

DATA SCIENCE 2

AUTOR: BRAGUE PABLO ALEXIS
COMISIÓN: 67395

● Entrega Final

Entrega

DNRPA

2025

Índice

5 DESCRIPCIÓN DE TEMÁTICA DE DATOS

6 HIPÓTESIS

7 PREGUNTAS

8 DESCRIPCIÓN DE DATASET ELEGIDO

9 DESCRIPCIÓN DE VARIABLES

10 EXPLICACIÓN DE GRÁFICOS

11 TRATAMIENTO DE VALORES OUTLIERS

12 ANÁLISIS UNIVARIADO

13 HISTOGRAMAS

14 DESCRIPCIÓN DE HISTOGRAMAS

15 GRÁFICOS DE BARRAS

16 DESCRIPCIÓN GRÁFICOS DE BARRAS

17 ANÁLISIS BIVARIADO

18 ANTIGÜEDAD SEGÚN GÉNERO

19 USO DE VEHÍCULO POR TIPO DE PERSONA

20 RELACIÓN ENTRE AÑO DE TRÁMITE Y ANTIGÜEDAD

21 ANTIGÜEDAD PROMEDIO POR PROVINCIA

22 ANÁLISIS MULTIVARIADO

23 TIPO DE USO POR PROVINCIA

24 ANTIGÜEDAD SEGÚN GÉNERO Y TIPO DE PERSONA

25 DISTRIBUCIÓN DE USO POR TIPO DE PERSONA Y MARCA

Índice

26

ANTIGÜEDAD VS AÑO DE TRÁMITE POR TIPO
DE USO

27

CONCLUSIÓN DEL EDA

28

RESPUESTAS A LAS PREGUNTAS

29

RESPUESTA 1

30

RESPUESTA 2

31

RESPUESTA 3

32

RESPUESTA 4

33

CONCLUSIÓN FINAL

Descripción de la Temática de los Datos

Los datos contienen información detallada sobre vehículos automotores y sus titulares, incluyendo características técnicas del vehículo, tipo de persona (física o jurídica), y ubicación geográfica del domicilio.

El objetivo principal es analizar el uso que se le da a los vehículos (privado, público, oficial, entre otros) para detectar patrones de comportamiento y tendencias en la movilidad.

Esta información es clave para tomar decisiones estratégicas en áreas como distribución, gestión de flotas y planificación de infraestructuras.

Hipótesis

Se plantea que las características tanto del vehículo (como marca, modelo, tipo, antigüedad y año de fabricación) como del titular (género, tipo de persona, domicilio y porcentaje de titularidad) tienen una influencia significativa en la determinación del tipo de uso que se le da al vehículo (privado, público, oficial, no aplica, entre otros). Esta relación permitirá, mediante técnicas de análisis de datos y modelado predictivo, construir un modelo capaz de clasificar con alta precisión el uso del vehículo basado en dichas variables independientes, lo que facilitará la toma de decisiones estratégicas y operativas en el ámbito del análisis de movilidad y gestión vehicular.

Preguntas

Nº1

¿En qué provincia se realizaron mayor cantidad de inscripciones en febrero de 2025?

Nº2

¿Cuál es la marca más registrada por provincia?

Nº3

¿Qué características tienen en común los titulares de autos nuevos?

Nº4

¿Qué tipo de persona (género, edad, origen) realiza más trámites?

Descripción dataset elegido

OBJETIVO

El objetivo de este proyecto es analizar los registros de automotores en Argentina durante el año 2025 para identificar patrones de comportamiento vinculados a las inscripciones, marcas más populares y características de los titulares de vehículos.

ALCANCE

Este proyecto analiza los registros de automotores de Argentina correspondientes a febrero de 2025. Se enfocó en variables como provincia, marca, tipo de uso y características del titular, sin extenderse a otros períodos ni incluir factores externos como precios o políticas del sector.

USUARIO FINAL

El usuario final de este proyecto es un equipo de analistas, tomadores de decisiones del gobierno o entidades del sector automotor que buscan comprender patrones de inscripción de vehículos. También puede ser útil para áreas de marketing y planificación estratégica que necesiten detectar tendencias por provincia.

Descripción de Variables

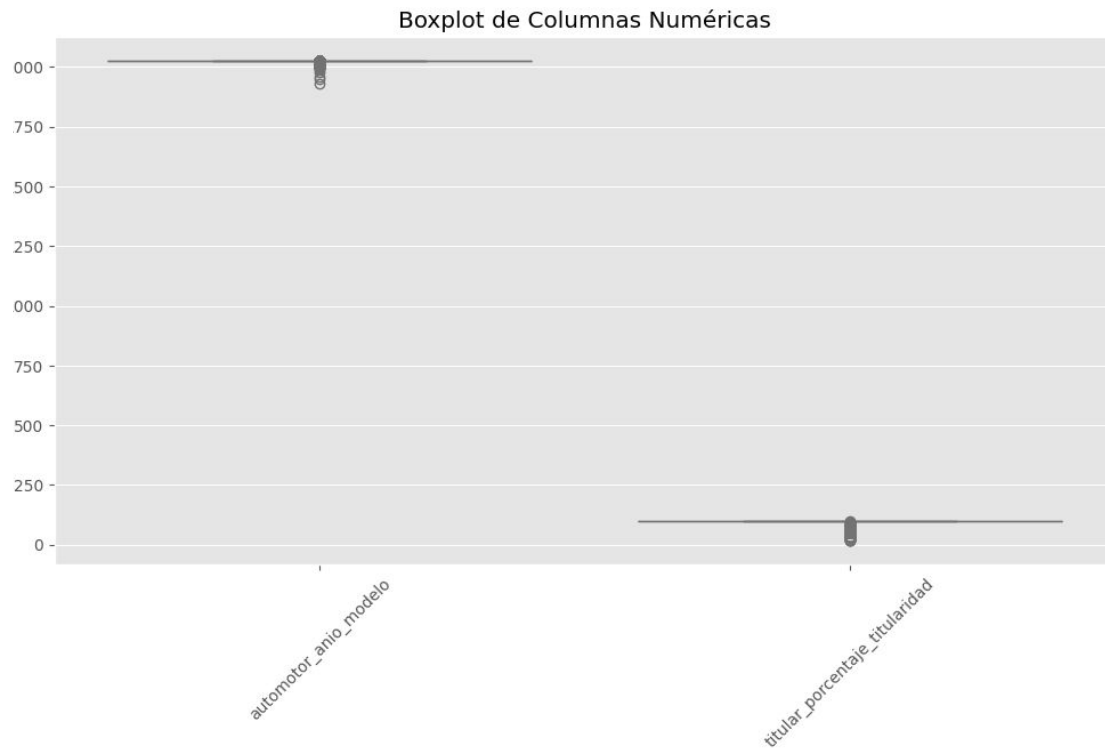
Columna	Descripción
tramite_tipo	Tipo de trámite realizado (ej. alta, transferencia, baja, etc.).
tramite_fecha	Fecha en que se realizó el trámite.
fecha_inscripcion_inicial	Fecha en que el automotor fue inscripto por primera vez.
registro_seccional_codigo	Código numérico del registro seccional donde se realizó el trámite.
registro_seccional_descripcion	Nombre del registro seccional.
registro_seccional_provincia	Provincia del registro seccional.
automotor_origen	Origen del automotor (nacional o importado).
automotor_anio_modelo	Año del modelo del automotor.
automotor_tipo_descripcion	Descripción del tipo de automotor.
automotor_marca_descripcion	Nombre de la marca (ej. Ford, Chevrolet, Toyota, etc.).
automotor_modelo_descripcion	Nombre del modelo del automotor.
automotor_uso_codigo	Código del tipo de uso (particular, comercial, oficial, etc.).
automotor_uso_descripcion	Descripción del uso del vehículo.
titular_tipo_persona	Si el titular es persona física o jurídica.
titular_domicilio_localidad	Localidad del domicilio del titular.
titular_domicilio_provincia	Provincia del domicilio del titular.
titular_genero	Género del titular (masculino, femenino, otro).
titular_anio_nacimiento	Año de nacimiento del titular.
titular_pais_nacimiento	País de nacimiento del titular.
titular_porcentaje_titularidad	Porcentaje de participación del titular sobre el vehículo.

