

Predicción del precio de los carros

ÍNDICE

.01

INTRODUCCIÓN

.02

LIMPIEZA
DE DATOS

.03

ANÁLISIS EXPLORATORIO

.04

DESAFIOS



Introducción

Se requiere predecir el precio de los carros con base en : el precio, los kilometros recorridos, el tipo de combustible, la transmisión, la cantidad de dueños que ha tenido, el año de manufactura, el tipo de motor, y la cantidad de asientos.

Conjunto de datos

Descripción del Conjunto de Datos

	car_name	car_prices_in_rupee	kms_driven	fuel_type	transmission	ownership	manufacture	engine	Seats
0	Jeep Compass 2.0 Longitude Option BSIV	10.03 Lakh	86,226 kms	Diesel	Manual	1st Owner	2017	1956 cc	5 Seats
1	Renault Duster RXZ Turbo CVT	12.83 Lakh	13,248 kms	Petrol	Automatic	1st Owner	2021	1330 cc	5 Seats

Limpieza de datos

- No había valores vacíos.
- Se revisó la columna de nombre del carro: esta tiene 1896 tipos diferentes de valores,
- La columna de precio del carro en rupias fue la mpas retante, pues tenía valores en rupias, en lakhs y en Crore. Fue necesario crear una columna nueva para separar el tipo de modena y luego poder convertir cada una en lacks. Como se muestra en la siguiente diapositiva

```
#df["car_prices_in_rupee"] = pd.to_numeric(df["car_prices_in_rupee"], downcast="float")
#aún no se puede convertir en valor número por el error: ValueError: Unable to parse string "1.10 Crore" at position 119
#El crore (cr)(del prácrito krodī) denota diez millones (10,000,000 o 107 en notación científica) y es igual a 100 lakhs
#en el sistema de numeración india.
```

```
#entonces hay que convertir los valores Crore para poder eliminar ese string
```

```
#primero creo otra columna "currency" y asigno la denominación de las monedas.
df.loc[df["car_prices_in_rupee"].str.contains("Lakh"), "currency"] = "Lakh"
df.loc[df["car_prices_in_rupee"].str.contains("Crore"), "currency"] = "Crore"
df.loc[df["currency"].isnull(), "currency"] = "Rupias"
```

```
#luego, elimino los string de la columna car prices
df["car_prices_in_rupee"] = df["car_prices_in_rupee"].str.replace("Lakh", "")
df["car_prices_in_rupee"] = df["car_prices_in_rupee"].str.replace("Crore", "")
```



```
# en donde la columna currency sea x le pido que haga la respectiva conversi3n de la moneda.  
#Todas quedan en Lakh porque en rupias quedaban valores muy grandes
```

```
df["car_prices_in_rupee"] = np.where(df["currency"] == "Lakh", df["car_prices_in_rupee"], 0)  
df["car_prices_in_rupee"] = np.where(df["currency"] == "Crore", df["car_prices_in_rupee"] * 100, df["car_prices_in_rupee"])  
df["car_prices_in_rupee"] = np.where(df["currency"] == "Rupias", df["car_prices_in_rupee"] / 100000, df["car_prices_in_rupee"])
```

```
[ ] #le cambio el nombre a la columna para mayor claridad  
df = df.rename(columns={"car_prices_in_rupee": "car_prices_in_lakh"})  
#y elimino la columna de currency  
df=df.drop(columns="currency",axis=1)
```

```
[ ] df.head()
```

Unnamed: 0		car_name	car_prices_in_lakh	kms_driven	fuel_type	transmission	ownership	manufacture	engine	Seats
0	0	Jeep Compass 2.0 Longitude Option BSIV	10.03	86,226 kms	Diesel	Manual	1st Owner	2017	1956 cc	5 Seats
1	1	Renault Duster RXZ Turbo CVT	12.83	13,248 kms	Petrol	Automatic	1st Owner	2021	1330 cc	5 Seats
2	2	Toyota Camry 2.5 G	16.40	60,343 kms	Petrol	Automatic	1st Owner	2016	2494 cc	5 Seats
3	3	Hyundai i20 N IV CVT	7.77	36,686 kms	Petrol	Automatic	1st Owner	2018	1499 cc	5 Seats

