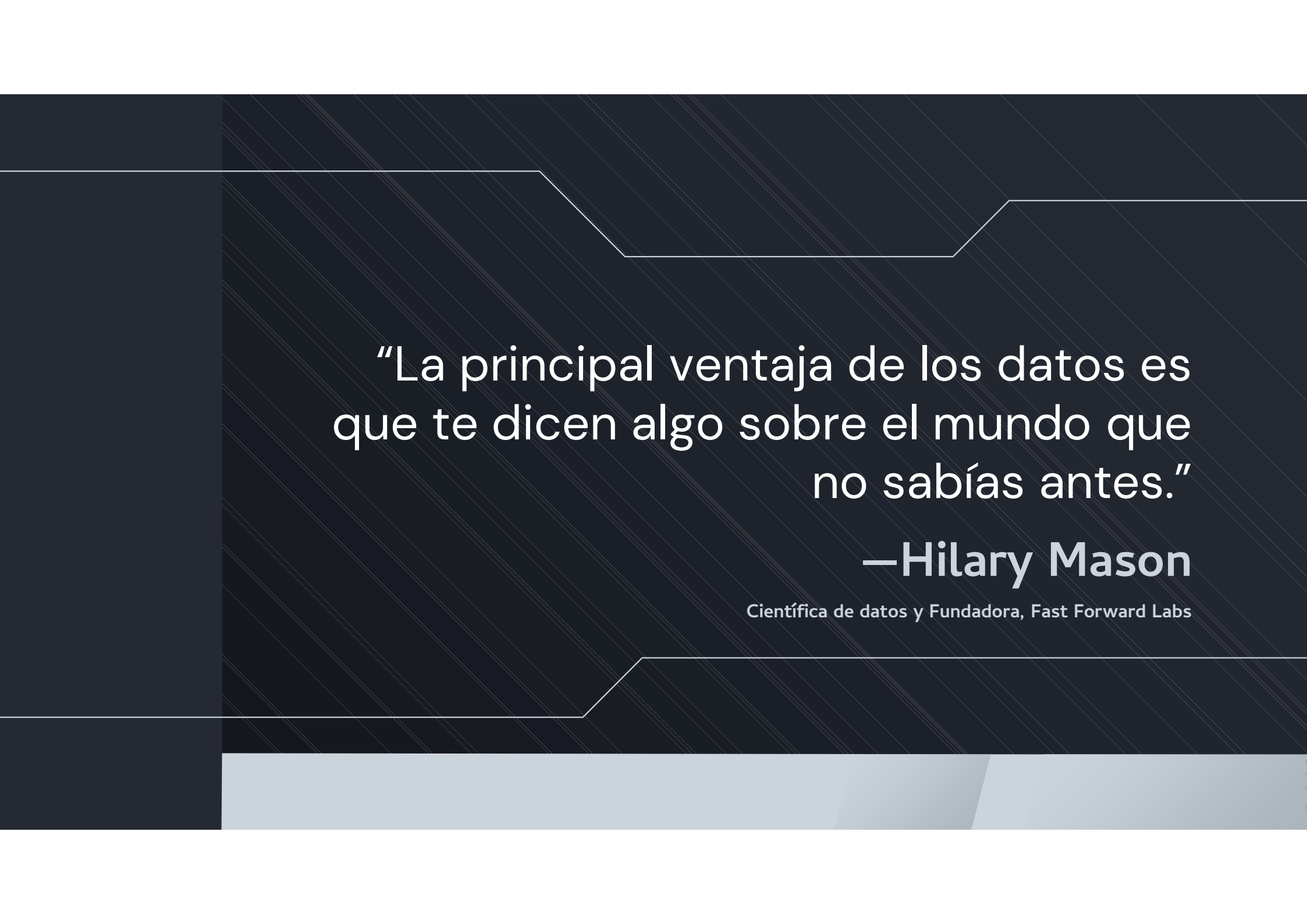


Modelo de Predicción del Precio de Vehículos Usados

Data Science II: Machine Learning para la Ciencia de Datos





“La principal ventaja de los datos es
que te dicen algo sobre el mundo que
no sabías antes.”