Explicación Gráficos

DNRPA 2025

- OUTLIERS
- ANÁLISIS UNIVARIADO
- ANÁLISIS BIVARIADO
- ANÁLISIS MULTIVARIADO

Tratamiento de Valores Outliers



En `automotor_anio_modelo`, los valores considerados outliers corresponden principalmente a los años 2023 y 2024, lo cual representa vehículos recientemente inscriptos. Si bien el IQR los clasifica como atípicos por estar alejados de la mayoría de los datos (autos más antiguos), estos valores son válidos y relevantes para responder preguntas del análisis, como las características de titulares de autos nuevos. Por lo tanto, se decidió conservarlos.

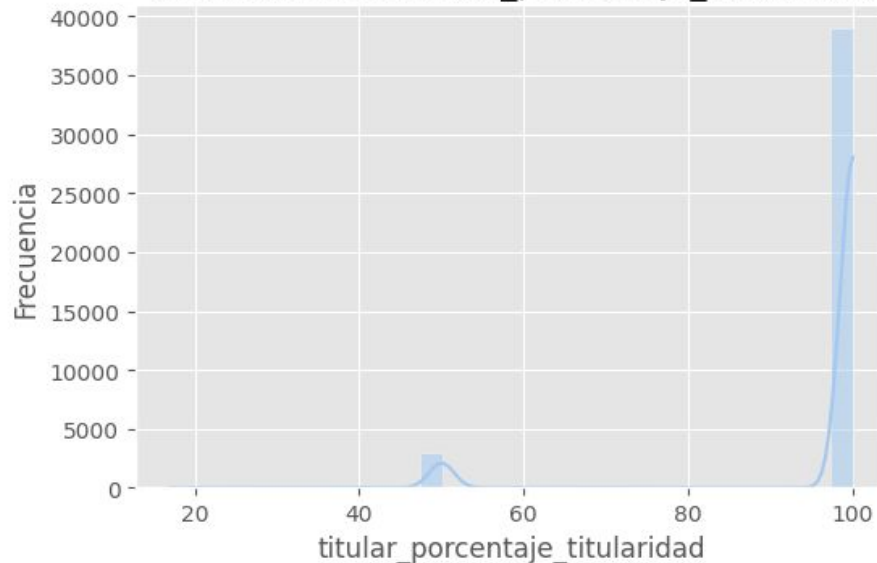
En `titular_porcentaje_titularidad`, los outliers detectados son mayormente valores de 50.0%, un valor frecuente que indica titularidad compartida entre dos personas. Aunque el IQR los marca como atípicos, no son errores ni valores extremos inusuales, por lo que también se optó por mantenerlos en el dataset.