Limpieza de datos

- A la columna de kilómetros recorridos y la de motor se les removió los strings y se convirtieron a valores numéricos.

```
#ahora convertiré a valores numéricos kms_driven
df["kms_driven"] = df["kms_driven"].str.replace("kms", "")
df["kms_driven"] = df["kms_driven"].str.replace(", ", "")
df["kms_driven"] = df["kms_driven"].str.replace(".", "")
df["kms_driven"] = df["kms_driven"].astype(int)
```

```
#ahora voy a a convertir lo valores de engine a numéricos
df["engine"] = df["engine"].str.replace("cc", "")
df["engine"] = df["engine"].astype(int)
```


Limpieza de datos

- En la columna de número de dueños se convirtió a número con el equivalente de número correspondiente

```
#ahora voy a a convertir lo valores de ownership a numéricos, como los valores aumentan, la secuencia de 0 a 5 funciona y  
#y no es necesario la codificación binaria.  
df["ownership"] = df["ownership"].str.replace("0th Owner", "0")  
df["ownership"] = df["ownership"].str.replace("1st Owner", "1")  
df["ownership"] = df["ownership"].str.replace("2nd Owner", "2")  
df["ownership"] = df["ownership"].str.replace("3rd Owner", "3")  
df["ownership"] = df["ownership"].str.replace("4th Owner", "4")  
df["ownership"] = df["ownership"].str.replace("5th Owner", "5")  
df["ownership"] = df["ownership"].astype(int)
```

Limpieza de datos

- En la columna de número de asientos también se convirtió al número correspondiente

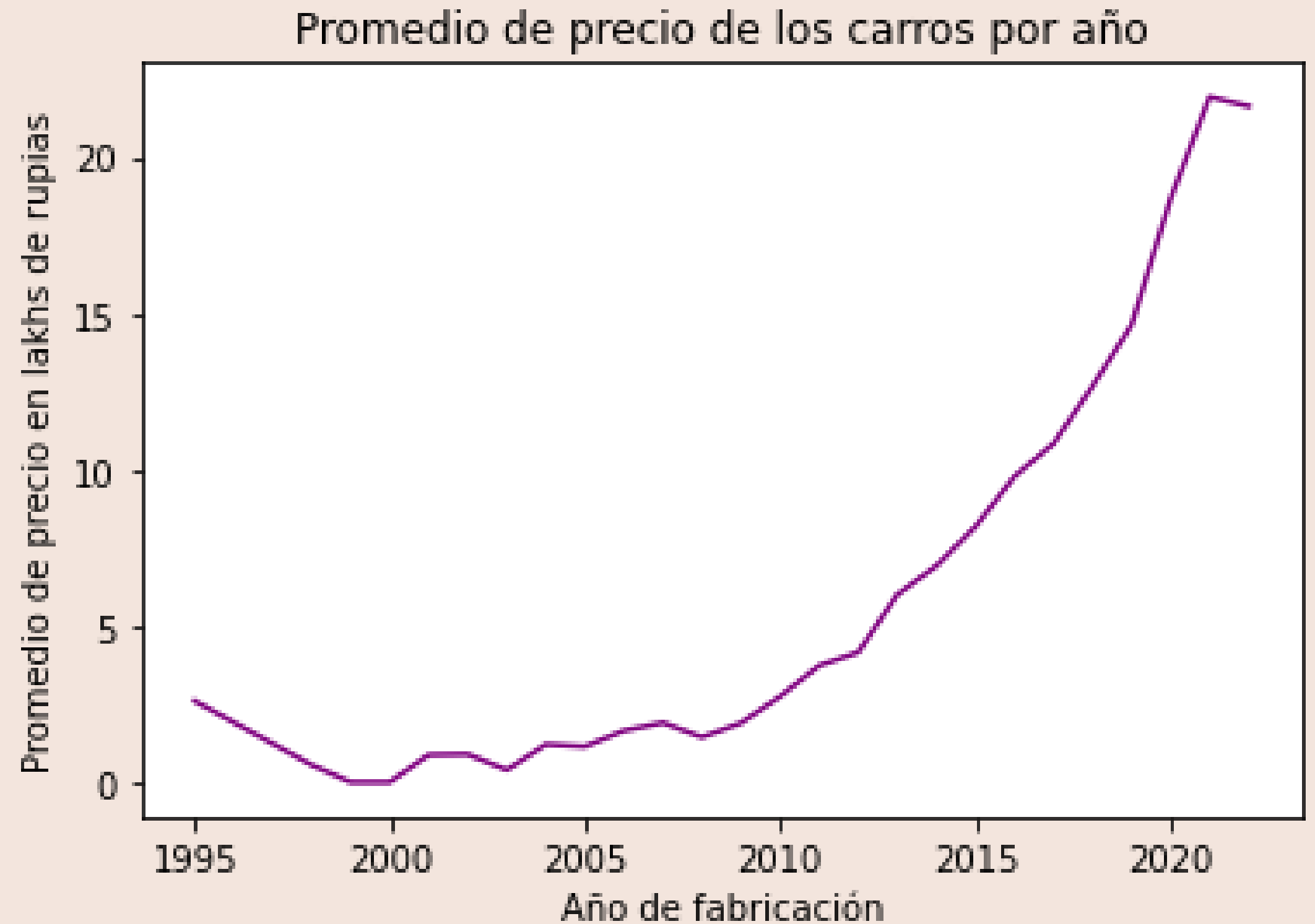
```
#ahora voy a a convertir lo valores de Seats a numéricos
df = df.rename(columns={"Seats": "seats"})
df["seats"] = df["seats"].str.replace("2 Seats", "2")
df["seats"] = df["seats"].str.replace("4 Seats", "4")
df["seats"] = df["seats"].str.replace("5 Seats", "5")
df["seats"] = df["seats"].str.replace("6 Seats", "6")
df["seats"] = df["seats"].str.replace("7 Seats", "7")
df["seats"] = df["seats"].str.replace("8 Seats", "8")
df["seats"] = df["seats"].astype(int)
```



