

```
In [1]: print("hello")
```

```
hello
```

```
In [2]: pip install numpy==2.2.0
```

```
Collecting numpy==2.2.0
  Downloading numpy-2.2.0-cp313-cp313-win_amd64.whl.metadata (60 kB)
  Downloading numpy-2.2.0-cp313-cp313-win_amd64.whl (12.6 MB)
    ----- 0.0/12.6 MB ? eta -:---:-
    ----- 0.3/12.6 MB ? eta -:---:-
    ----- 1.3/12.6 MB 4.5 MB/s eta 0:00:03
    ----- 2.4/12.6 MB 4.6 MB/s eta 0:00:03
    ----- 2.9/12.6 MB 4.6 MB/s eta 0:00:03
    ----- 3.9/12.6 MB 4.5 MB/s eta 0:00:02
    ----- 4.7/12.6 MB 4.2 MB/s eta 0:00:02
    ----- 5.5/12.6 MB 4.1 MB/s eta 0:00:02
    ----- 6.3/12.6 MB 4.2 MB/s eta 0:00:02
    ----- 6.8/12.6 MB 4.0 MB/s eta 0:00:02
    ----- 7.3/12.6 MB 3.8 MB/s eta 0:00:02
    ----- 8.4/12.6 MB 3.9 MB/s eta 0:00:02
    ----- 9.4/12.6 MB 4.0 MB/s eta 0:00:01
    ----- 10.7/12.6 MB 4.2 MB/s eta 0:00:01
    ----- 11.8/12.6 MB 4.3 MB/s eta 0:00:01
    ----- 12.6/12.6 MB 4.3 MB/s 0:00:03

Installing collected packages: numpy
  Attempting uninstall: numpy
    Found existing installation: numpy 2.2.6
    Uninstalling numpy-2.2.6:
      Successfully uninstalled numpy-2.2.6
Successfully installed numpy-2.2.0
Note: you may need to restart the kernel to use updated packages.
```

```
In [3]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [4]: import matplotlib
np.__version__, pd.__version__, sns.__version__, matplotlib.__version__
```

```
Out[4]: ('2.2.0', '2.3.2', '0.13.2', '3.10.5')
```

```
In [5]: df=pd.read_csv("Cars.csv")
```

```
In [6]: df.head()
```

Out[6]:

| | name | year | selling_price | km_driven | fuel | seller_type | transmission | owner | mileage |
|---|------------------------------|------|---------------|-----------|--------|-------------|--------------|--------------|---------|
| 0 | Maruti Swift Dzire VDI | 2014 | 450000 | 145500 | Diesel | Individual | Manual | First Owner | 2 years |
| 1 | Skoda Rapid 1.5 TDI Ambition | 2014 | 370000 | 120000 | Diesel | Individual | Manual | Second Owner | 2 years |
| 2 | Honda City 2017-2020 EXi | 2006 | 158000 | 140000 | Petrol | Individual | Manual | Third Owner | 1 year |
| 3 | Hyundai i20 Sportz Diesel | 2010 | 225000 | 127000 | Diesel | Individual | Manual | First Owner | 2 years |
| 4 | Maruti Swift VXI BSIII | 2007 | 130000 | 120000 | Petrol | Individual | Manual | First Owner | 1 year |



In [7]: df.shape

Out[7]: (8128, 13)

In [8]: df.describe()

Out[8]:

| | year | selling_price | km_driven | seats |
|--------------|-------------|---------------|--------------|-------------|
| count | 8128.000000 | 8.128000e+03 | 8.128000e+03 | 7907.000000 |
| mean | 2013.804011 | 6.382718e+05 | 6.981951e+04 | 5.416719 |
| std | 4.044249 | 8.062534e+05 | 5.655055e+04 | 0.959588 |
| min | 1983.000000 | 2.999900e+04 | 1.000000e+00 | 2.000000 |
| 25% | 2011.000000 | 2.549990e+05 | 3.500000e+04 | 5.000000 |
| 50% | 2015.000000 | 4.500000e+05 | 6.000000e+04 | 5.000000 |
| 75% | 2017.000000 | 6.750000e+05 | 9.800000e+04 | 5.000000 |
| max | 2020.000000 | 1.000000e+07 | 2.360457e+06 | 14.000000 |

In [9]: df.rename(columns={'seller_type':'seller',
'fuel_type':'fuel',
'body_type':'body',
'engine_type':'engine',
'transmission_type':'transmission',

```
'mileage':'mileage',
'max_power':'power',
'km_driven':'driven',
'selling_price':'price'
}, inplace=True)
```

In [10]: df.columns

Out[10]: Index(['name', 'year', 'price', 'driven', 'fuel', 'seller', 'transmission',
'owner', 'mileage', 'engine', 'power', 'torque', 'seats'],
dtype='object')

In [11]: df.head(15)

