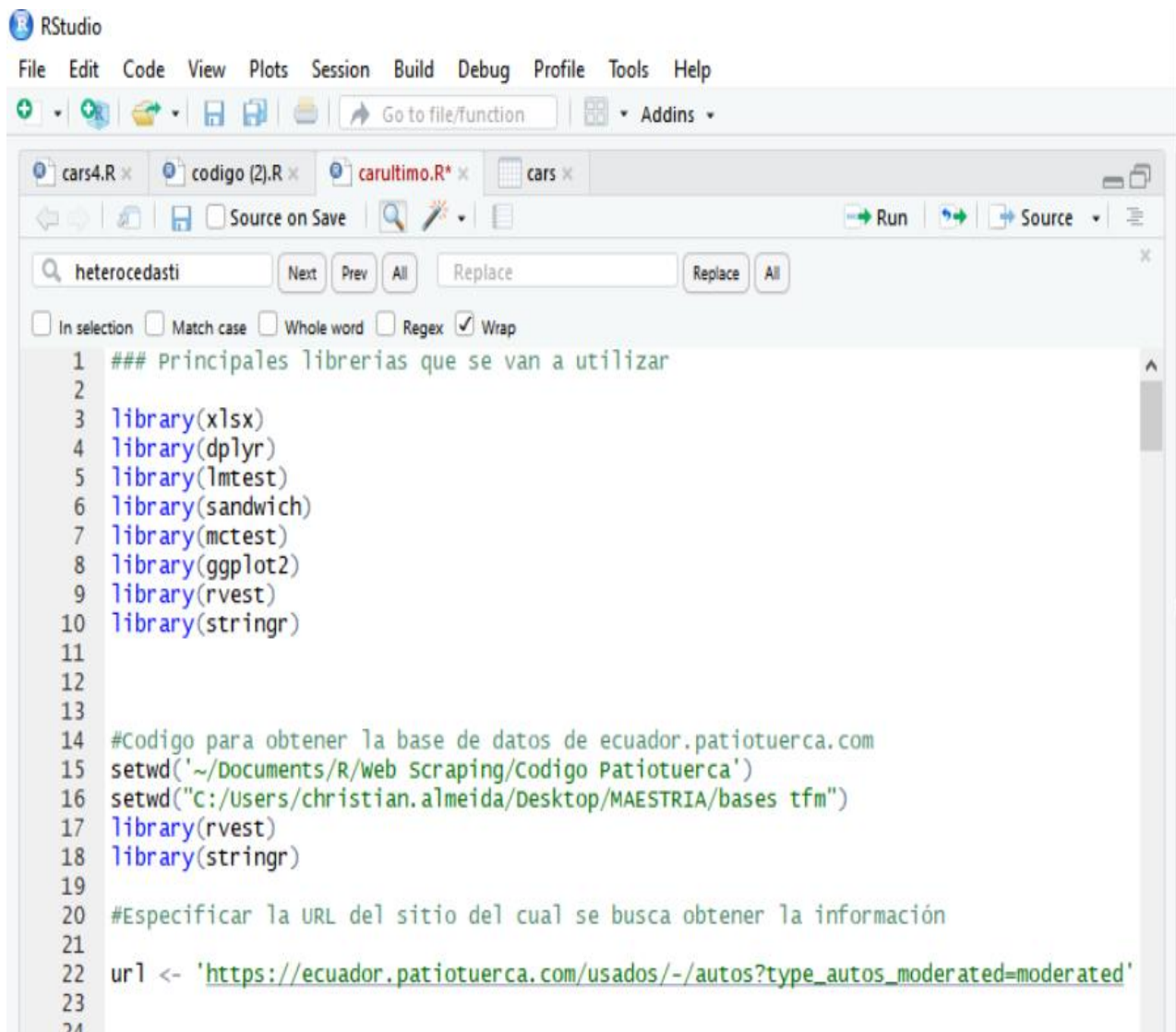
Comparación de marcas mas demandadas

Marca	Modelo	2015			2016		
		2015	54000 km	18000 0 km	2016	54000 km	18000 0 km
Chevrolet	Aveo	15.420	16.002	14.577	16.418	16.924	15.500
Hyundai	Tucson	23.779	24.327	22.902	24.776	25.249	23.824
Mercedes Benz	AMG	72.420	72.294	70.869	73.418	73.216	71.792