Análisis

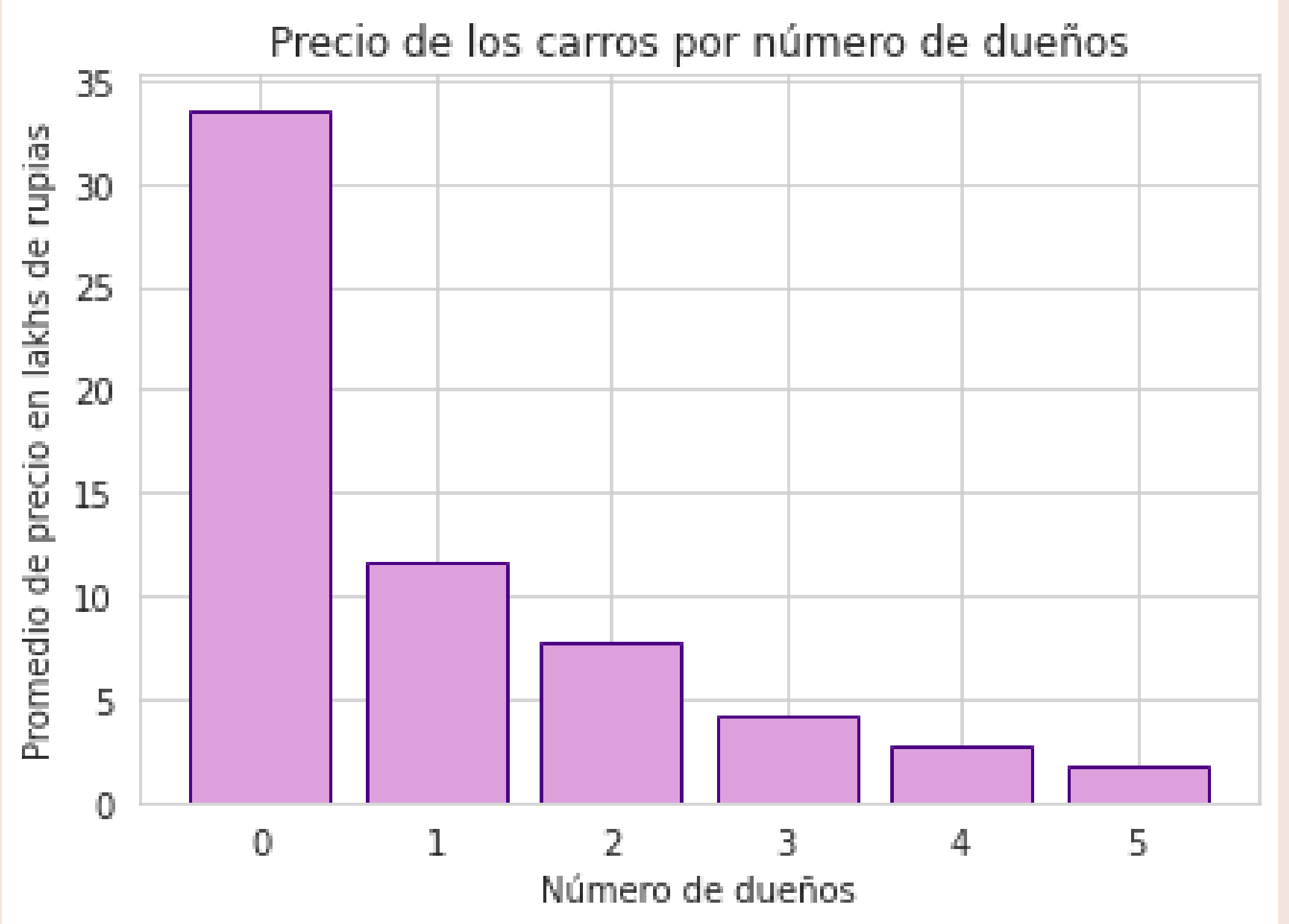
exploratorio

Promedio del precio de los carros en el tiempo



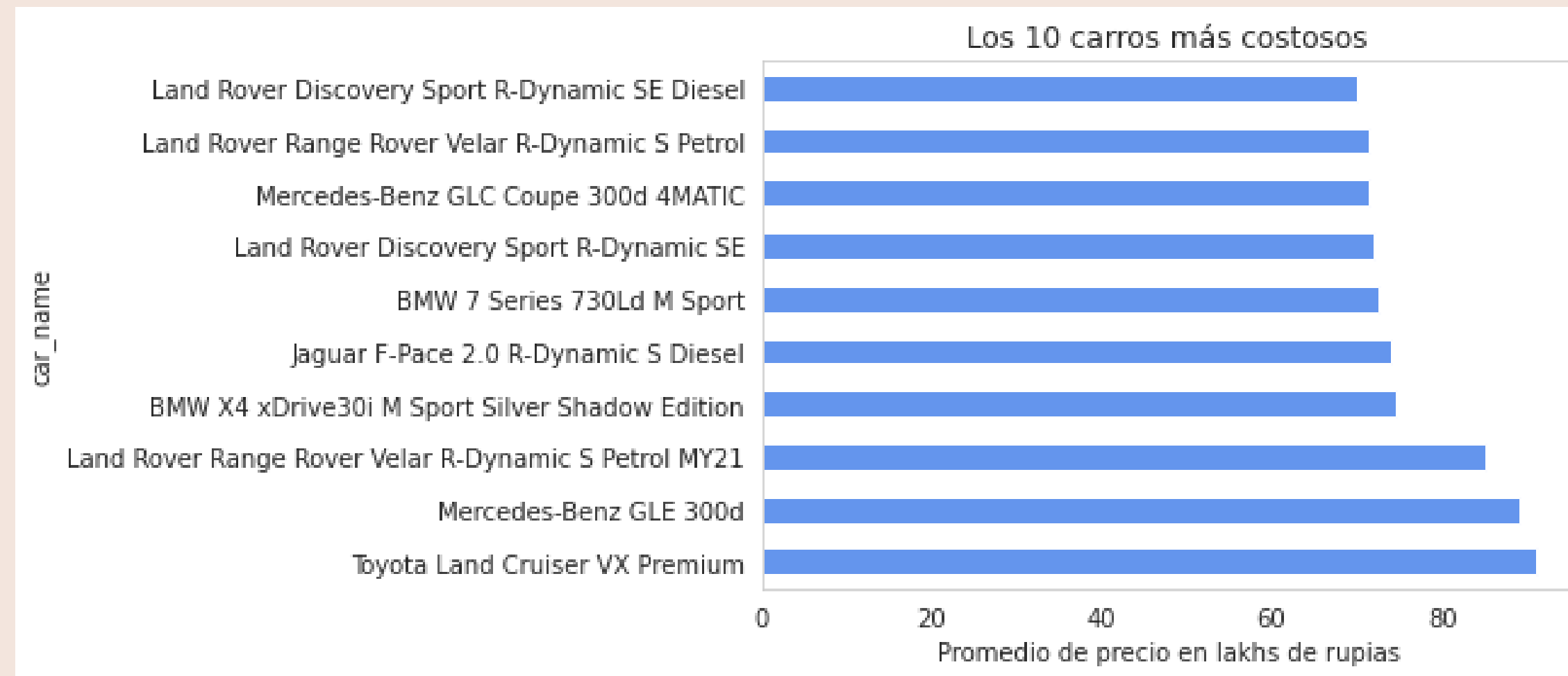
- Se puede observar que el promedio del precio de carros ha aumentado con el paso dle tiempo, lo cual es lógico e intuitivo, pero pudo ser confirmado.

