

# Predicción de precios de automóviles



## DATA SCIENCE

Coderhouse - Comisión 42365

## Índice

Descripción del dataset.....	2
Hipótesis.....	2
Objetivos.....	3
Conclusiones.....	3
Preguntas de interés y visualizaciones.....	4
Cual es el precio promedio de los automóviles.....	4
Cual es el precio promedio por tipo de automóviles.....	4
Cual es el precio promedio por marca de los automóviles.....	5
Cual es el precio promedio por año de fabricación.....	6
Cual es el precio promedio de los vehículos por condición.....	7
Un auto con mas kms tiende a ser más económico, pero la transmisión importa.....	8
Como comparar las predicciones de un modelo lineal y polinomial para predecir el precio de vehículos según su año.....	9
Como se comparan los modelos de regresión lineal y polinomial para predecir el precio de vehículos.....	10
Como predecir el precio de vehículos utilizando un modelo de regresión lineal.....	11
Hay alguna relación entre la transmisión de un vehículo y sus otras características.....	12
Recomendaciones e Insights.....	13
Insights.....	13
Recomendaciones.....	13

## Descripción del Dataset



El siguiente set de datos descargado desde el sitio web de kaggle trata sobre el mercado de vehículos usados, es importante entender qué factores afectan los precios de los automóviles. Los compradores desean obtener el vehículo más completo por su dinero, mientras que los vendedores buscan establecer precios razonables para sus vehículos. Para facilitar la toma de decisiones, en este análisis exploratorio de datos, utilizaremos un dataset que contiene información sobre vehículos usados a la venta en Dallas, Texas.

Exploraremos las diferentes variables que se encuentran disponibles, como el año de fabricación, la marca, el estado, el tipo de transmisión, el kilometraje, etc. A través de visualizaciones y análisis, identificaremos patrones y relaciones entre las variables con el objetivo de obtener una comprensión más profunda de los factores que afectan los precios de estos vehículos.

Este estudio proporcionará insights valiosos tanto para vendedores como compradores, permitiéndoles tomar decisiones informadas en el mercado de vehículos usados. Con una mejor comprensión de los determinantes del precio, los compradores podrán elegir el vehículo que mejor se ajuste a sus necesidades y presupuesto, mientras que los vendedores podrán establecer precios competitivos para sus productos.

En conclusión, este análisis contribuirá a un mercado más eficiente para la comercialización.

## Hipótesis



Se plantea la hipótesis de que el precio de los autos en venta está relacionado con una serie de factores interconectados. En primer lugar, el año de fabricación desempeña un papel fundamental, ya que se espera que los vehículos más recientes tengan precios más altos debido a su valor percibido. Sin embargo, esta relación entre el año de fabricación y el precio puede estar influenciada por otros factores, como el estado general del vehículo, ya que los autos en mejores condiciones tienden a tener un mayor valor de mercado. Además, la marca del automóvil también ejerce influencia, ya que algunas marcas premium tienden a tener precios más altos debido a su reputación y calidad percibida.

El tipo de transmisión y el sistema de combustión son otros factores a considerar, ya que los vehículos con transmisiones automáticas y sistemas de combustión eficientes pueden ser más demandados en el mercado, lo que podría tener un efecto positivo en sus precios. Por último, la relación entre el precio y los kilómetros recorridos permite evaluar la depreciación y la demanda en el uso del vehículo, lo que puede ser un factor importante para determinar su valor en el mercado de autos usados.

En conjunto, esta investigación tiene como objetivo proporcionar una comprensión completa de cómo estos factores interrelacionados afectan los precios de los autos en venta. Esto, a su vez, ayudará a los compradores y vendedores a tomar decisiones más informadas en el mercado automotriz.

## Objetivos



El objetivo de este trabajo es realizar un análisis exploratorio de datos de un dataset que contiene información de vehículos usados a la venta en el estado de Dallas, Texas, USA. El análisis nos permitirá entender mejor los factores que influyen en los precios de los vehículos usados y proporcionar información relevante para los potenciales compradores y vendedores.

## Conclusiones



El análisis del mercado de vehículos usados en Dallas ha revelado patrones importantes que influyen en los precios de los automóviles. En primer lugar, se observa una tendencia clara de que la antigüedad del automóvil está inversamente relacionada con su precio, lo que significa que los vehículos más antiguos tienden a tener precios más bajos. Esta relación es un hallazgo significativo para los compradores y vendedores de vehículos usados en la región.

Además, se destaca la importancia de la condición del vehículo y el kilometraje en la determinación de los precios. Los automóviles en mejores condiciones y con menor kilometraje tienden a tener precios más altos, lo que sugiere que los compradores valoran la calidad y el estado general de un automóvil al tomar decisiones de compra.

En términos de análisis predictivo, se emplearon modelos de regresión lineal y polinomial para predecir los precios de los vehículos usados. Se encontró que el modelo polinomial ofreció una mayor precisión al ajustarse a los datos, lo que puede ser valioso para estimar con mayor exactitud los precios de los automóviles en el mercado.