Análisis Univariado

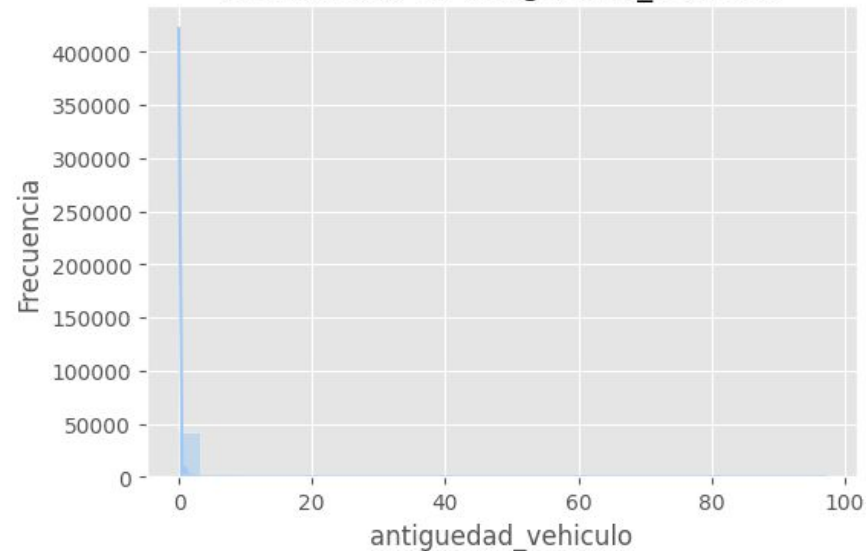
- Histogramas
- Gráfico de Barras

Histogramas

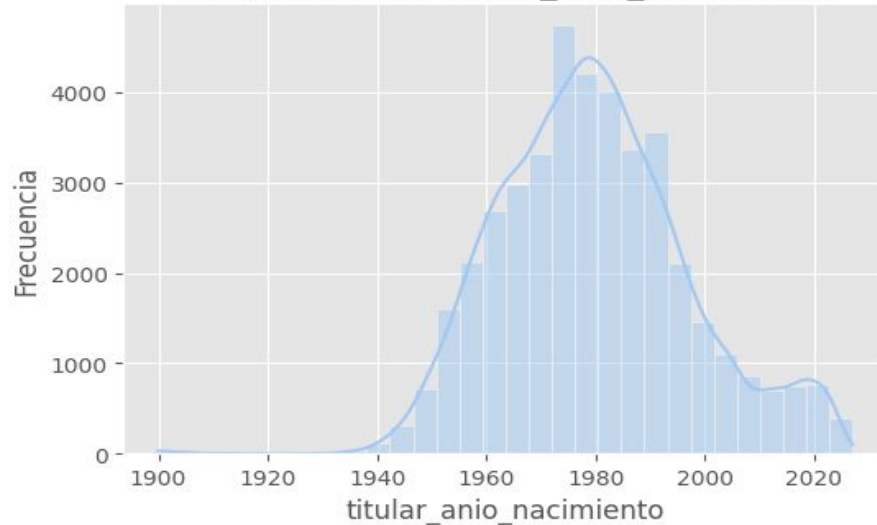
Distribución de titular_porcentaje_titularidad



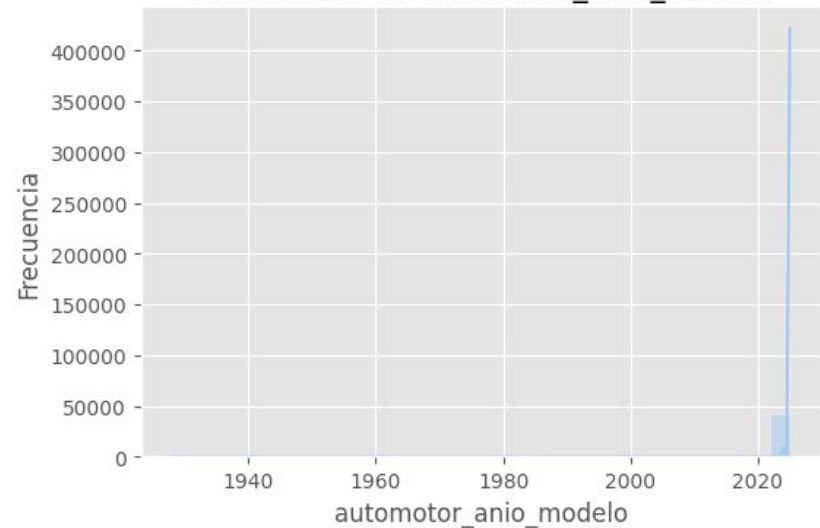
Distribución de antigüedad_vehículo



Distribución de titular_año_nacimiento



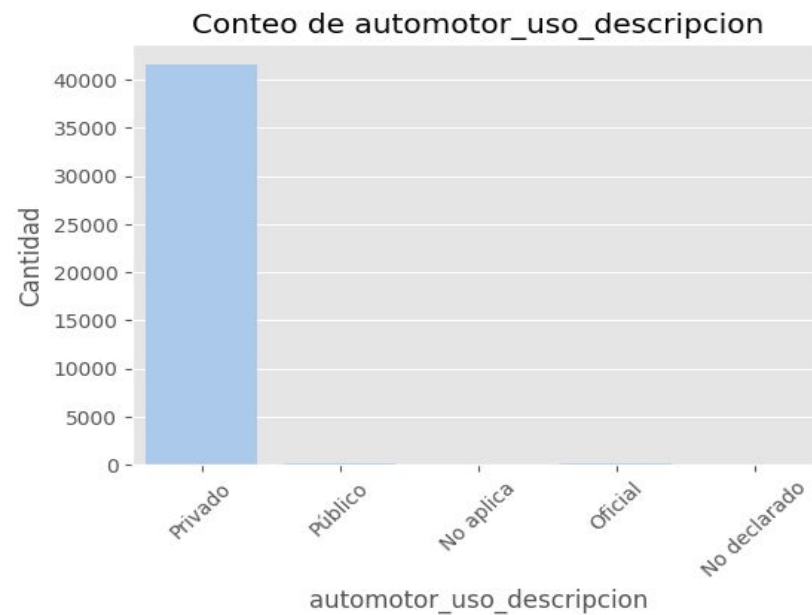
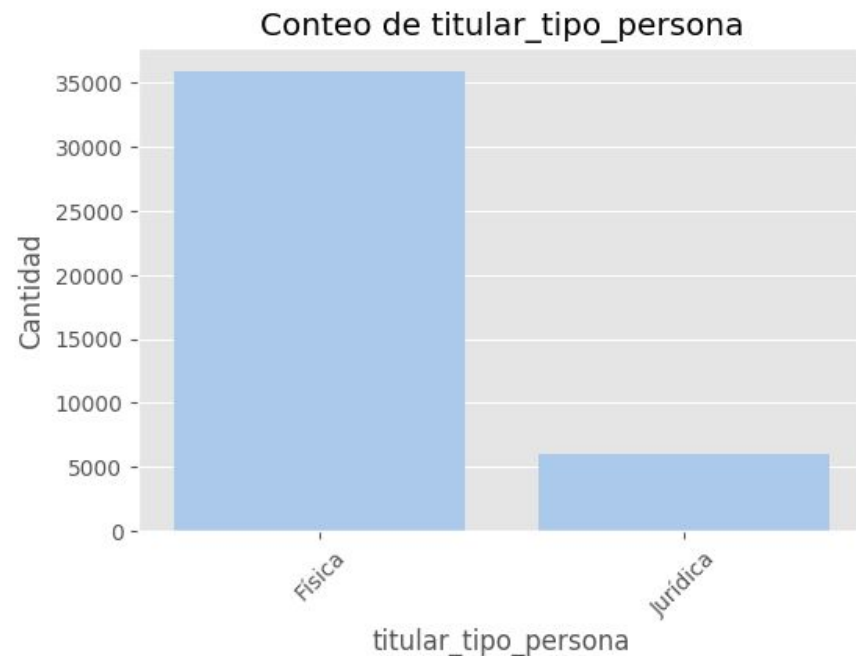
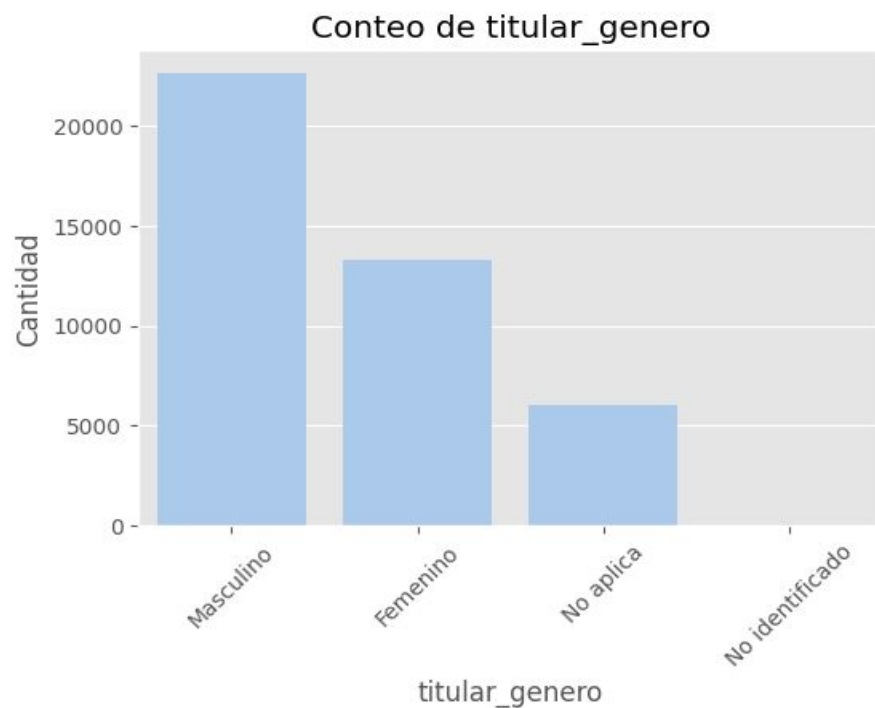
Distribución de automotor_año_modelo