| | | name | year | price | driven | fuel | seller | transmission | owner | mileage | ei |
|----|--------------------------------------|------|--------|--------|--------|------------|--------|--------------|--------------|------------|----|
| 0 | Maruti Swift Dzire VDI | 2014 | 450000 | 145500 | Diesel | Individual | | Manual | First Owner | 23.4 kmpl | |
| 1 | Skoda Rapid 1.5 TDI Ambition | 2014 | 370000 | 120000 | Diesel | Individual | | Manual | Second Owner | 21.14 kmpl | |
| 2 | Honda City 2017-2020 EXi | 2006 | 158000 | 140000 | Petrol | Individual | | Manual | Third Owner | 17.7 kmpl | |
| 3 | Hyundai i20 Sportz Diesel | 2010 | 225000 | 127000 | Diesel | Individual | | Manual | First Owner | 23.0 kmpl | |
| 4 | Maruti Swift VXI BSIII | 2007 | 130000 | 120000 | Petrol | Individual | | Manual | First Owner | 16.1 kmpl | |
| 5 | Hyundai Xcent 1.2 VTi E Plus | 2017 | 440000 | 45000 | Petrol | Individual | | Manual | First Owner | 20.14 kmpl | |
| 6 | Maruti Wagon R LXI DUO BSIII | 2007 | 96000 | 175000 | LPG | Individual | | Manual | First Owner | 17.3 km/kg | |
| 7 | Maruti 800 DX BSII | 2001 | 45000 | 5000 | Petrol | Individual | | Manual | Second Owner | 16.1 kmpl | 7 |
| 8 | Toyota Etios VXD | 2011 | 350000 | 90000 | Diesel | Individual | | Manual | First Owner | 23.59 kmpl | |
| 9 | Ford Figo Diesel Celebration Edition | 2013 | 200000 | 169000 | Diesel | Individual | | Manual | First Owner | 20.0 kmpl | |
| 10 | Renault Duster 110PS Diesel RxL | 2014 | 500000 | 68000 | Diesel | Individual | | Manual | Second Owner | 19.01 kmpl | |
| 11 | Maruti Zen LX | 2005 | 92000 | 100000 | Petrol | Individual | | Manual | Second Owner | 17.3 kmpl | 9 |
| 12 | Maruti Swift Dzire VDI | 2009 | 280000 | 140000 | Diesel | Individual | | Manual | Second Owner | 19.3 kmpl | |

| | | name | year | price | driven | fuel | seller | transmission | owner | mileage | ei |
|----|--|--------------------------|------|--------|--------|--------|------------|--------------|--------------|-----------|----|
| 13 | | Maruti Swift 1.3 VXi | 2007 | 200000 | 80000 | Petrol | Individual | Manual | Second Owner | NaN | |
| 14 | | Maruti Wagon R LXI Minor | 2009 | 180000 | 90000 | Petrol | Individual | Manual | Second Owner | 18.9 kmpl | |

In [12]: `print(df["owner"].unique())`

```
['First Owner' 'Second Owner' 'Third Owner' 'Fourth & Above Owner'
 'Test Drive Car']
```

In [13]: `OwnerMap = {`

```
'First Owner':1,
'Second Owner':2,
'Third Owner':3,
'Fourth & Above Owner':4,
'Test Drive Car':5
```

```
}  
df['owner'] = df['owner'].map(OwnerMap)
```

In [14]: `dropdata=df[df['fuel'].isin(['CNG', 'LPG'])].index
df=df.drop(dropdata)`

In [15]: `df.shape`

Out[15]: (8033, 13)

In [16]: `df.head(20)`

Out[16]:

| | | name | year | price | driven | fuel | seller | transmission | owner | mileage | en |
|----|--|--------------------------------------|------|--------|--------|--------|------------|--------------|-------|------------|----|
| 0 | | Maruti Swift Dzire VDI | 2014 | 450000 | 145500 | Diesel | Individual | Manual | 1 | 23.4 kmpl | |
| 1 | | Skoda Rapid 1.5 TDI Ambition | 2014 | 370000 | 120000 | Diesel | Individual | Manual | 2 | 21.14 kmpl | |
| 2 | | Honda City 2017-2020 EXi | 2006 | 158000 | 140000 | Petrol | Individual | Manual | 3 | 17.7 kmpl | |
| 3 | | Hyundai i20 Sportz Diesel | 2010 | 225000 | 127000 | Diesel | Individual | Manual | 1 | 23.0 kmpl | |
| 4 | | Maruti Swift VXI BSIII | 2007 | 130000 | 120000 | Petrol | Individual | Manual | 1 | 16.1 kmpl | |
| 5 | | Hyundai Xcent 1.2 VTIVT E Plus | 2017 | 440000 | 45000 | Petrol | Individual | Manual | 1 | 20.14 kmpl | |
| 7 | | Maruti 800 DX BSII | 2001 | 45000 | 5000 | Petrol | Individual | Manual | 2 | 16.1 kmpl | 79 |
| 8 | | Toyota Etios VXD | 2011 | 350000 | 90000 | Diesel | Individual | Manual | 1 | 23.59 kmpl | |
| 9 | | Ford Figo Diesel Celebration Edition | 2013 | 200000 | 169000 | Diesel | Individual | Manual | 1 | 20.0 kmpl | |
| 10 | | Renault Duster 110PS Diesel RxL | 2014 | 500000 | 68000 | Diesel | Individual | Manual | 2 | 19.01 kmpl | |
| 11 | | Maruti Zen LX | 2005 | 92000 | 100000 | Petrol | Individual | Manual | 2 | 17.3 kmpl | 99 |
| 12 | | Maruti Swift Dzire VDi | 2009 | 280000 | 140000 | Diesel | Individual | Manual | 2 | 19.3 kmpl | |
| 13 | | Maruti Swift 1.3 VXi | 2007 | 200000 | 80000 | Petrol | Individual | Manual | 2 | NaN | |