Sin embargo, es importante destacar que este estudio podría beneficiarse de un análisis más exhaustivo que considere una variedad de variables adicionales o que agrupe las variables existentes de manera diferente. Esto podría permitir la creación de modelos aún más eficientes y precisos para comprender completamente el mercado de vehículos usados en Dallas. En conjunto, estos hallazgos ofrecen información valiosa para compradores y vendedores, ayudándoles a tomar decisiones más informadas en este mercado en constante cambio.

## Preguntas de interés y visualizaciones



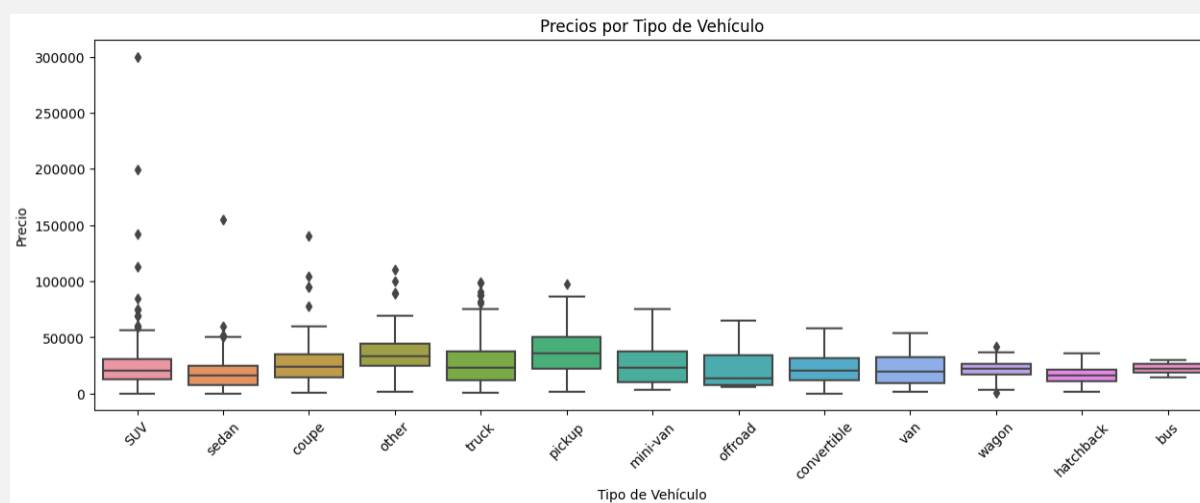
### 1. ¿Cuál es el precio promedio de los automóviles?

El precio promedio de un automóvil es: \$22561.38

Media: \$16134.58

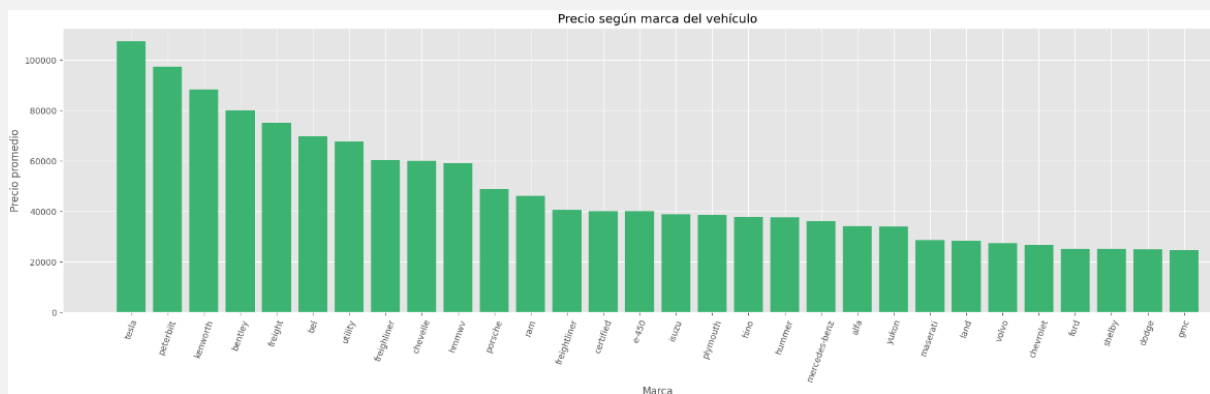
### 2. ¿Cuál es el precio promedio por tipo de automóvil?

pickup	\$37.134,65	other	\$36.103,36
Truck	\$27.707,02	Coupe	\$27.133,35
mini-van	\$24.780,02	SUV	\$24.371,01
offroad	\$24.030,50	convertible	\$22.345,05
bus	\$22.225,00	van	\$21.492,14
wagon	\$21.053,25	sedan	\$17.942,71
hatchback	\$16.180,15		