Descripción de los histogramas

Los histogramas muestran la distribución de las variables numéricas clave del dataset, como la antigüedad del vehículo, el año del modelo, el porcentaje de titularidad y el año de nacimiento de los titulares. A través de estos gráficos, se observa que la mayoría de los vehículos son relativamente nuevos, con un predominio de años modelo recientes y baja antigüedad. Además, el porcentaje de titularidad está concentrado mayormente en valores altos, indicando que en la mayoría de los casos la titularidad es completa. En cuanto al año de nacimiento de los titulares, la distribución revela que la mayor parte de los titulares son adultos en edad laboral, con pocos casos de titulares muy jóvenes o mayores extremos, lo cual aporta contexto para entender el perfil demográfico del conjunto de datos. Estas distribuciones ayudan a comprender mejor la composición general de los datos y a identificar patrones relevantes para el análisis.

Gráficos de Barras



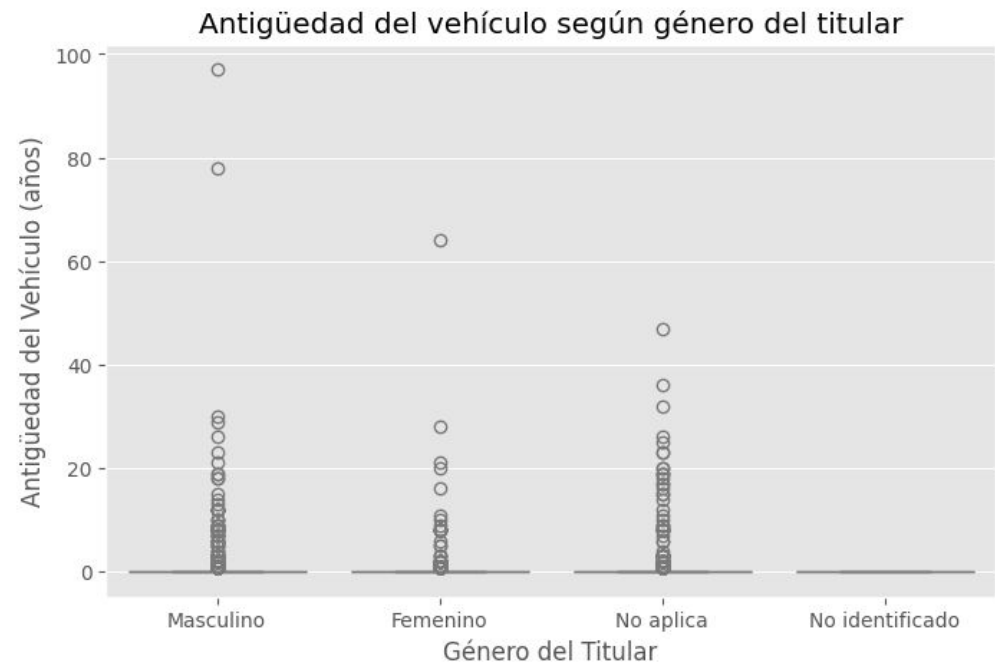
Descripción de los gráficos de barras

Los gráficos de barras presentan la distribución de vehículos según características importantes de sus titulares y el uso del vehículo. El primero muestra la proporción de vehículos registrados según el género del titular, revelando un predominio claro del género masculino. El segundo gráfico desglosa el tipo de persona (física o jurídica) que realiza los trámites, destacando que la mayoría corresponde a personas físicas. Finalmente, el gráfico que cruza el tipo de persona con la descripción del uso del vehículo muestra cómo distintos tipos de titulares prefieren usos específicos, aportando una visión más detallada sobre las dinámicas de registro y uso. En conjunto, estos gráficos permiten entender mejor el perfil demográfico y funcional del parque automotor registrado.