| | | name | year | price | driven | fuel | seller | transmission | owner | mileage | en |
|----|--|--------------------------------------|------|--------|--------|--------|------------|--------------|-------|------------|----|
| 14 | | Maruti Wagon R LXI Minor | 2009 | 180000 | 90000 | Petrol | Individual | Manual | 2 | 18.9 kmpl | |
| 15 | | Mahindra KUV 100 mFALCON G80 K8 5str | 2016 | 400000 | 40000 | Petrol | Individual | Manual | 1 | 18.15 kmpl | |
| 16 | | Maruti Ertiga SHVS VDI | 2016 | 778000 | 70000 | Diesel | Individual | Manual | 2 | 24.52 kmpl | |
| 17 | | Hyundai i20 1.4 CRDi Asta | 2012 | 500000 | 53000 | Diesel | Individual | Manual | 2 | 23.0 kmpl | |
| 18 | | Maruti Alto LX | 2002 | 150000 | 80000 | Petrol | Individual | Manual | 2 | 19.7 kmpl | 79 |
| 19 | | Hyundai i20 2015-2017 Asta 1.4 CRDi | 2016 | 680000 | 100000 | Diesel | Individual | Manual | 1 | 22.54 kmpl | |
| 20 | | Mahindra Verito 1.5 D4 BSIII | 2011 | 174000 | 100000 | Diesel | Individual | Manual | 2 | 21.0 kmpl | |

```
In [17]: split=df.mileage.str.split(" ").str[0].astype(float)
df.mileage=split
```

```
In [18]: df.head()
```

Out[18]:

| | | name | year | price | driven | fuel | seller | transmission | owner | mileage | engin |
|----------|--|------------------------------|-------------|--------------|---------------|-------------|---------------|---------------------|--------------|----------------|--------------|
| 0 | | Maruti Swift Dzire VDI | 2014 | 450000 | 145500 | Diesel | Individual | Manual | 1 | 23.40 | 124 C |
| 1 | | Skoda Rapid 1.5 TDI Ambition | 2014 | 370000 | 120000 | Diesel | Individual | Manual | 2 | 21.14 | 149 C |
| 2 | | Honda City 2017-2020 EXi | 2006 | 158000 | 140000 | Petrol | Individual | Manual | 3 | 17.70 | 149 C |
| 3 | | Hyundai i20 Sportz Diesel | 2010 | 225000 | 127000 | Diesel | Individual | Manual | 1 | 23.00 | 139 C |
| 4 | | Maruti Swift VXI BSIII | 2007 | 130000 | 120000 | Petrol | Individual | Manual | 1 | 16.10 | 129 C |



In [19]: `split_engine=df.engine.str.split(" ").str[0].astype(float)
df.engine=split_engine`

In [20]: `df.head()`

Out[20]:

| | name | year | price | driven | fuel | seller | transmission | owner | mileage | engin |
|---|------------------------------|-------------|--------------|---------------|-------------|---------------|---------------------|--------------|----------------|--------------|
| 0 | Maruti Swift Dzire VDI | 2014 | 450000 | 145500 | Diesel | Individual | Manual | 1 | 23.40 | 1248. |
| 1 | Skoda Rapid 1.5 TDI Ambition | 2014 | 370000 | 120000 | Diesel | Individual | Manual | 2 | 21.14 | 1498. |
| 2 | Honda City 2017-2020 EXi | 2006 | 158000 | 140000 | Petrol | Individual | Manual | 3 | 17.70 | 1497. |
| 3 | Hyundai i20 Sportz Diesel | 2010 | 225000 | 127000 | Diesel | Individual | Manual | 1 | 23.00 | 1396. |
| 4 | Maruti Swift VXI BSIII | 2007 | 130000 | 120000 | Petrol | Individual | Manual | 1 | 16.10 | 1298. |



In [21]: `split_power=df.power.str.split(" ").str[0].astype(float)
df.power=split_power`

In [22]: `df.head()`

Out[22]:

| | name | year | price | driven | fuel | seller | transmission | owner | mileage | engin |
|----------|------------------------------|-------------|--------------|---------------|-------------|---------------|---------------------|--------------|----------------|--------------|
| 0 | Maruti Swift Dzire VDI | 2014 | 450000 | 145500 | Diesel | Individual | Manual | 1 | 23.40 | 1248. |
| 1 | Skoda Rapid 1.5 TDI Ambition | 2014 | 370000 | 120000 | Diesel | Individual | Manual | 2 | 21.14 | 1498. |
| 2 | Honda City 2017-2020 EXi | 2006 | 158000 | 140000 | Petrol | Individual | Manual | 3 | 17.70 | 1497. |
| 3 | Hyundai i20 Sportz Diesel | 2010 | 225000 | 127000 | Diesel | Individual | Manual | 1 | 23.00 | 1396. |
| 4 | Maruti Swift VXI BSIII | 2007 | 130000 | 120000 | Petrol | Individual | Manual | 1 | 16.10 | 1298. |



In [23]: `df.shape`

Out[23]: (8033, 13)

In [24]: `split_name = df.name.str.split(" ").str[0]`
`df.name = split_name`

In [25]: `df.head()`

Out[25]:

| | name | year | price | driven | fuel | seller | transmission | owner | mileage | engine |
|---|---------|------|--------|--------|--------|------------|--------------|-------|---------|--------|
| 0 | Maruti | 2014 | 450000 | 145500 | Diesel | Individual | Manual | 1 | 23.40 | 1248.0 |
| 1 | Skoda | 2014 | 370000 | 120000 | Diesel | Individual | Manual | 2 | 21.14 | 1498.0 |
| 2 | Honda | 2006 | 158000 | 140000 | Petrol | Individual | Manual | 3 | 17.70 | 1497.0 |
| 3 | Hyundai | 2010 | 225000 | 127000 | Diesel | Individual | Manual | 1 | 23.00 | 1396.0 |
| 4 | Maruti | 2007 | 130000 | 120000 | Petrol | Individual | Manual | 1 | 16.10 | 1298.0 |