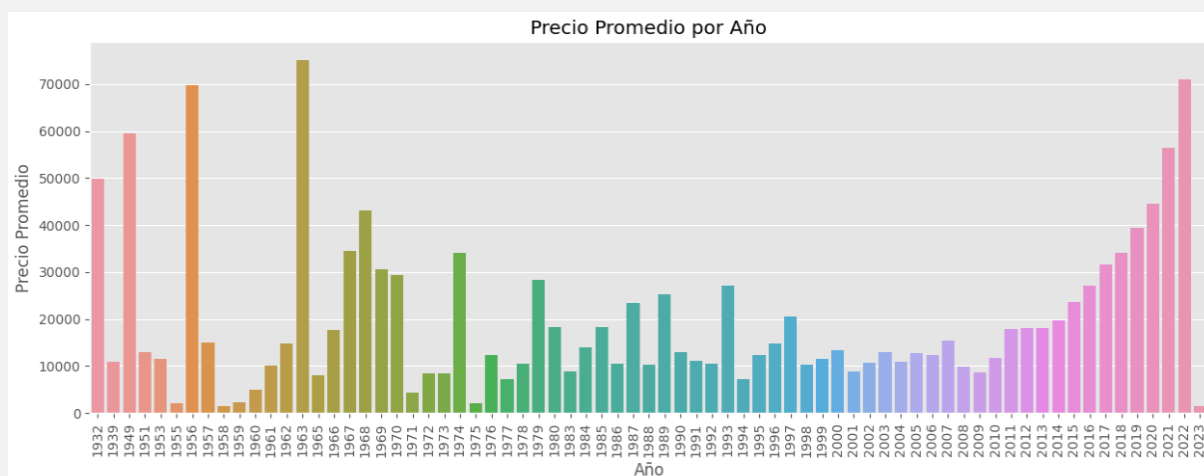
### 3. ¿Cuál es el precio promedio por marca de los automóviles?

tesla	\$107.362,75	peterbilt	\$94.892,5
kenworth	\$88.266,67	bentley	\$79.900
freight	\$75.000	bel	\$69.750
utility	\$67.500	freighliner	\$60.250
chevelle	\$60.000	hmmwv	\$59.000
porsche	\$48.681,36	ram	\$46.729,49
freightliner	\$40.496,15	certified	\$40.000
e-450	\$39.950	isuzu	\$38.701,25
plymouth	\$38.500	mercedes-benz	\$37.970,52
hino	\$37.725	hummer	\$37.630,33
alfa	\$34.172,50	yukon	\$33.785
maserati	\$28.495	chevrolet	\$28.347,46
land	\$28.221,11	volvo	\$26.865,67
gmc	\$26.735,26	ford	\$26.634,42



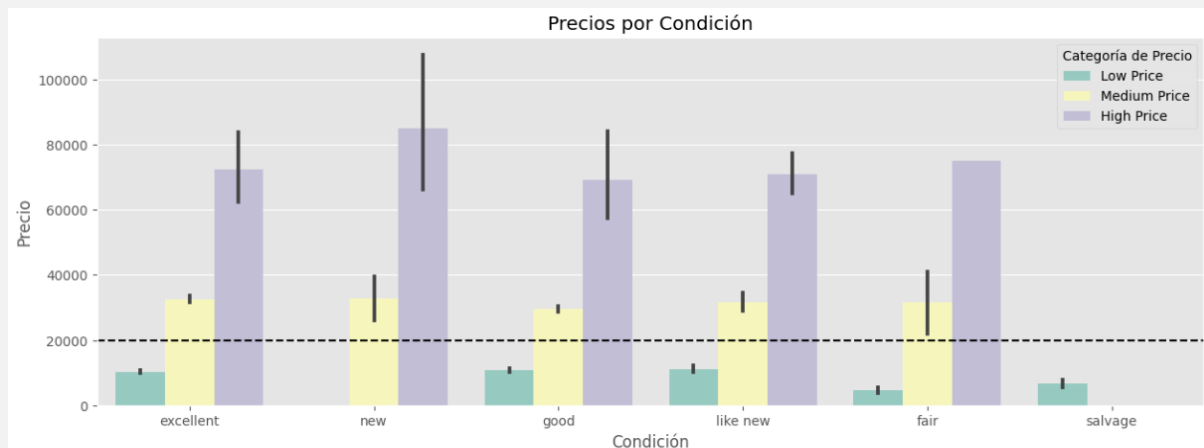
#### 4. ¿Cuál es el precio promedio por año de fabricación?