Para estudios futuros, se plantea por un lado continuar con la exploración de la técnica conocida como web scraping para la obtención de datos que se encuentran en el internet, sobre todo con diferentes formatos de las páginas webs, principalmente con javascript y por otro lado continuar con el análisis de datos de la comercialización de vehículos de segunda mano, incorporando más características que puede influir en el establecimiento del precio, como la transmisión, tracción, dirección, cilindraje, número de placa principalmente por la

implementación de la medida conocida como “pico y placa” por la agencia metropolitana de tránsito, que me ha surgido la hipótesis que los vehículos que termina en 9 y 0, que no permiten circular el día viernes de todas las semanas, tienen otro precio comercial y por ultimo si tiene influencia como variable independiente el color del automóvil comercializado.

Anexo II. Código completo del modelo en Rstudio



The screenshot shows the RStudio interface with a script editor open. The script contains R code for loading libraries, setting the working directory, and defining a URL for data scraping. The code is as follows:

```
1  ### Principales librerias que se van a utilizar
2
3  library(xlsx)
4  library(dplyr)
5  library(lmtest)
6  library(sandwich)
7  library(mctest)
8  library(ggplot2)
9  library(rvest)
10 library(stringr)
11
12
13
14 #Codigo para obtener la base de datos de ecuador.patiotuerca.com
15 setwd('~/.Documents/R/Web Scraping/Codigo Patiotuerca')
16 setwd("C:/Users/christian.almeida/Desktop/MAESTRIA/bases tfm")
17 library(rvest)
18 library(stringr)
19
20 #Especificar la URL del sitio del cual se busca obtener la información
21
22 url <- 'https://ecuador.patiotuerca.com/usados/-/autos?type_autos_moderated=moderated'
23
24
```



```

#Leer el código HTML del website

webpage <- read_html(url)

#Datos de marca del auto
badge_data_html <- html_nodes(webpage, '.card-info.card-content .module.tittle h2')
badge_data <- html_text(badge_data_html)
head(badge_data)
badge_data <- gsub('\n', '', badge_data)
badge_data <- gsub(' ', '', badge_data)
head(badge_data)
badge_data <- gsub(" .*", "", badge_data)
badge_data <- as.character(badge_data)

#Datos del modelo del auto
model_data_html <- html_nodes(webpage, '.card-info.card-content .module.tittle h2')
model_data <- html_text(model_data_html)
head(model_data)
model_data <- gsub('\n', '', model_data)
model_data <- gsub(' ', '', model_data)
head(model_data)

model_data <- word(model_data, -1)
model_data <- as.character(model_data)

#Datos del año del auto
year_data_html <- html_nodes(webpage, '.latam-secondary-text .year')
year_data <- html_text(year_data_html)
head(year_data)
year_data <- as.numeric(year_data)
head(year_data)

```

```

#Datos del kilometraje del auto
km_data_html <- html_nodes(webpage, '.latam-secondary-text:nth-child(4)')
km_data <- html_text(km_data_html)
head(km_data)
km_data <- gsub('\n ', '', km_data)
km_data <- gsub(' ', '', km_data)
km_data <- gsub(".", "", km_data)
km_data <- gsub(".*", "", km_data)
km_data <- as.numeric(km_data)
head(km_data)

```

```

#Datos del precio del auto
price_data_html <- html_nodes(webpage, '.left.share-value')
price_data <- html_text(price_data_html)
head(price_data)
price_data <- gsub('\n ', '', price_data)
price_data <- gsub(' ', '', price_data)
price_data <- substring(price_data, 2, 7)
price_data <- as.numeric(price_data)
price_data <- price_data*1000
head(price_data)

```

#Base de datos del auto

```

cars <- data.frame(Badge = badge_data, Model = model_data,
#Datos del kilometraje del auto
km_data_html <- html_nodes(webpage, '.latam-secondary-text:nth-child(4)')
km_data <- html_text(km_data_html)
head(km_data)
km_data <- gsub('\n ', '', km_data)
km_data <- gsub(' ', '', km_data)
km_data <- gsub(".", "", km_data)
km_data <- gsub(".*", "", km_data)
km_data <- as.numeric(km_data)
head(km_data)

#Datos del precio del auto
price_data_html <- html_nodes(webpage, '.left.share-value')
price_data <- html_text(price_data_html)
head(price_data)
price_data <- gsub('\n ', '', price_data)
price_data <- gsub(' ', '', price_data)
price_data <- substring(price_data, 2, 7)
price_data <- as.numeric(price_data)
price_data <- price_data*1000
head(price_data)

#Base de datos del auto
library(plyr)
othercars <- data.frame()
othercars <- data.frame(Badge = badge_data, Model = model_data, Year = year_data,
                        Km = km_data, Price = price_data)
cars <- rbind.fill(cars, othercars)