In [26]: `del df['torque']`

In [27]: `df.head()`

Out[27]:

| | name | year | price | driven | fuel | seller | transmission | owner | mileage | engine |
|---|---------|------|--------|--------|--------|------------|--------------|-------|---------|--------|
| 0 | Maruti | 2014 | 450000 | 145500 | Diesel | Individual | Manual | 1 | 23.40 | 1248.0 |
| 1 | Skoda | 2014 | 370000 | 120000 | Diesel | Individual | Manual | 2 | 21.14 | 1498.0 |
| 2 | Honda | 2006 | 158000 | 140000 | Petrol | Individual | Manual | 3 | 17.70 | 1497.0 |
| 3 | Hyundai | 2010 | 225000 | 127000 | Diesel | Individual | Manual | 1 | 23.00 | 1396.0 |
| 4 | Maruti | 2007 | 130000 | 120000 | Petrol | Individual | Manual | 1 | 16.10 | 1298.0 |



In [28]: `dropowner=df[df['owner'].isin([5])].index
df=df.drop(dropowner)`

In [29]: `df['owner'].isin([5]).sum()`

Out[29]: `np.int64(0)`

In [30]: `df.head()`

Out[30]:

| | name | year | price | driven | fuel | seller | transmission | owner | mileage | engine |
|---|---------|------|--------|--------|--------|------------|--------------|-------|---------|--------|
| 0 | Maruti | 2014 | 450000 | 145500 | Diesel | Individual | Manual | 1 | 23.40 | 1248.0 |
| 1 | Skoda | 2014 | 370000 | 120000 | Diesel | Individual | Manual | 2 | 21.14 | 1498.0 |
| 2 | Honda | 2006 | 158000 | 140000 | Petrol | Individual | Manual | 3 | 17.70 | 1497.0 |
| 3 | Hyundai | 2010 | 225000 | 127000 | Diesel | Individual | Manual | 1 | 23.00 | 1396.0 |
| 4 | Maruti | 2007 | 130000 | 120000 | Petrol | Individual | Manual | 1 | 16.10 | 1298.0 |



In [31]: `df['price']`

Out[31]:

```
0      450000
1      370000
2      158000
3      225000
4      130000
       ...
8123    320000
8124    135000
8125    382000
8126    290000
8127    290000
Name: price, Length: 8028, dtype: int64
```

In [32]: `df.head()`

Out[32]:

| | name | year | price | driven | fuel | seller | transmission | owner | mileage | engine |
|---|---------|------|--------|--------|--------|------------|--------------|-------|---------|--------|
| 0 | Maruti | 2014 | 450000 | 145500 | Diesel | Individual | Manual | 1 | 23.40 | 1248.0 |
| 1 | Skoda | 2014 | 370000 | 120000 | Diesel | Individual | Manual | 2 | 21.14 | 1498.0 |
| 2 | Honda | 2006 | 158000 | 140000 | Petrol | Individual | Manual | 3 | 17.70 | 1497.0 |
| 3 | Hyundai | 2010 | 225000 | 127000 | Diesel | Individual | Manual | 1 | 23.00 | 1396.0 |
| 4 | Maruti | 2007 | 130000 | 120000 | Petrol | Individual | Manual | 1 | 16.10 | 1298.0 |