1932	\$49.900	1939	\$10.900	1949	\$59.500	1951	\$13.000
1953	\$11.466,67	1955	\$2.000	1956	\$69.750	1957	\$15.000
1958	\$1.500	1959	\$2.225	1960	\$5.000	1961	\$10.000
1962	\$14.819,4	1963	\$75.166,67	1965	\$7.966,67	1966	\$17.775
1967	\$34.500,17	1968	\$43.000	1969	\$30.560	1970	\$29.300
1971	\$4.450	1972	\$8.416,67	1973	\$8.500	1974	\$33.998,75
1975	\$2.000	1976	\$12.416,67	1977	\$7.300	1978	\$10.500
1979	\$28.400	1980	\$18.287	1983	\$8.900	1984	\$14.075
1985	\$18.243,75	1986	\$10.577,86	1987	\$23.500	1988	\$10.356
1989	\$25.250	1990	\$12.920	1991	\$11.190,2	1992	\$10.583,33
1993	\$27.174,37	1994	\$7.133,33	1995	\$12.292,86	1996	\$14.900
1997	\$20.612,5	1998	\$10.235,45	1999	\$11.492,5	2000	\$13.431,56
2001	\$8.874,1	2002	\$10.645,93	2003	\$12.922	2004	\$11.006,80
2005	\$12.704,90	2006	\$12.249,32	2007	\$15.382,81	2008	\$9.795,58
2009	\$8.589,64	2010	\$11.704,12	2011	\$17.937,48	2012	\$18.148,30
2013	\$18.112,59	2014	\$19.731,21	2015	\$23.561,82	2016	\$27.104,80
2017	\$31675,08	2018	\$34090,44	2019	\$39481,10	2020	\$44602,54
2021	\$56521,07	2022	\$71067,29	2023	\$1500		



## 5. ¿Cuál es el precio promedio de los vehículos por condición?

Contamos con 4 tipos de estados de vehículos; Excelente, Bueno, Regular y Malo. Para esta ocasión subdividimos cada estado en 3 categorías de precios; Precio bajo, Precio medio y Precio alto. El primer elemento contempla de \$0 a \$19.999, el segundo de \$20.000 a \$49.999 y el tercero igual o superior de \$50.000.





## 6. ¿ Un auto con más kms tiende a ser más económico, pero la transmisión importa?

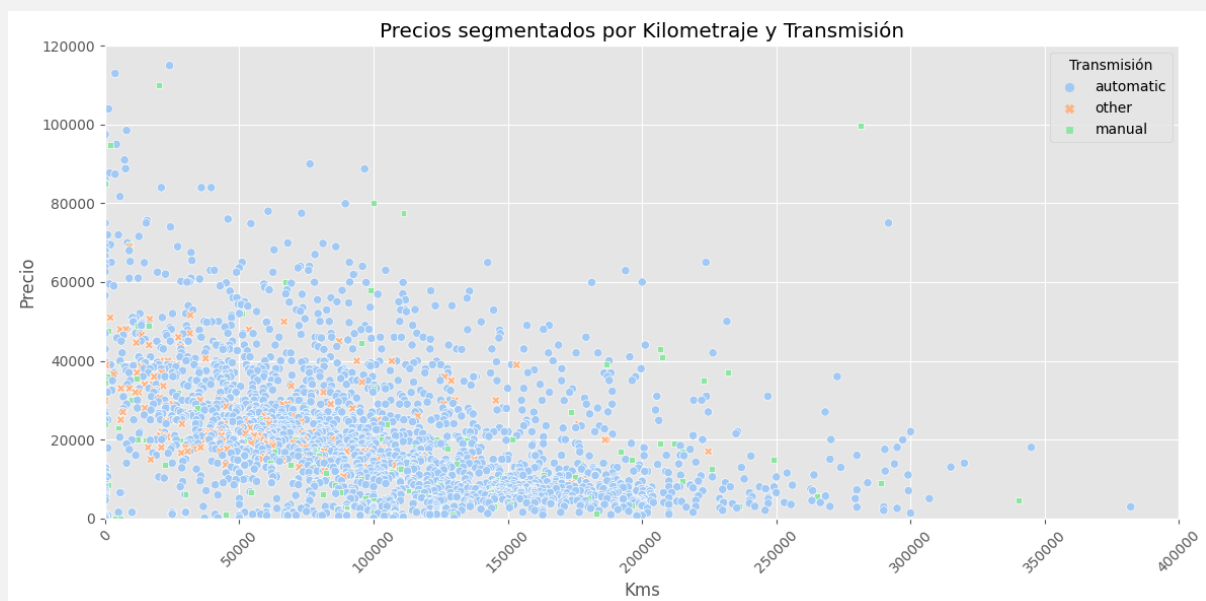
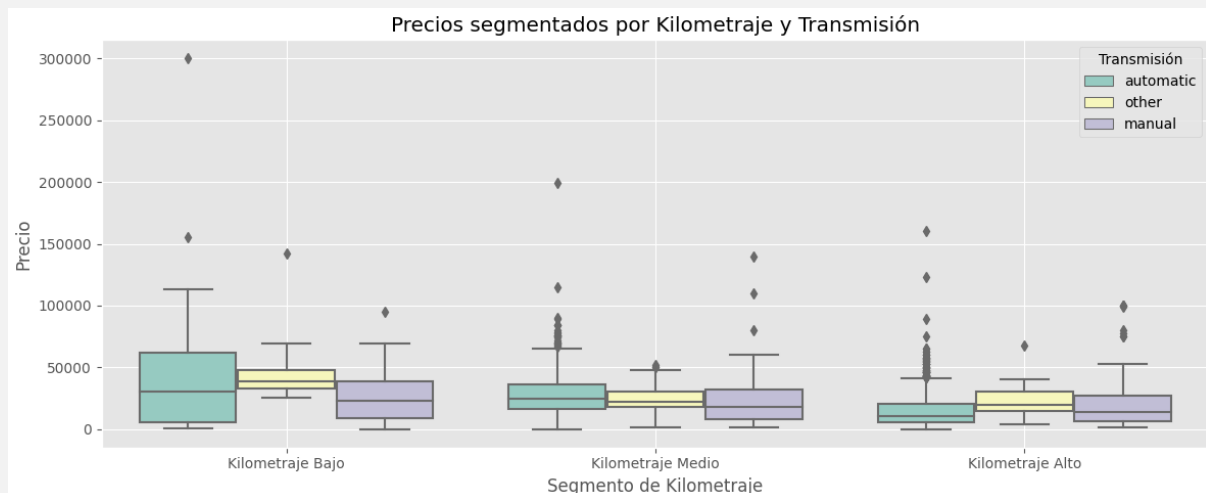
Categorizamos los Kilometrajes vehiculares en 3 segmentos;