—Hilary Mason

Científica de datos y Fundadora, Fast Forward Labs



Ventas de vehículos usados

- Mercado en constante crecimiento a nivel mundial (8% anual).
- Ingresos de más de 140 mil millones de USD. para el 2021.
- Digitalización e inclusión tecnológica del mercado.
- Generación de gran cantidad de datos explotables.



\$284 mil millones

Ingresos proyectados para el 2030.

Principales razones de este crecimiento



Mercado Organizado

Participación de fabricantes y empresas organizadas



Financiamiento

Creditos y aumentos en los ingresos.



Digitalizacion

Inclusión tecnológica y experiencias emergentes



Nuevas tendencias

Periodos de propiedad más cortos, demanda de autos de lujo, uso de compactos.

Modelo predictivo del precio de automóviles usados

Dirigido a personal operativo de una empresa vendedora o para desarrollar un aplicativo web o celular que facilite la experiencia al cliente que desee vender un automovil.





Dataset

- 19.237 muestras con 17 columnas y 1 variable objetivo.
- Datos Fabricante: marca, modelo, categoría y año de producción.
- Datos Técnicos: tipo de combustible, volumen del motor, cantidad de cilindros, tipo de transmisión, tipo de tracción, ubicación del volante y cantidad de airbags.
- Datos Estéticos: Interior de cuero, cantidad de puertas y color.
- Datos de Uso: Kilometraje

Hipótesis

Con toda esta información se busca identificar la relación de cada uno de estos factores con la valoración final del precio de un automóvil. Una vez contemos con este análisis nos gustaría poder responder las siguientes preguntas:

01

¿Qué categoría tiene el mayor impacto?

02

¿Este impacto se repite en todas las marcas?

03

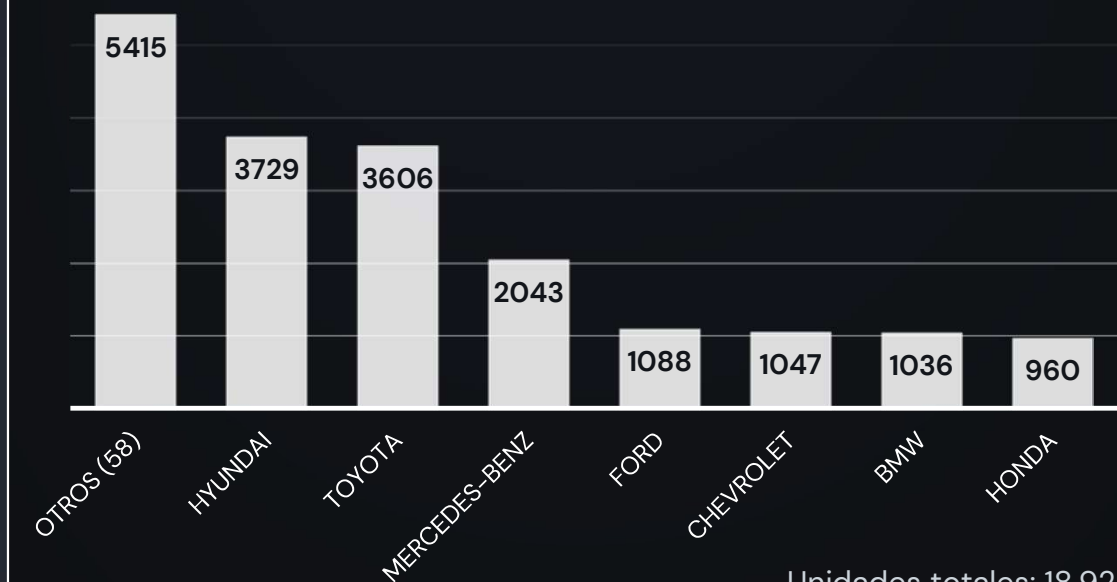
¿Vale la pena considerar las características técnicas y estéticas pensando en el valor de reventa?

Objetivo

Desarrollar un modelo de data science que pueda predecir el valor de venta de un automóvil usado basándose en diversas características. Para lograr esto, usaremos diversos algoritmos, para poder comparar la efectividad de cada uno de ellos e inclinarnos por el que genere los mejores resultados en la predicción.