In [33]: `df.shape`

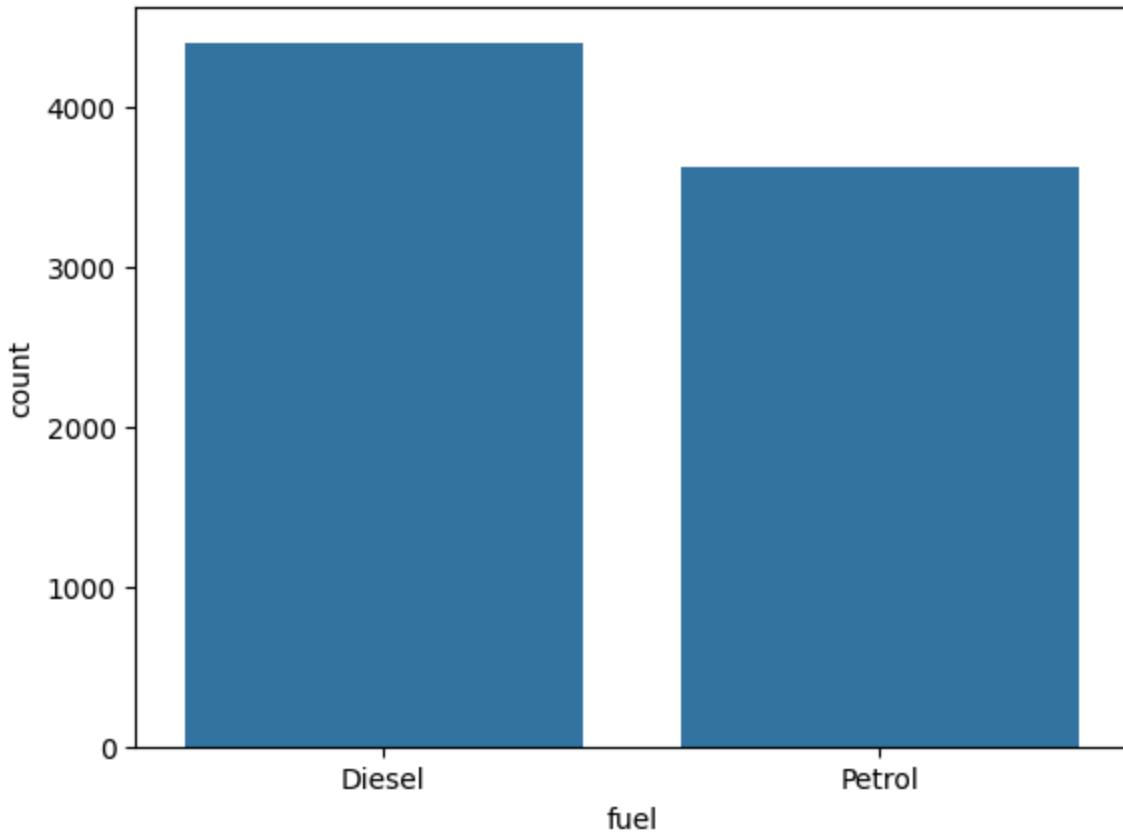
Out[33]: `(8028, 12)`

In [34]: `df.info()`

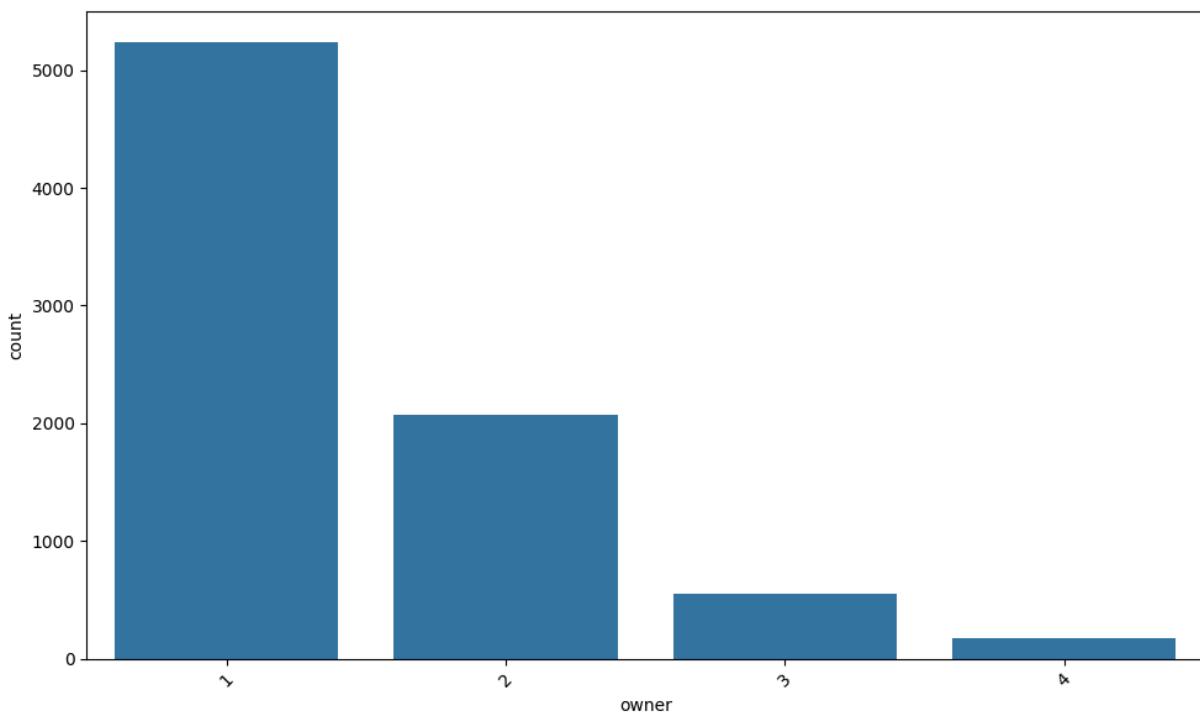
```
<class 'pandas.core.frame.DataFrame'>
Index: 8028 entries, 0 to 8127
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   name        8028 non-null   object 
 1   year         8028 non-null   int64  
 2   price        8028 non-null   int64  
 3   driven       8028 non-null   int64  
 4   fuel          8028 non-null   object 
 5   seller        8028 non-null   object 
 6   transmission 8028 non-null   object 
 7   owner         8028 non-null   int64  
 8   mileage       7814 non-null   float64
 9   engine        7814 non-null   float64
 10  power         7820 non-null   float64
 11  seats         7814 non-null   float64
dtypes: float64(4), int64(4), object(4)
memory usage: 815.3+ KB
```

```
In [35]: sns.countplot(data=df, x='fuel')
```

```
Out[35]: <Axes: xlabel='fuel', ylabel='count'>
```

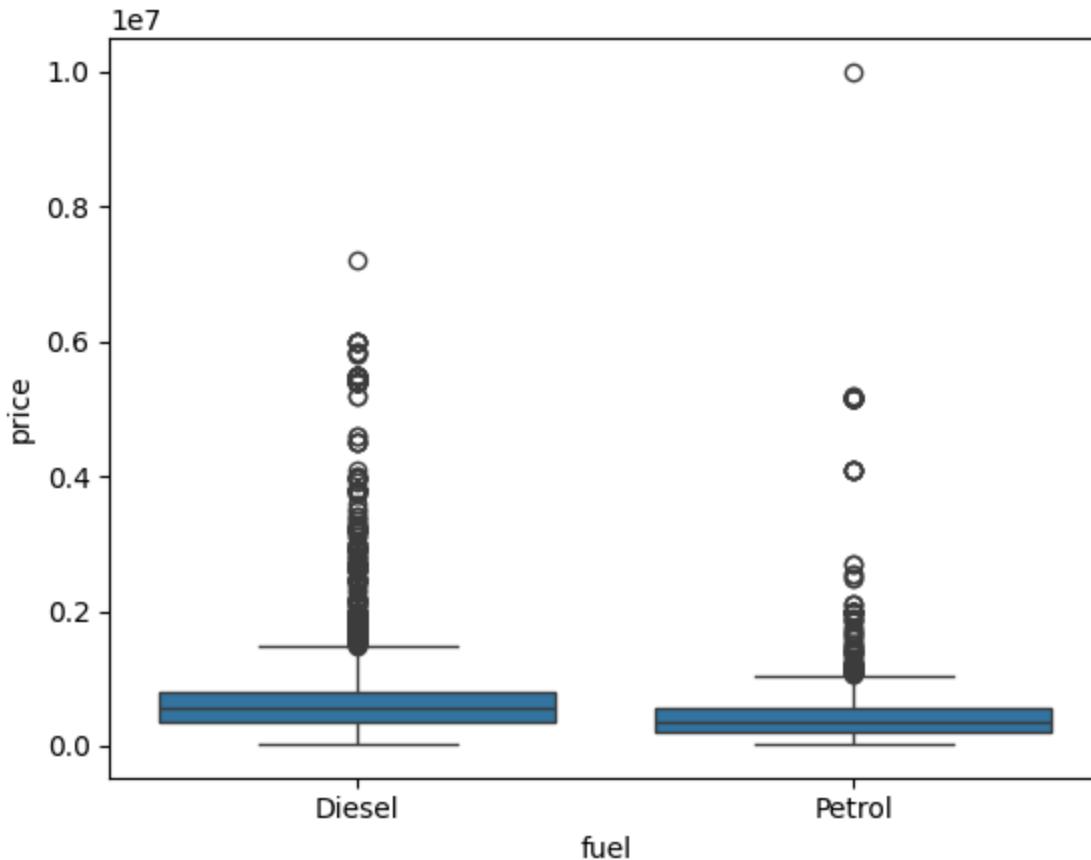


```
In [36]: plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='owner')
plt.xticks(rotation=45)
plt.tight_layout()
```