Kilometraje Bajo: menor a 10,000 kms.

Kilometraje Medio: Entre 10,000 y 100,000 kms.

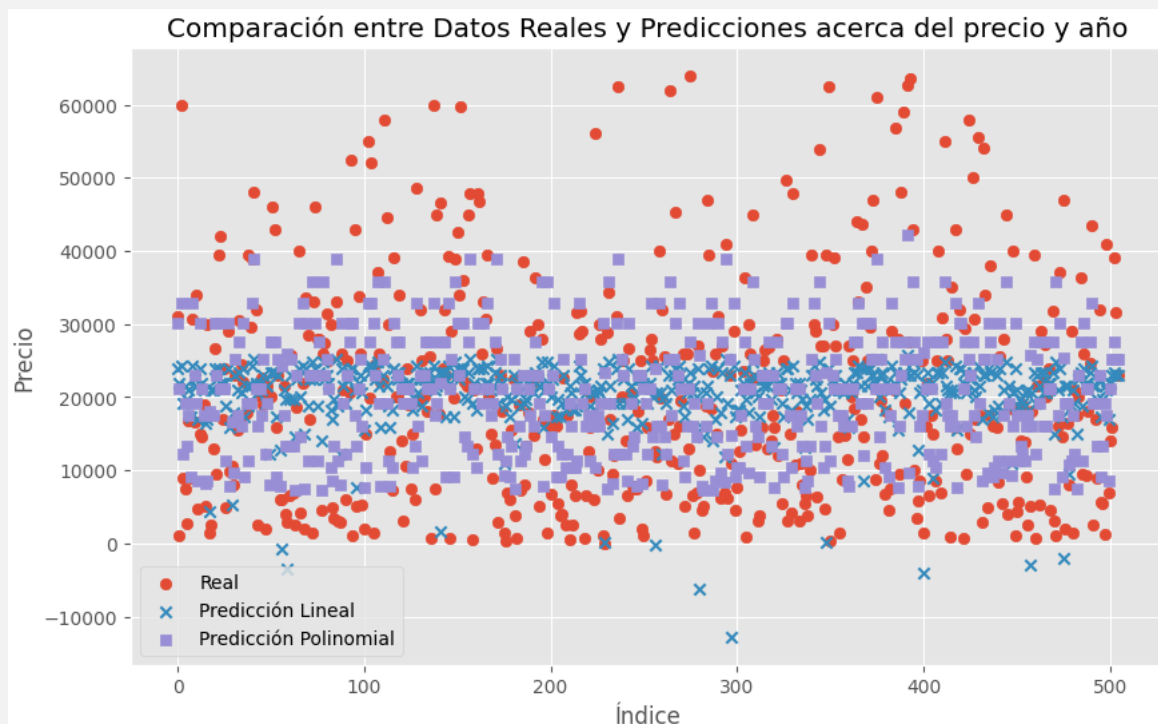
Kilometraje Alto: Mayor a 100,000 kms.

El análisis de los datos y el gráfico indican que si bien el kilometraje de un automóvil puede ser un factor que influye en su precio, no se puede concluir de manera definitiva que un auto con mayor kilometraje sea necesariamente más económico. La transmisión también desempeña un papel levemente significativo en la determinación del precio de un vehículo, como se evidencia en la distribución de precios segmentados por segmentos de kilometraje y tipos de transmisión en el gráfico presentado. Por lo tanto, la relación entre el kilometraje y el precio de un automóvil está influenciada por otros factores, en este caso su transmisión.



## 7. ¿Cómo comparar las predicciones de un modelo lineal y polinomial para predecir el precio de vehículos según su año?

La comparación de modelos lineal y polinomial para predecir el precio de vehículos según su año implica cargar los datos, dividirlos en conjuntos de entrenamiento y prueba, entrenar ambos modelos, realizar predicciones y medir su precisión. Además, se crea un gráfico para visualizar las predicciones frente a los valores reales, lo que proporciona una comprensión visual del rendimiento de ambos modelos en la tarea de predicción de precios de vehículos en función del año.