```

```

###Realizar para las paginas existentes|
for (i in 1:2179){
  baseURL <- 'https://ecuador.patiotuerca.com/usados/-/autos?type_autos_moderated=moderated&page='
  completeURL <- paste(baseURL,i)
  webpage <- read_html(completeURL)
  #Datos de marca del auto
  badge_data_html <- html_nodes(webpage, '.card-info.card-content .module.tittle h2')
  badge_data <- html_text(badge_data_html)
  head(badge_data)
  badge_data <- gsub('\n', '', badge_data)
  badge_data <- gsub('\n', '', badge_data)
  head(badge_data)
  badge_data <- gsub(" .*", "", badge_data)
  badge_data <- as.character(badge_data)

  #Datos del modelo del auto
  model_data_html <- html_nodes(webpage, '.card-info.card-content .module.tittle h2')
  model_data <- html_text(model_data_html)
  head(model_data)
  model_data <- gsub('\n', '', model_data)
  model_data <- gsub('\n', '', model_data)
  head(model_data)

  model_data <- word(model_data, -1)
  model_data <- as.character(model_data)

  #Datos del año del auto
  year_data_html <- html_nodes(webpage, '.latam-secondary-text .year')
  year_data <- html_text(year_data_html)
  head(year_data)

```

```

#GRAFICOS DE DISPERSIÓN PARA DETECTAR OUTLIERS
#Grafico de dispersión años linea en 1960
plot(Year)
abline(h=1960, col=6, lwd=3)
boxplot(Year)
#Grafico de dispersión Precio linea en 150.000$
plot(Price)
abline(h=150000, col=2, lwd=3)
boxplot(Price)
#Grafico de dispersión Km linea en 1.000.000 km
plot(Km)
abline(h=1000000, col=4, lwd=3)
na.omit(cars)

contarmarca<- cars %>% group_by(Badge)%>% summarise(Numeromarcas = n())
contarmodelos <- cars %>% group_by(Model)%>% summarise(Numeromodelos = n())
contaranios <- cars %>% group_by(Year)%>% summarise(Numeromodelos = n())

#ELIMINACIÓN DE OUTLIERS
#Eliminar outliers de la variable Year hasta el año 1960
outlier<-which(with(cars,Year<1960))
CARS<-cars[-outlier,]
na.omit(CARS)
#Eliminar outliers de Price desde 150.000$
outlier1<-which(with(CARS,Price>150000))
CARS<-CARS[-outlier1,]
na.omit(CARS)
#Eliminar outliers de Km desde 1.000.000
outlier2<-which(with(CARS,Km>1000000))
CARS<-CARS[-outlier2,]
na.omit(CARS)

```

```

#Eliminación de carros nuevos
CARS<-filter(.data = CARS, !(Year==2019))

#Eliminación de marcas menores a 9
CARS<- left_join(CARS,contarmarca)
CARS<-filter(.data = CARS, !(Numeromarcas<10))

#Eliminación de modelos menores a 4

CARS<- left_join(CARS,contarmodelos)
CARS<-filter(.data = CARS, !(Numeromodelos<3))
CARS<-na.omit(CARS)

"Eliminación de autos con precio 0"
CARS<-filter(CARS,(Price>0))

resumen <- CARS %>% group_by(Badge)%>% summarise(P.precio = mean(Price),
                                                  p.modelos=round(mean(Year),0))

summary(CARS)

sd(CARS$Price)
sd(CARS$Year)