Impacto en el mercado por marca

Total de Unidades



Unidades totales: 18.924

*Otros contiene las marcas restantes 28% de la muestra.

Fabricante	Modelo	Unidades
OTROS (58)	Optima (Kia)	230
HYUNDAI	Sonata	1067
TOYOTA	Prius	1069
MERCEDES-BENZ	E 350	534
FORD	Fusion	295
CHEVROLET	Cruze	295
BMW	X5	351
HONDA	FIT	444

TOP 3 Ventas Marcas

\$83,2 mil millones

Ventas Hyundai: 3729 unidades.

Precio promedio: \$22.300.

\$51,6 mil millones

Ventas Toyota: 3606 unidades.

Precio promedio: \$14.300.

\$38,0 mil millones

Ventas Mercedes-Benz: 2043 unidades.

Precio promedio: \$18.600.

\$327,9 mil millones

Ventas Totales: 18.924 unidades.

Precio promedio: \$17.300.

Información Observada EDA.

- Hay muestras que pertenecen a marcas, modelos o categorías de automóviles muy específicas que podríamos descartar ya que pueden traer más problemas al modelo que beneficios.
- Algunas marcas tienen suficientes muestras lo que nos permitiría comparar el comportamiento del modelo aplicado a todas las marcas o segmentándolo por algunas, viendo su comportamiento.
- Al ver la correlación de variables no se observa alguna relación lineal con nuestra variable objetivo, por lo que podríamos suponer que las especificaciones técnicas o estéticas, por si solas no tienen un efecto considerable en el valor del precio.
- La cantidad de dinero que genera el mercado de automóviles usados hace que nuestro proyecto sea de gran interés para diversas empresas que podrían sacar beneficios económicos considerables además de ventajas competitivas.





Modelado

- Se realizaron pruebas con los modelos de regresión lineal, árboles de decisión, random forest, xgboost y gradient boosting .
- Xgboost y gradient boosting obtuvieron muy buenos performance, pero para el proyecto nos decantamos por la opción de gradient boosting .
- Los parámetros usados fueron de max_depth=7 y n_estimators=200.
- El valor de R^2 obtenido fue de 0,81.
- Las variables que más impacto tienen en el precio son: la cantidad de airbags, la edad del auto, el volumen del motor, el impuesto asociado a este, el tipo de combustible (Diesel) y el kilometraje.

Respondiendo a las preguntas de nuestra hipótesis

01

¿Qué categoría tiene el mayor impacto?

La categoría que tiene mayor impacto en el valor del automovil es la cantidad de airbags.

02

¿Este impacto se repite en todas las marcas?

El impacto de la cantidad de airbags se mantiene sin importar la marca del fabricante.

03

¿Vale la pena considerar las características técnicas y estéticas pensando en el valor de reventa?

Entre las 20 características con mayor impacto en el precio 9 son técnicas, 3 son estéticas y el resto estan asociadas a la marca, año de fabricación y referentes al uso de este.

Conclusiones

Para nuestro caso de negocio, en el cual buscamos tener un valor aproximado del precio de un automóvil usado sin que se requiera de la presencia de un perito o una revisión a fondo de este, se considera que este modelo de ML cumple con el requerimiento ya que el mismo tiene mejor rendimiento para los autos en el sector de precio medio por lo que una empresa podría agilizar el proceso de forma considerable.

Al tener el listado de la importancia de las variables y cómo actúan estas en el precio se podría diseñar un formulario de uso tanto para la empresa que vende o para el cliente que desea vender, el cual le daría un rango de precio estimado del valor de su automóvil con solo indicar algunas características de este.

Otra manera de aplicarlo sería que basado en un presupuesto del futuro comprador se podrían indicar un listado de las posibles combinaciones de características que puede tener un automóvil para mantener ese rango de precio o el caso contrario que definiendo características específicas le dé un valor aproximado del costo que tendría ese automóvil.

Gracias!

¿Preguntas?

clementeelias85@gmail.com

+54 9 11 2265 2777



Clemente E. Márquez A.

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**