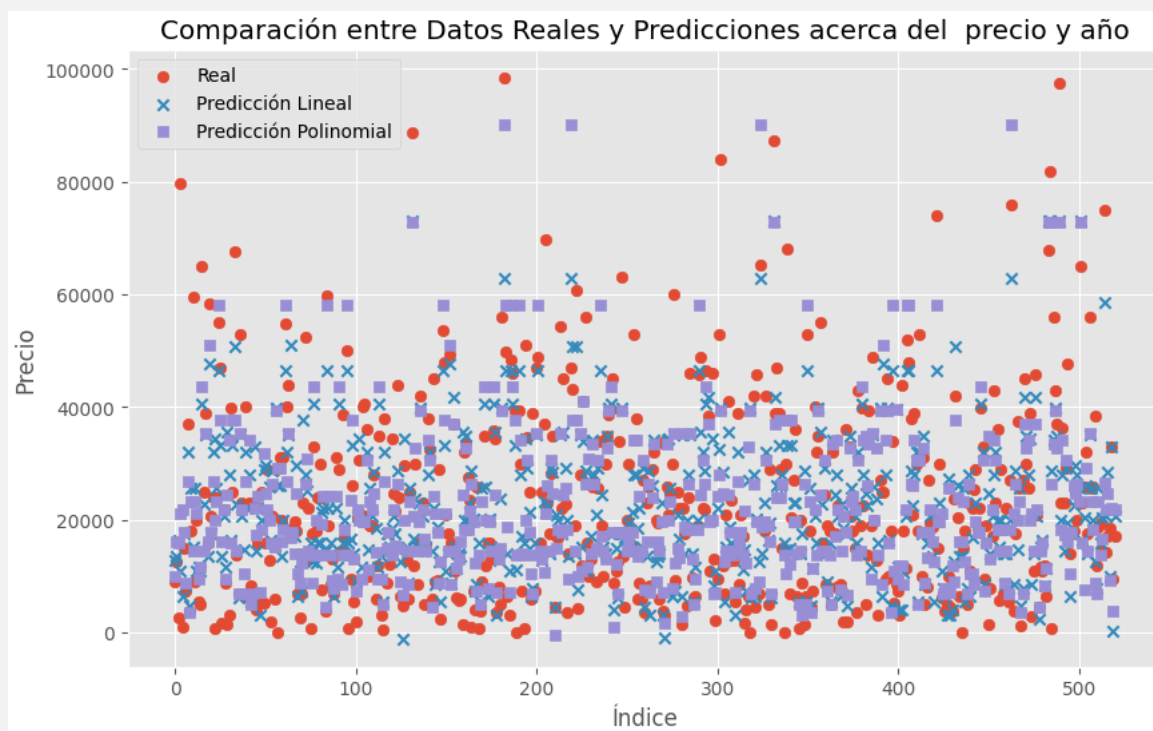
Precisión del modelo polinomial: 36%

Precisión del modelo lineal: 15%

## 8. ¿Cómo se comparan los modelos de regresión lineal y polinomial para predecir el precio de vehículos?

En la tarea de modelado y predicción de precios de vehículos, se implementaron dos enfoques distintos: un modelo de regresión lineal y otro de regresión polinomial. Tras la carga y preprocesamiento de los datos, que incluyó la codificación `OneHotEncoder` de las características categóricas y la imputación de valores faltantes en la variable objetivo (precio), se procedió a dividir el conjunto de datos en conjuntos de entrenamiento y prueba. El modelo de regresión lineal se desplegó para efectuar predicciones lineales convencionales, mientras que se aplicó un modelo de regresión polinomial de grado 3 para explorar posibles relaciones no lineales en los datos. Se calculó la métrica de precisión del modelo polinomial en el conjunto de prueba, y posteriormente se efectuó una comparación visual entre las predicciones generadas por ambos modelos y los valores reales de los precios de los vehículos. Este análisis proporcionó una base sólida para evaluar y determinar cuál de los dos modelos ofreció un rendimiento superior en la estimación de los precios de los vehículos.

La predicción de precios de vehículos basada en características como el title status, drive, año.