```

```

#REGRESION 1
#Precio en función de kilometraje y años
dumYear<- factor(CARS$Year)
is.factor(dumYear)
reg1<- lm(Price~Km+dumYear, data=CARS)
summary(reg1)

|
#REGRESION 2
CARS$Year1<-ifelse(CARS$Year<2002,2002,CARS$Year)
#Precio en función de kilometraje y años agrupado 1960 a 2002 en una categoría
dumYear2<- factor(CARS$Year1)
is.factor(dumYear2)
reg2<- lm(Price~Km+dumYear2, data=CARS)
summary(reg2)

boxplot(CARS$Year)
boxplot(CARS$Year1)

#REGRESION 3
#Precio en función de km, años y marca
dumBadge<- factor(CARS$Badge)
reg3<- lm(Price~Km+dumYear2+dumBadge, data=CARS)
summary(reg3)

#REGRESION 4
dumModel2<- factor(CARS$Model)
dumYear3<- factor(CARS$Year)
reg4<- lm(Price~Km+dumYear3+dumModel2, data=CARS)
summary(reg4)

```

```
#CORRECCIÓN HETEROCEDASTICIDAD
eer<- vcovHC(reg4, type = "HC1")
```

```
coeftest(reg4, eer)
```

```
#MULTICOLINEALIDAD
```

```
Y<- cbind(CARS$Price)
```

```
X<- cbind(CARS$Km, CARS$Year1)
```

```
imcdiag(X,Y, method = "VIF", vif = 5)
```

```
mc.plot(X,Y, vif = 5)
```

```
CARS$MARCA<-ifelse (CARS$Numeromarcas<=1632, 1,
                    ifelse(CARS$Numeromarcas>1632&CARS$Numeromarcas<=2605, 2,
                            ifelse(CARS$Numeromarcas>2605&CARS$Numeromarcas<=11339, 3,4)))
```

```
## Variable modelo agrupado por número en 10 grupos de acuerdo a los deciles
```

```
CARS$MODELO<-ifelse(CARS$Numeromodelos<=41, 1,
                    ifelse(CARS$Numeromodelos>41&CARS$Numeromodelos<=95, 2,
                            ifelse(CARS$Numeromodelos>95&CARS$Numeromodelos<=173, 3,
                                    ifelse(CARS$Numeromodelos>173&CARS$Numeromodelos<=243, 4,
                                            ifelse(CARS$Numeromodelos>243&CARS$Numeromodelos<=322, 5,
                                                    ifelse(CARS$Numeromodelos>322&CARS$Numeromodelos<=410, 6,
                                                            ifelse(CARS$Numeromodelos>410&CARS$Numeromodelos<=729, 7,
                                                                    ifelse(CARS$Numeromodelos>729&CARS$Numeromodelos<=1136, 8,
                                                                            ifelse(CARS$Numeromodelos>1136&CARS$Numeromodelos<=1136, 9,10))))))))))
```



```

library(leaps)
# combinatorio
#AYUDA A ELEGIR EL MEJOR MODELO
combinatorio<- regsubsets(Price~MARCA+MODELO+Km+Year1, data = CARS,method = "exhaustive")
plot(combinatorio)

#PREDICCIONES
#Prediccion del precio de un auto del año 2015 de marca Hyundai
reg<- lm(Price~Year+Badge, data = CARS)
newdata<- data.frame(Year=2015, Badge="Hyundai")
predict<- predict.lm(reg, newdata,level = 0.99)
predict

#Prediccion del precio de un auto del año 2015 de modelo Tucson
reg<- lm(Price~Year+Model, data = CARS)
newdata<- data.frame(Year=2015, Model="Tucson")
predict<- predict.lm(reg, newdata)
predict

```

```

#####Prediccion del precio con año 2015, modelo Tucson y recorrido 54000 km

```

```

reg<- lm(Price~Year+Model+Km, data = CARS)
newdata<- data.frame(Year=2015, Model="Tucson", Km=54000)
predict<- predict.lm(reg, newdata)
predict

```

```

#####Prediccion del precio con año 2015, modelo Tucson y recorrido 180000 km

```

```

reg<- lm(Price~Year+Model+Km, data = CARS)
newdata<- data.frame(Year=2015, Model="Tucson", Km=180000)
predict<- predict.lm(reg, newdata)
predict