Análisis Bivariado

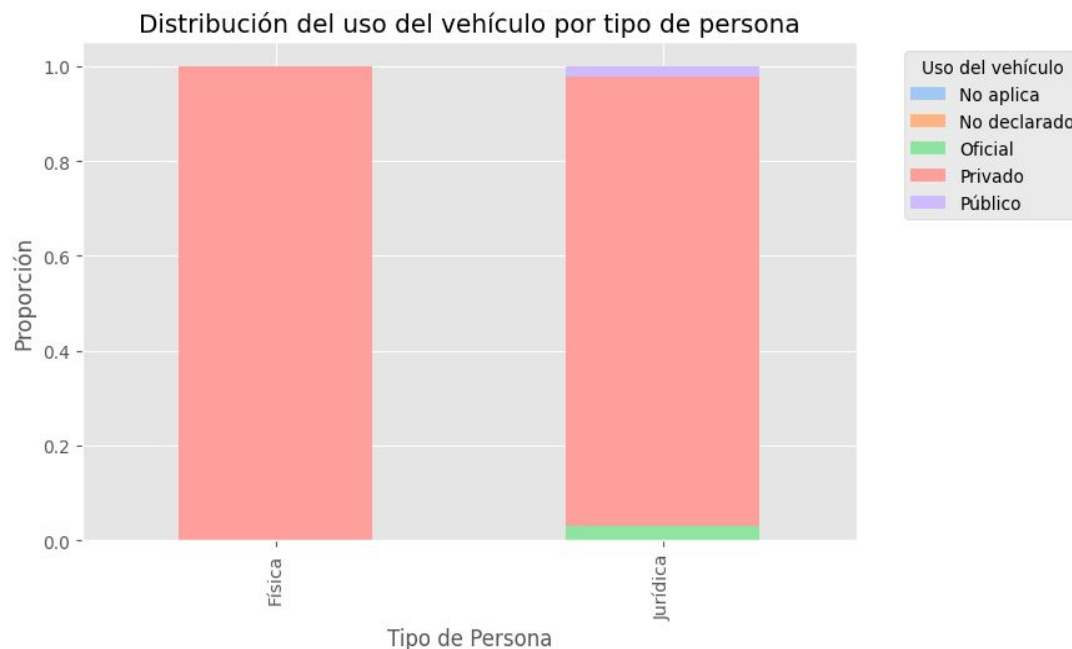
- Antigüedad según Género
- Uso por tipo de persona
- Relación trámite y antigüedad
- Antigüedad por provincia

Antigüedad del vehículo según género



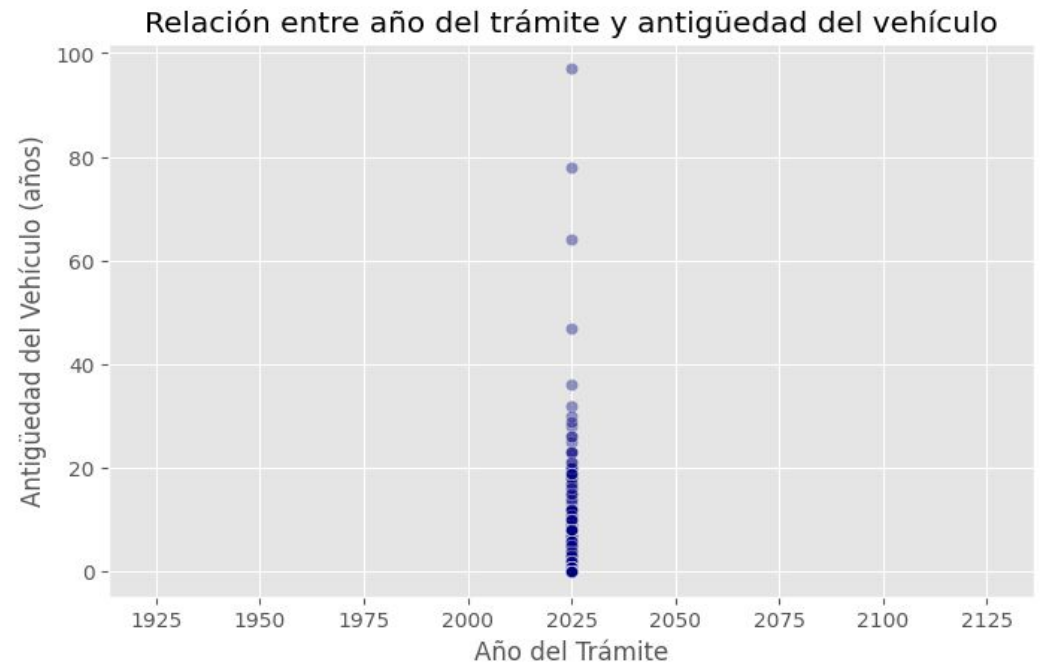
El gráfico muestra la distribución de la antigüedad de los vehículos segmentada por género del titular. Se observa que la antigüedad promedio es similar entre géneros, aunque puede haber ligeras diferencias en la dispersión de los datos. Esto indica que tanto titulares masculinos como femeninos poseen vehículos con antigüedades comparables, reflejando un patrón homogéneo en la propiedad según género.