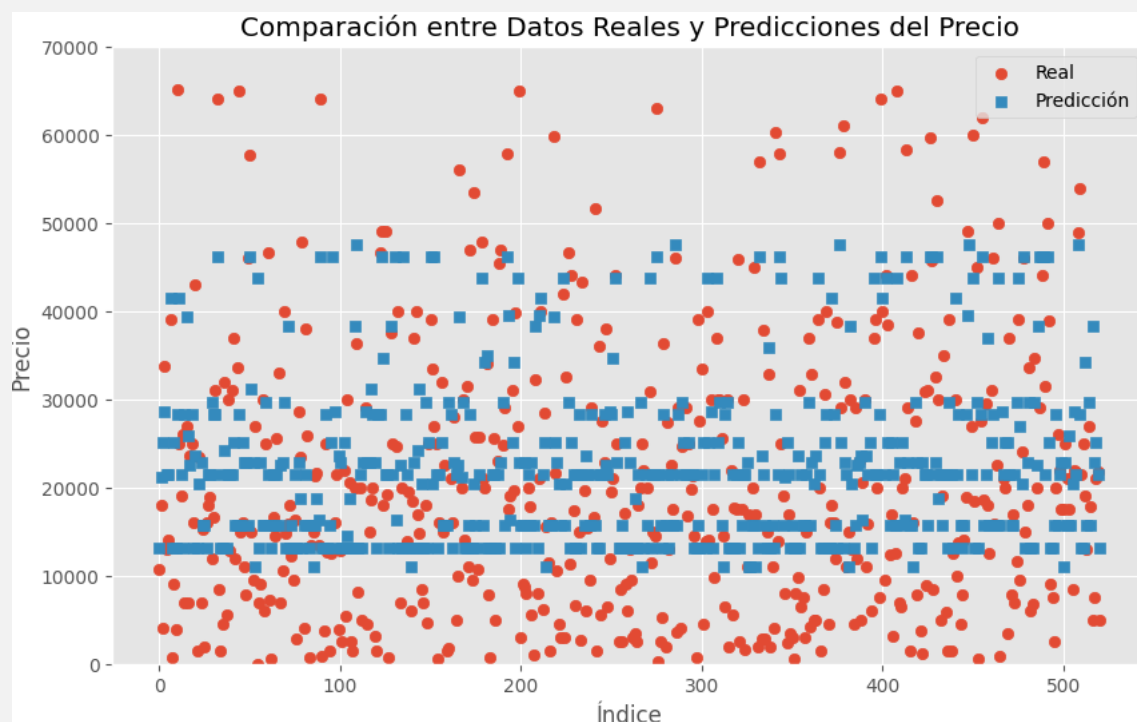
Precisión del modelo: 45%

## 9. ¿Cómo predecir el precio de vehículos utilizando un modelo de regresión lineal?

Se utiliza un enfoque de ingeniería de características mediante `ColumnTransformer` y `OneHotEncoder` para transformar variables categóricas en representaciones numéricas. Luego se crea un pipeline que combina el preprocesamiento con un modelo de regresión lineal utilizando `LinearRegression`.

El código calcula y muestra la precisión del modelo en los datos de prueba, lo que proporciona una métrica cuantitativa de su capacidad para predecir el precio de los vehículos. Además, se visualiza de manera profesional la comparación entre los valores reales y las predicciones mediante un gráfico de dispersión, lo que facilita la interpretación de la calidad del modelo.

La predicción de precios de vehículos está basada en características como el tipo y el combustible.