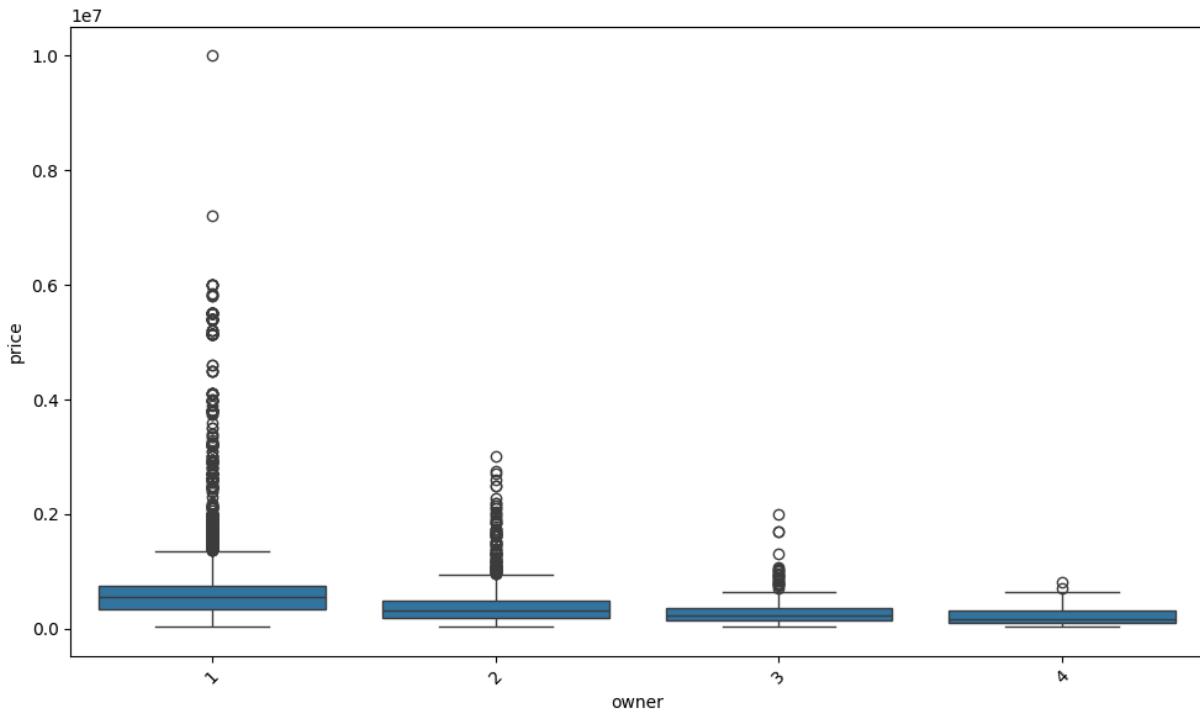
```
In [37]: sns.boxplot(data=df, x='fuel', y='price')
```

```
Out[37]: <Axes: xlabel='fuel', ylabel='price'>
```