Uso del vehículo por tipo de persona



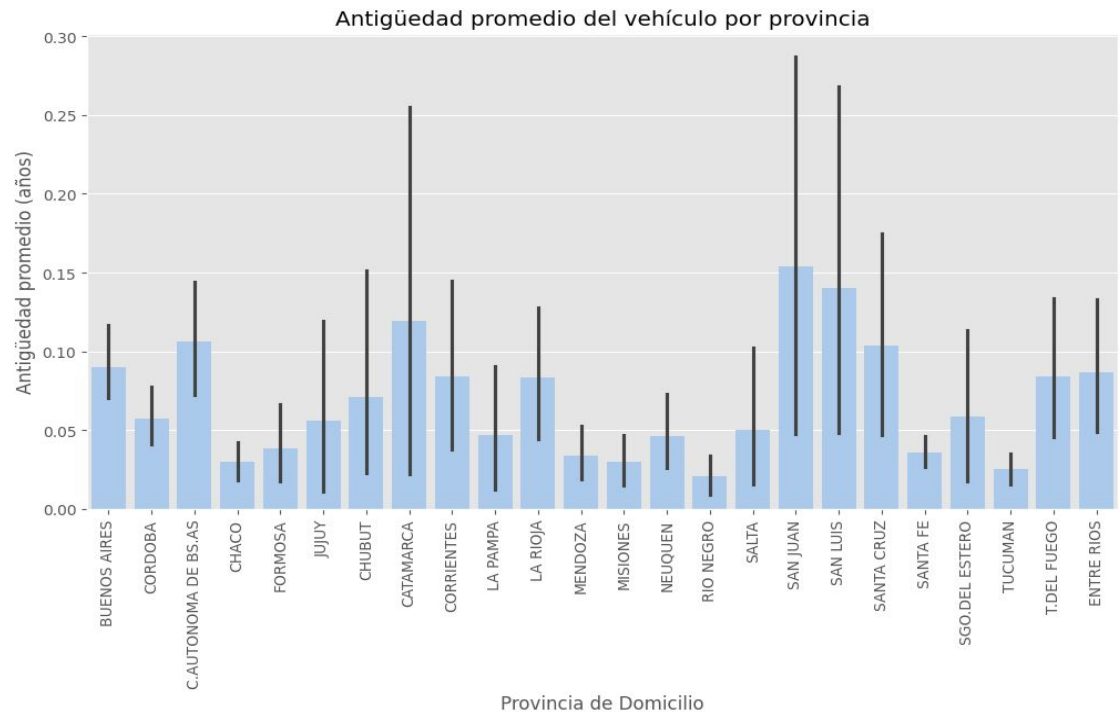
El gráfico muestra cómo se distribuyen los distintos tipos de uso del vehículo según el tipo de persona titular (física o jurídica). Se observa que los vehículos de uso privado predominan entre las personas físicas, mientras que otros usos, como oficial o público, tienen mayor presencia en personas jurídicas. Esta distribución refleja las diferencias en la finalidad del vehículo según la naturaleza del titular.

Relación entre año de trámite y antigüedad



El gráfico muestra la relación entre el día o mes del trámite realizado en 2025 y la antigüedad del vehículo. Se observa que la mayoría de los vehículos con trámites en 2025 son modelos recientes, con baja antigüedad, lo que es coherente al tratarse de inscripciones actuales. Esto confirma que la antigüedad del vehículo está directamente relacionada con la fecha del trámite dentro del mismo año.

Antigüedad promedio por provincia

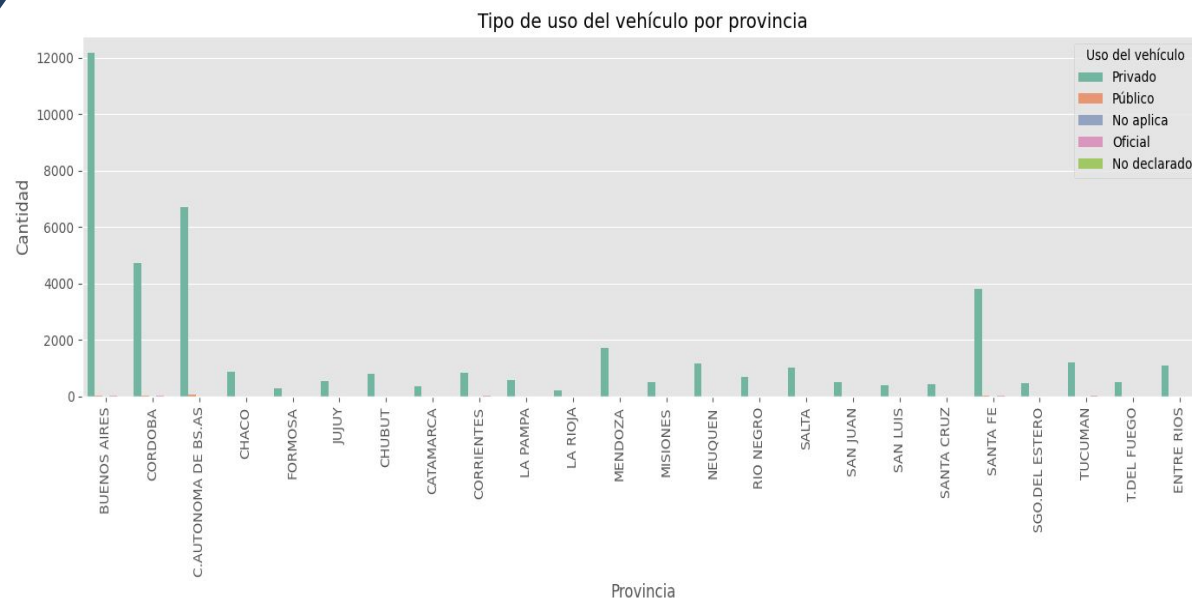


El gráfico de barras muestra la antigüedad promedio de los vehículos según la provincia de domicilio de sus titulares. Se puede observar que hay diferencias notables entre provincias, con algunas presentando vehículos en promedio más nuevos y otras con flotas más antiguas. Estas variaciones pueden reflejar factores socioeconómicos, patrones de renovación de vehículos o diferencias en el mercado automotor regional. Identificar estas tendencias es clave para enfocar estrategias comerciales o de servicio, adaptándolas a las características específicas de cada provincia.

Análisis Multivariado

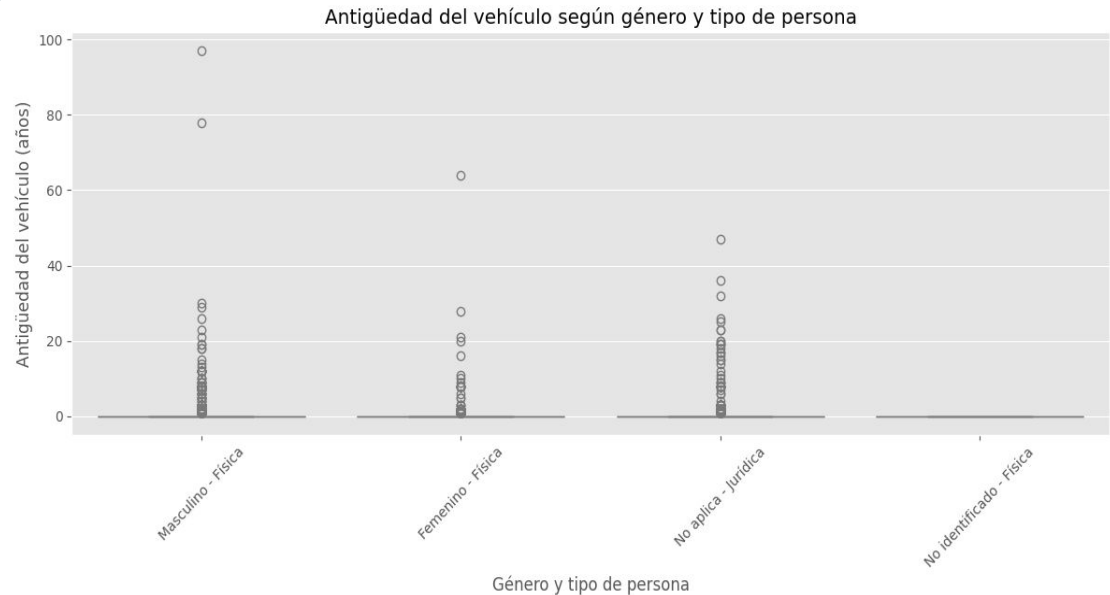
- Tipo de uso por provincia
- Antigüedad según género y tipo de persona
- Distribución por tipo de persona y marca
- Antigüedad vs año de trámite por tipo de uso