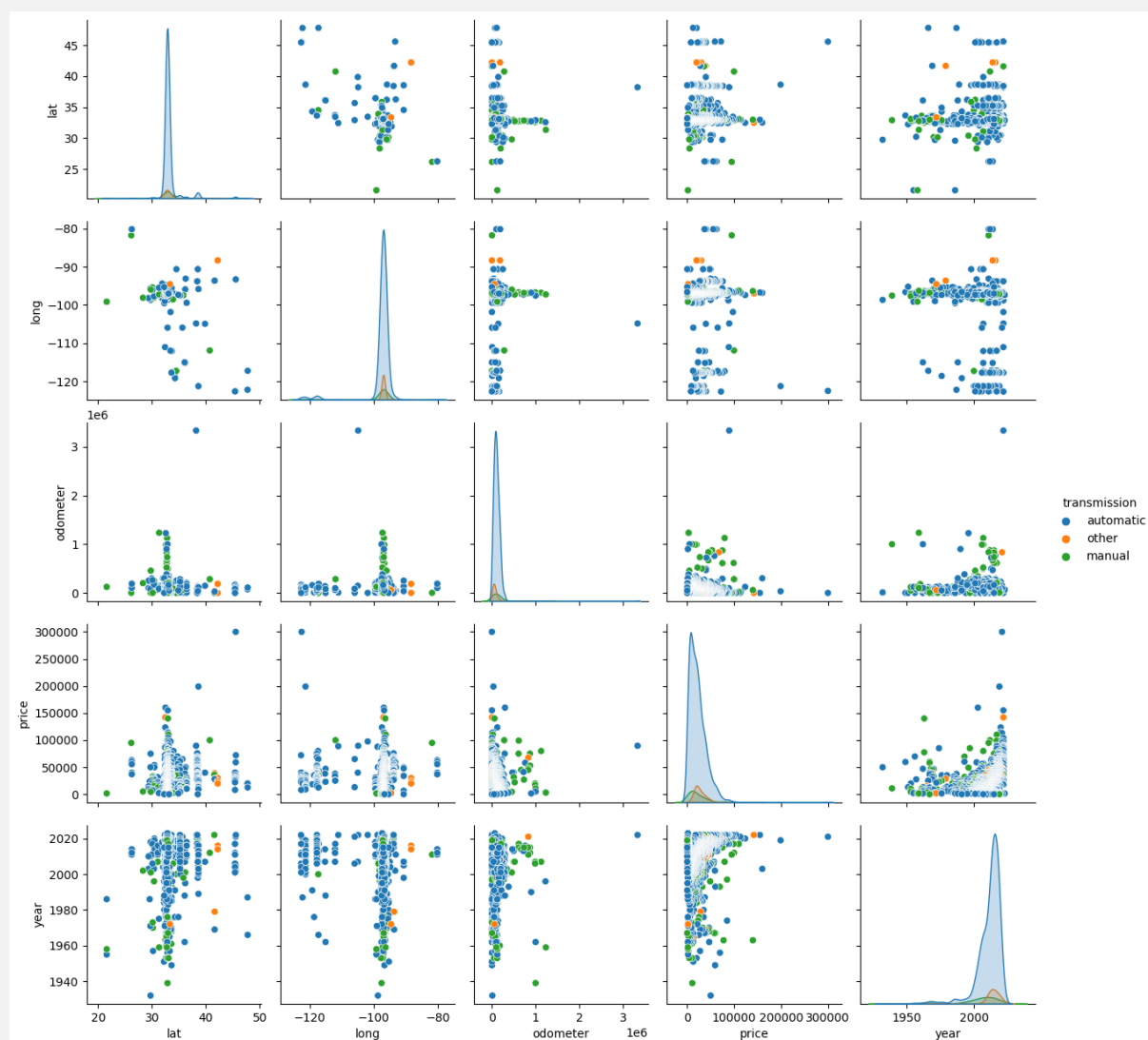
Precisión del modelo: 32%

En este caso final observamos como la inclusión de variables no concluyentes afecta de forma negativa la eficiencia del modelo creado. Es importante seleccionar de forma cuidadosa las características relevantes de modo tal que la calidad del modelo de predicción se mejore y optimice en su rendimiento y no disminuya su precisión.

## 10. ¿Hay alguna relación entre la transmisión de un vehículo y sus otras características?

El diagrama de dispersión de pares ayudara a mostrar cómo se distribuyen los datos para cada variable. Por ejemplo, si la transmisión automática está asociada con un mayor precio de venta, el diagrama de dispersión de pares mostrará que los puntos de datos para los vehículos con transmisión automática están más cerca del extremo superior del rango de precios de venta.

Se muestran que si existen algunas diferencias entre los vehículos con transmisión automática contra los vehículos de transmisión manual. Por ejemplo, los vehículos con transmisión automática tienden a tener un precio de venta más alto, un kilometraje más bajo y un aumento de su producción desde los 1980 en adelante.



## Recomendaciones e Insights



### Insights

1. Relación entre el año de fabricación y el precio: Confirmar si existe una relación inversa significativa entre el año de fabricación y el precio de los vehículos. Se puede hacer un gráfico de dispersión o un análisis de correlación para respaldar esta hipótesis. Esto podría indicar si los vehículos más nuevos tienden a tener precios más altos.
2. Influencia de la marca: Analiza la influencia de la marca en los precios de los vehículos. ¿Las marcas premium tienen precios significativamente más altos que las marcas menos conocidas? Esto podría confirmar la hipótesis de que la marca influye en el precio.
3. Tipo de transmisión y precio: Examina cómo el tipo de transmisión se relaciona con los precios de los vehículos. ¿Los vehículos con transmisión automática tienen precios más altos que los manuales? Esto podría respaldar la hipótesis de que el tipo de transmisión afecta los precios.
4. Condición y kilometraje: Analiza la relación entre la condición del vehículo y el kilometraje con respecto a los precios. ¿Los vehículos en mejores condiciones y con menor kilometraje tienen precios más altos? Esto podría confirmar la importancia de la calidad y el estado general en la determinación de los precios.
5. Predicción de precios: Utiliza modelos de regresión lineal y polinomial para predecir los precios de los vehículos usados. Compara la precisión de estos modelos y considera utilizar el modelo polinomial si ofrece una mayor exactitud en las predicciones.

### Recomendaciones

1. Segmentación de datos: Divide los datos en segmentos por marca, año de fabricación y tipo de transmisión para obtener una comprensión más detallada de cómo estos factores específicos influyen en los precios.
2. Análisis geoespacial: Utiliza la información para determinar si la ubicación en Dallas tiene un impacto en los precios de los vehículos. Es posible que ciertas áreas tengan precios más altos o bajos debido a la demanda de cada estado.
3. Variables adicionales: Considerar la posibilidad de agregar más variables en los análisis, como el tipo de combustible, el tamaño del vehículo o el título del estado, para obtener una imagen más completa de los factores que afectan los precios.
4. Agrupación de datos: Agrupa modelos de automóviles similares para analizar las tendencias de precios dentro de categorías específicas. Por ejemplo, se podría agrupar los vehículos por tipo, es decir; sedán, SUV, camioneta, etc. y/o por capacidad de cilindros.
5. Visualizaciones avanzadas: Utiliza gráficos avanzados como mapas de calor y gráficos de dispersión con color para identificar patrones y relaciones de manera más efectiva.
6. Actualización continua: Dado que el mercado de vehículos usados es dinámico, considera realizar actualizaciones regulares de tu análisis para mantenerlo relevante y útil para compradores y vendedores.