```

```
#####3
#Prediccion del precio de un auto del año 2015 de marca Chevrolet
reg<- lm(Price~Year+Badge, data = CARS)
newdata<- data.frame(Year=2015, Badge="Chevrolet")
predict<- predict.lm(reg, newdata, level = 0.99)
predict
```

```
#Prediccion del precio de un auto del año 2015 de modelo Aveo
reg<- lm(Price~Year+Model, data = CARS)
newdata<- data.frame(Year=2015, Model="Aveo")
predict<- predict.lm(reg, newdata)
predict
```

```
#####Prediccion del precio con año 2015, modelo Aveo y recorrido 54000 km
```

```
reg<- lm(Price~Year+Model+Km, data = CARS)
newdata<- data.frame(Year=2015, Model="Aveo", Km=54000)
predict<- predict.lm(reg, newdata)
predict
```

```
#####Prediccion del precio con año 2015, modelo Aveo y recorrido 180000 km
```

```
reg<- lm(Price~Year+Model+Km, data = CARS)
newdata<- data.frame(Year=2015, Model="Aveo", Km=180000)
predict<- predict.lm(reg, newdata)
predict
```

```
#Prediccion del precio de un auto del año 2016 de marca Chevrolet
reg<- lm(Price~Year+Badge, data = CARS)
newdata<- data.frame(Year=2016, Badge="Chevrolet")
predict<- predict.lm(reg, newdata, level = 0.99)
predict
```

```
#Prediccion del precio de un auto del año 2016 de modelo Aveo
reg<- lm(Price~Year+Model, data = CARS)
newdata<- data.frame(Year=2016, Model="Aveo")
predict<- predict.lm(reg, newdata)
predict
```

```
#####Prediccion del precio con año 2016, modelo Aveo y recorrido 54000 km
```

```
reg<- lm(Price~Year+Model+Km, data = CARS)
newdata<- data.frame(Year=2016, Model="Aveo", Km=54000)
predict<- predict.lm(reg, newdata)
predict
```

```
#####Prediccion del precio con año 2015, modelo Aveo y recorrido 180000 km
```

```
reg<- lm(Price~Year+Model+Km, data = CARS)
newdata<- data.frame(Year=2016, Model="Aveo", Km=180000)
predict<- predict.lm(reg, newdata)
predict
```



```
#####
```

```
#Prediccion del precio de un auto del año 2015 de marca Mercedes Benz
```

```
reg<- lm(Price~Year+Badge, data = CARS)
```

```
newdata<- data.frame(Year=2015, Badge="Mercedes")
```

```
predict<- predict.lm(reg, newdata)
```

```
predict
```

```
#Prediccion del precio de un auto del año 2015 de modelo AMG
```

```
reg<- lm(Price~Year+Model, data = CARS)
```

```
newdata<- data.frame(Year=2015, Model="AMG")
```

```
predict<- predict.lm(reg, newdata)
```

```
predict
```

```
#####Prediccion del precio con año 2015, modelo AMG y recorrido 54000 km
```

```
reg<- lm(Price~Year+Model+Km, data = CARS)
```

```
newdata<- data.frame(Year=2015, Model="AMG", Km=54000)
```

```
predict<- predict.lm(reg, newdata)
```

```
predict
```

```
#####Prediccion del precio con año 2015, modelo AMG y recorrido 180000 km
```

```
reg<- lm(Price~Year+Model+Km, data = CARS)
```

```
newdata<- data.frame(Year=2015, Model="AMG", Km=180000)
```

```
predict<- predict.lm(reg, newdata)
```

```
predict
```

```
#Prediccion del precio de un auto del año 2016 de marca Mercedes
reg<- lm(Price~Year+Badge, data = CARS)
newdata<- data.frame(Year=2016, Badge="Mercedes")
predict<- predict.lm(reg, newdata, level = 0.99)
predict
```

```
#Prediccion del precio de un auto del año 2016 de modelo AMG
reg<- lm(Price~Year+Model, data = CARS)
newdata<- data.frame(Year=2016, Model="AMG")
predict<- predict.lm(reg, newdata)
predict
```

```
#####Prediccion del precio con año 2016, modelo AMG y recorrido 54000 km
```

```
reg<- lm(Price~Year+Model+Km, data = CARS)
newdata<- data.frame(Year=2016, Model="AMG", Km=54000)
predict<- predict.lm(reg, newdata)
predict
```

```
#####Prediccion del precio con año 2015, modelo AMG y recorrido 180000 km
```

```
reg<- lm(Price~Year+Model+Km, data = CARS)
newdata<- data.frame(Year=2016, Model="AMG", Km=180000)
predict<- predict.lm(reg, newdata)
predict
```





UNIVERSIDAD
BICENTENARIA

¡Sueña, haz que suceda!