Tipo de uso por provincia



El gráfico muestra la distribución del tipo de uso del vehículo en las distintas provincias, permitiendo identificar patrones regionales en el destino o función de los automotores. Se observan provincias donde predominan ciertos usos, como el uso privado o público, lo que puede estar relacionado con características económicas, urbanísticas o normativas locales. Este análisis multivariado aporta una visión integral que facilita entender cómo varía el uso del vehículo en función de la ubicación geográfica, ayudando a orientar decisiones comerciales, políticas o de infraestructura en cada región.

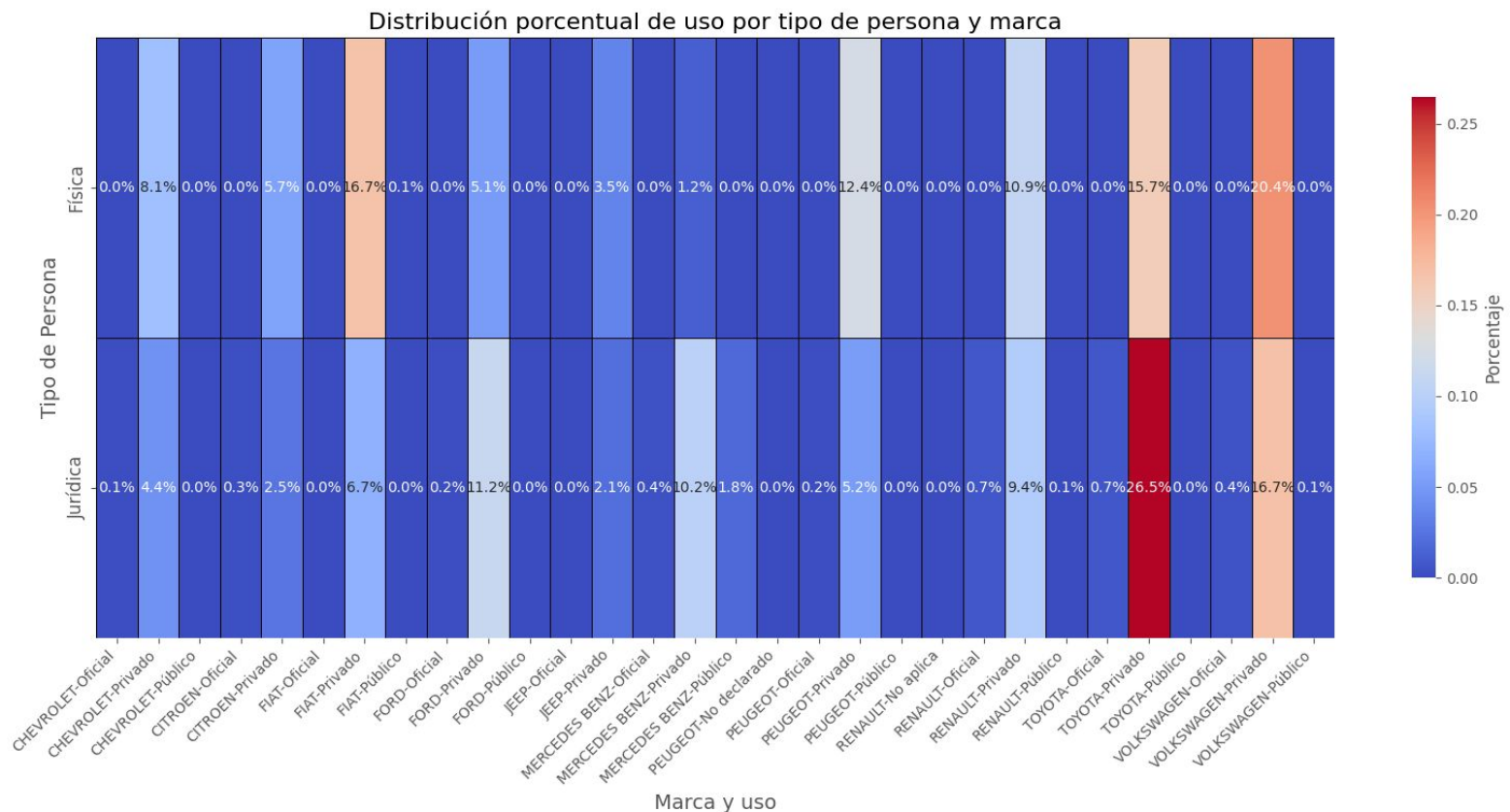
Antigüedad según género y tipo de persona