Influencia del número de dueños que ha tenido un carro en el costo

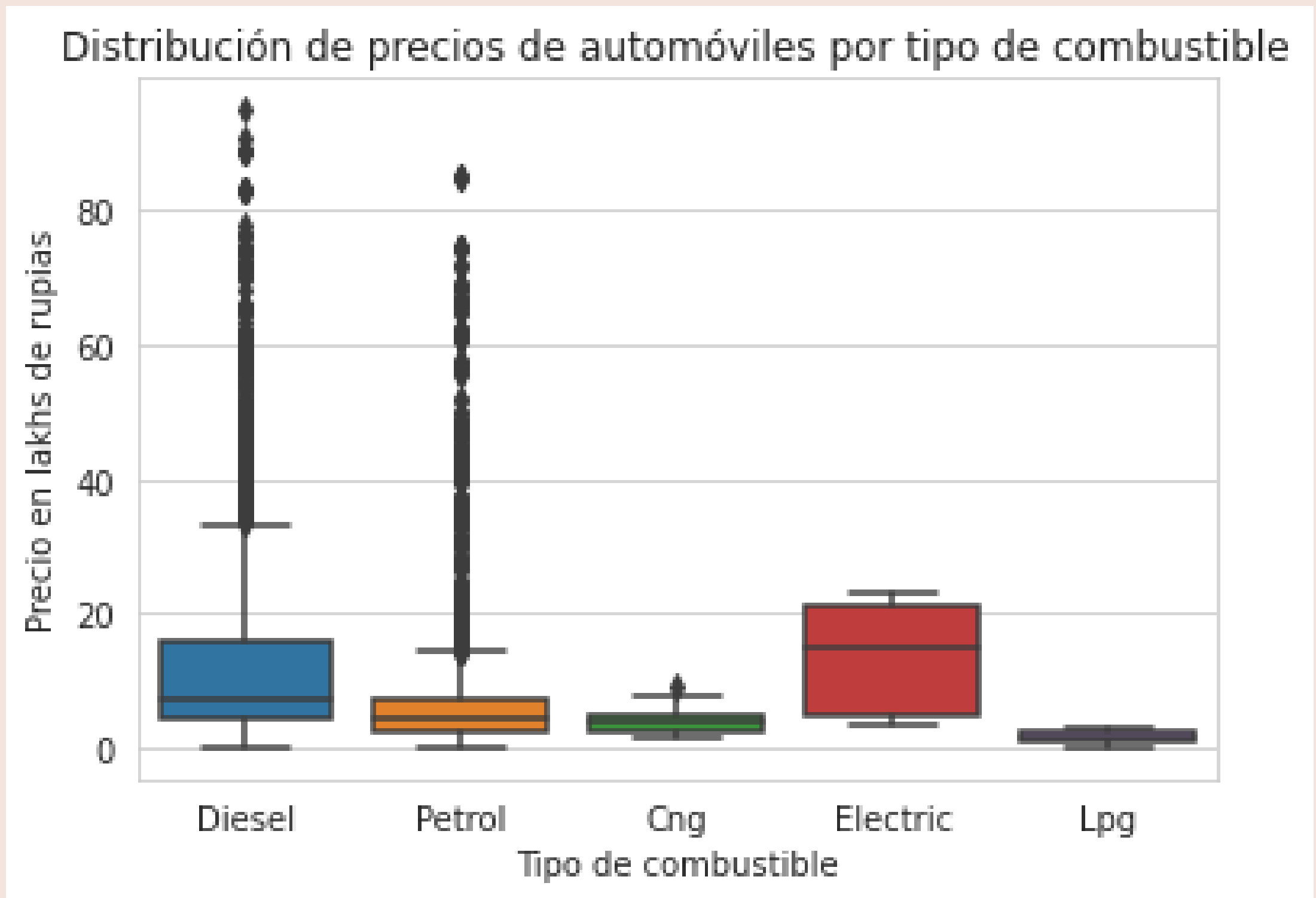


- El precio de los carros tiende a disminuir a medida que ha tenido más dueños. Se evidencia un salto grande de 0 a 1; lo que quiere decir que cuando es nuevo, el precio es bastante más alto; cuando ha tenido de 1 en adelante la diferencia empieza a ser menos significativa.