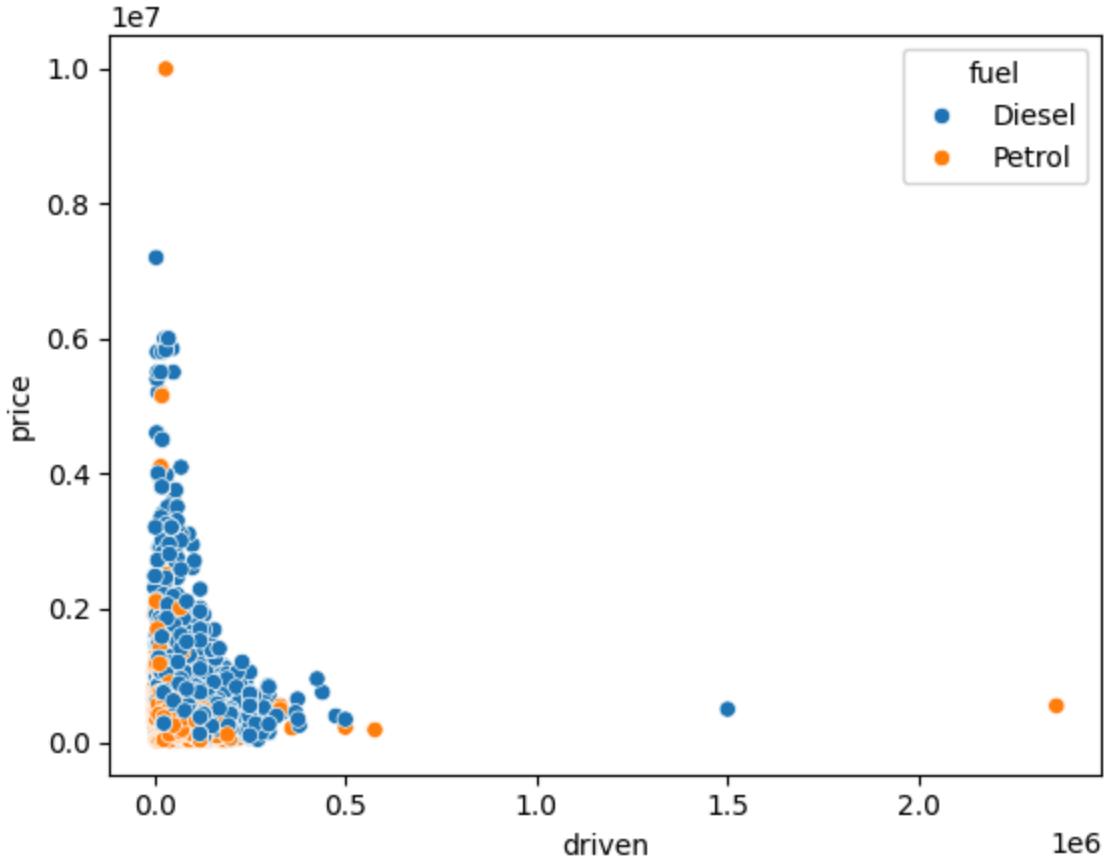
```
In [38]: plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='owner', y='price')
```

```
plt.xticks(rotation=45)  
plt.tight_layout()
```

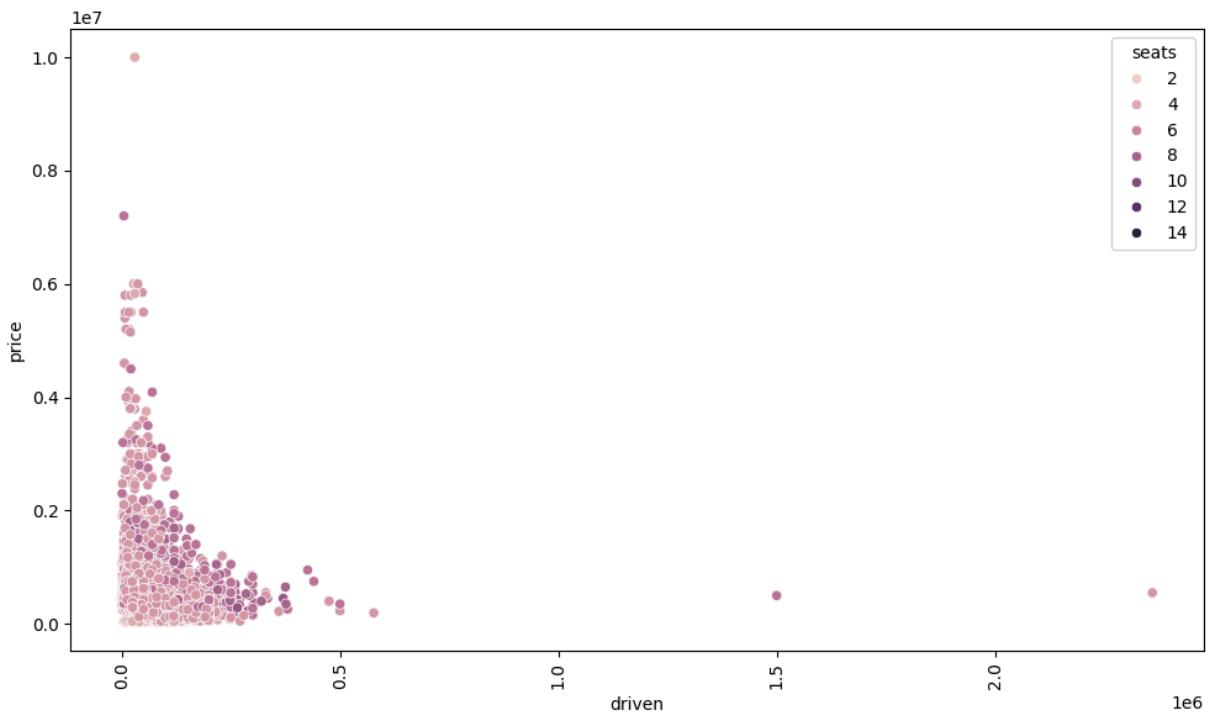


```
In [39]: sns.scatterplot(data=df, x='driven', y='price', hue='fuel')
```

```
Out[39]: <Axes: xlabel='driven', ylabel='price'>
```