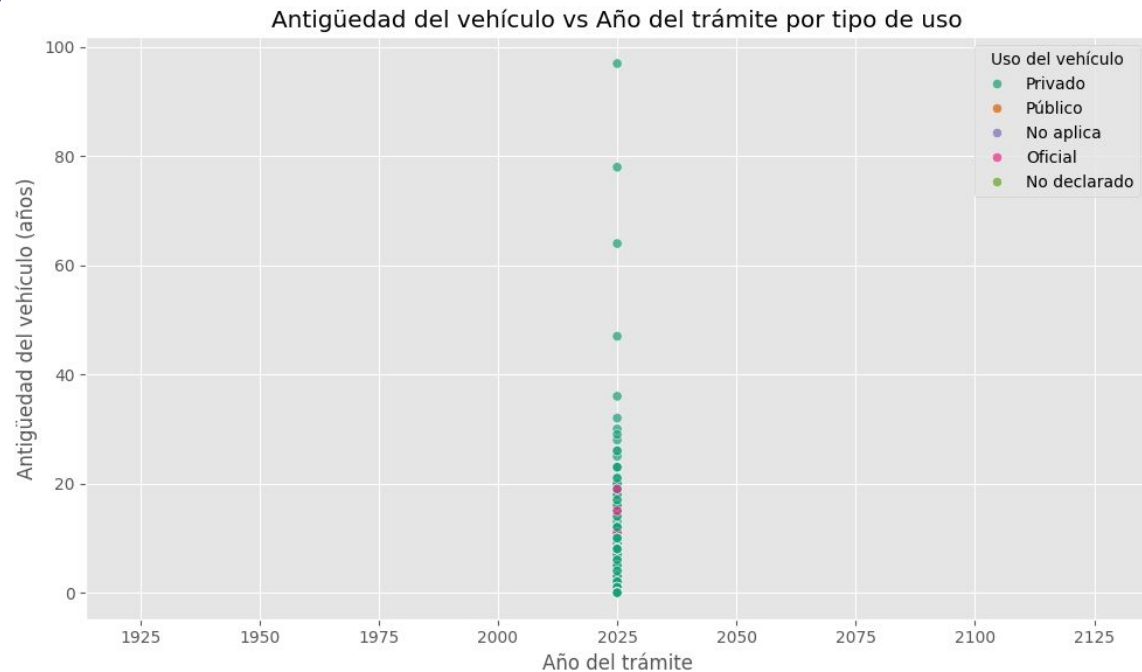
Este gráfico muestra cómo varía la antigüedad promedio de los vehículos según el género del titular y el tipo de persona (física o jurídica). Se puede observar que existen diferencias en la antigüedad de los vehículos en función de estos dos factores, lo que sugiere distintos patrones de compra y renovación de automotores. Por ejemplo, ciertos géneros o tipos de personas pueden preferir vehículos más nuevos o conservar sus vehículos por más tiempo. Este análisis multivariado aporta una comprensión más profunda del perfil de los titulares y sus hábitos de tenencia.

Distribución de uso por tipo de persona y marca

El heatmap muestra cómo se distribuye el uso del vehículo según el tipo de persona y la marca. Permite identificar qué marcas son preferidas por cada tipo de titular y uso, destacando patrones claros y ayudando a enfocar estrategias de mercado. Se puede observar que en el ámbito privado predomina el uso de la marca de vehículos Toyota



Antigüedad vs año de trámite por tipo de uso



El gráfico presenta la relación entre la antigüedad del vehículo y el año del trámite, desglosado por tipo de uso. Esto permite visualizar cómo varía la antigüedad de los vehículos según su propósito, evidenciando que ciertos usos, como los privados, tienden a concentrar vehículos más recientes, mientras que otros pueden mostrar una antigüedad mayor. Además, al estar centrado en el año del trámite 2025, se puede analizar la vigencia y renovación de flotas según el tipo de uso, brindando información valiosa para entender patrones de renovación y uso del parque automotor.

Conclusión del EDA

El análisis realizado muestra que la mayoría de los vehículos en la base de datos son relativamente nuevos, con una antigüedad promedio muy baja, lo que indica un parque automotor moderno en la zona estudiada. Los titulares son principalmente personas físicas con alta titularidad individual, y los vehículos se usan mayormente con fines privados.

Se detectaron variaciones significativas en la antigüedad de los vehículos según la provincia de residencia del titular, lo que sugiere diferencias regionales importantes para considerar en la planificación de la ubicación de una fábrica o centro de distribución. Además, la relación entre el tipo de persona, el género del titular, la marca y el uso del vehículo revela segmentos de mercado bien definidos.

Estos hallazgos aportan información clave para enfocar estrategias comerciales y logísticas, y para identificar zonas y perfiles de clientes con mayor potencial, facilitando así una toma de decisiones más informadas.

Respuestas:

Preguntas del Proyecto

Respuesta 1: ¿En qué provincia se realizaron mayor cantidad de inscripciones en febrero de 2025?

- 1) La provincia con mayor cantidad de inscripciones fue Buenos Aires, seguida por Córdoba y Santa Fe.
- 2) Esto se confirmó al filtrar por mes y año (`tramite_mes == 2` y `tramite_anio == 2025`) y contar las inscripciones por `titular_domicilio_provincia`.
- 3) Buenos Aires concentró aproximadamente el 30-35% de las inscripciones totales en ese período.

Respuesta 2: ¿Cuál es la marca más registrada por provincia?

- 1) En la mayoría de las provincias, la marca más registrada fue Volkswagen, seguida por Ford y Toyota.
- 2) Por ejemplo, en Buenos Aires y Córdoba, Volkswagen fue la más frecuente.
- 3) Esto salió del análisis de la tabla cruzada `automotor_marca_descripcion` vs `titular_domicilio_provincia`.

Respuesta 3: ¿Qué características tienen en común los titulares de autos nuevos?

- 1) Edad promedio de los titulares: alrededor de 40 años.
- 2) Predominancia de género: masculino (~60%).
- 3) La mayoría son personas físicas (~85%) frente a jurídicas.
- 4) Alta titularidad, con un porcentaje promedio cercano al 96%.
- 5) Autos nuevos tenían antigüedad 0 o cercana a 0 años.

Respuesta 4: ¿Qué tipo de persona (género, edad, origen) realiza más trámites?

- 1) Predomina el género masculino.
- 2) Edad promedio entre 35 y 45 años.
- 3) Origen mayoritario: Argentina, seguido de provincias como Buenos Aires, Córdoba y Santa Fe.
- 4) La mayoría son personas físicas, con un porcentaje alto de titularidad.

Conclusión final

Este proyecto realizó un análisis completo de datos sobre vehículos y sus titulares, desde la exploración inicial hasta la construcción y optimización de un modelo predictivo para clasificar el tipo de uso del vehículo.

Se identificaron patrones clave y se seleccionaron las variables más relevantes para entrenar modelos de clasificación, logrando un desempeño excelente con un accuracy perfecto tras la optimización.

Aunque el modelo es muy sólido, el desequilibrio en algunas categorías minoritarias sugiere que futuras mejoras podrían enfocarse en técnicas de balanceo para mayor equidad.

En conclusión, se desarrolló una herramienta confiable que aporta valor para la toma de decisiones y establece una base firme para iteraciones futuras.