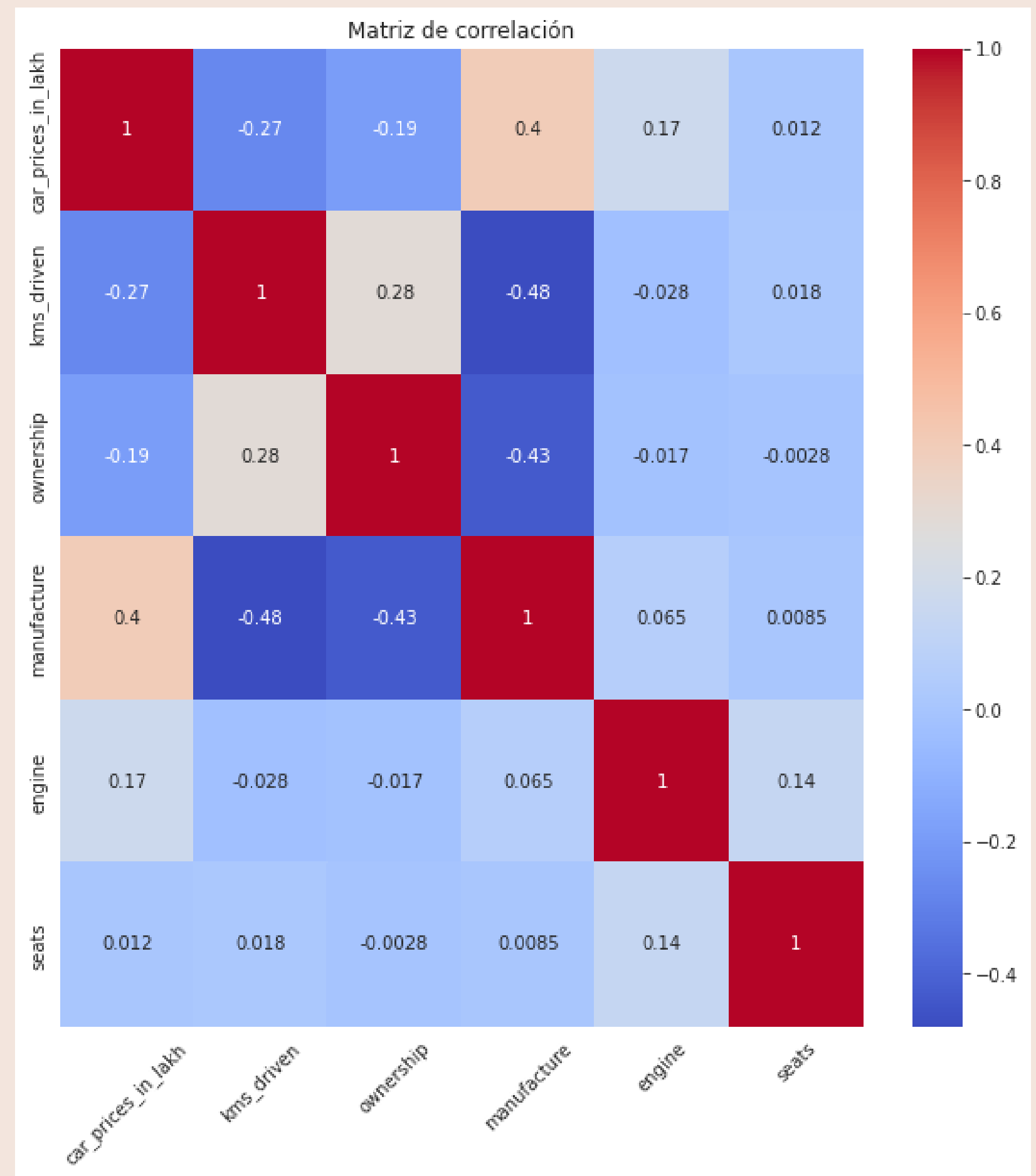
Los 10 carros más caros



Distribución de precios por tipo de combustible



Mapa de calor de correlación



Mapa de calor de correlación

- Se evidencia una correlación fuerte y positiva entre "engine" y "car prices", lo que indica que a medida que el tamaño del motor de un automóvil aumenta, el precio también aumenta
- El cuadro de correlación manufacture con kms_driven tiene un valor de 0.48 indica que existe una correlación moderada negativa entre las dos variables, lo que significa que a medida que el año de fabricación aumenta, la cantidad de kilómetros que el automóvil ha sido conducido disminuye
-