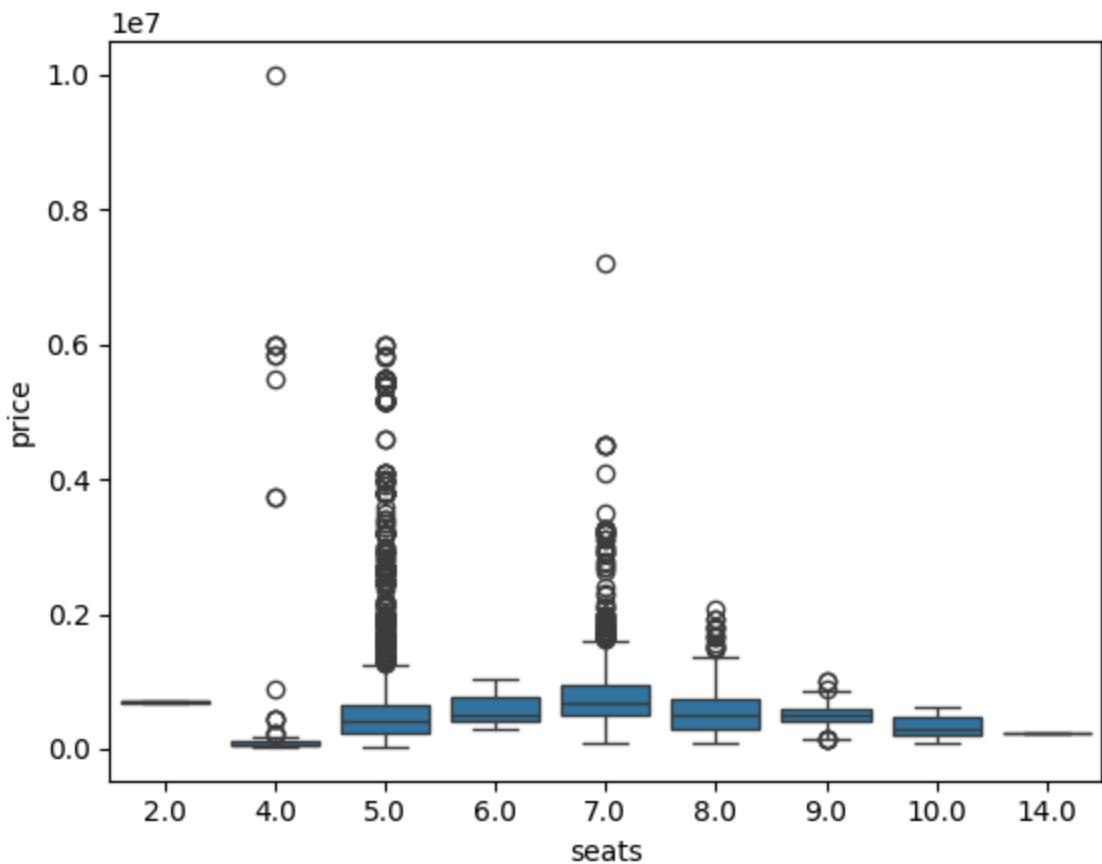


```
In [40]: plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='driven', y='price', hue='seats')
plt.xticks(rotation=90)
plt.tight_layout()
```



```
In [41]: sns.boxplot(data=df, x='seats', y='price')
```

```
Out[41]: <Axes: xlabel='seats', ylabel='price'>
```



In [42]: `df.head()`

Out[42]:

| | name | year | price | driven | fuel | seller | transmission | owner | mileage | engine |
|----------|-------------|-------------|--------------|---------------|-------------|---------------|---------------------|--------------|----------------|---------------|
| 0 | Maruti | 2014 | 450000 | 145500 | Diesel | Individual | Manual | 1 | 23.40 | 1248.C |
| 1 | Skoda | 2014 | 370000 | 120000 | Diesel | Individual | Manual | 2 | 21.14 | 1498.C |
| 2 | Honda | 2006 | 158000 | 140000 | Petrol | Individual | Manual | 3 | 17.70 | 1497.C |
| 3 | Hyundai | 2010 | 225000 | 127000 | Diesel | Individual | Manual | 1 | 23.00 | 1396.C |
| 4 | Maruti | 2007 | 130000 | 120000 | Petrol | Individual | Manual | 1 | 16.10 | 1298.C |

In [43]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
Index: 8028 entries, 0 to 8127
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   name        8028 non-null    object 
 1   year         8028 non-null    int64  
 2   price        8028 non-null    int64  
 3   driven       8028 non-null    int64  
 4   fuel          8028 non-null    object 
 5   seller        8028 non-null    object 
 6   transmission 8028 non-null    object 
 7   owner         8028 non-null    int64  
 8   mileage       7814 non-null    float64
 9   engine        7814 non-null    float64
 10  power         7820 non-null    float64
 11  seats         7814 non-null    float64
dtypes: float64(4), int64(4), object(4)
memory usage: 815.3+ KB
```

```
In [44]: from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df["name"] = le.fit_transform(df["name"])
df["fuel"] = le.fit_transform(df["fuel"])
df["seller"] = le.fit_transform(df["seller"])
df["transmission"] = le.fit_transform(df["transmission"])

df["name"].unique()
df["fuel"].unique()
df["seller"].unique()
df["transmission"].unique()
```

```
Out[44]: array([1, 0])
```

```
In [45]: df.head()
```

| | name | year | price | driven | fuel | seller | transmission | owner | mileage | engine | power | |
|---|------|------|--------|--------|------|--------|--------------|-------|---------|--------|--------|-------|
| 0 | 20 | 2014 | 450000 | 145500 | 0 | 1 | | 1 | 1 | 23.40 | 1248.0 | 74.0 |
| 1 | 27 | 2014 | 370000 | 120000 | 0 | 1 | | 1 | 2 | 21.14 | 1498.0 | 103.5 |
| 2 | 10 | 2006 | 158000 | 140000 | 1 | 1 | | 1 | 3 | 17.70 | 1497.0 | 78.0 |
| 3 | 11 | 2010 | 225000 | 127000 | 0 | 1 | | 1 | 1 | 23.00 | 1396.0 | 90.0 |
| 4 | 20 | 2007 | 130000 | 120000 | 1 | 1 | | 1 | 1 | 16.10 | 1298.0 | 88.2 |

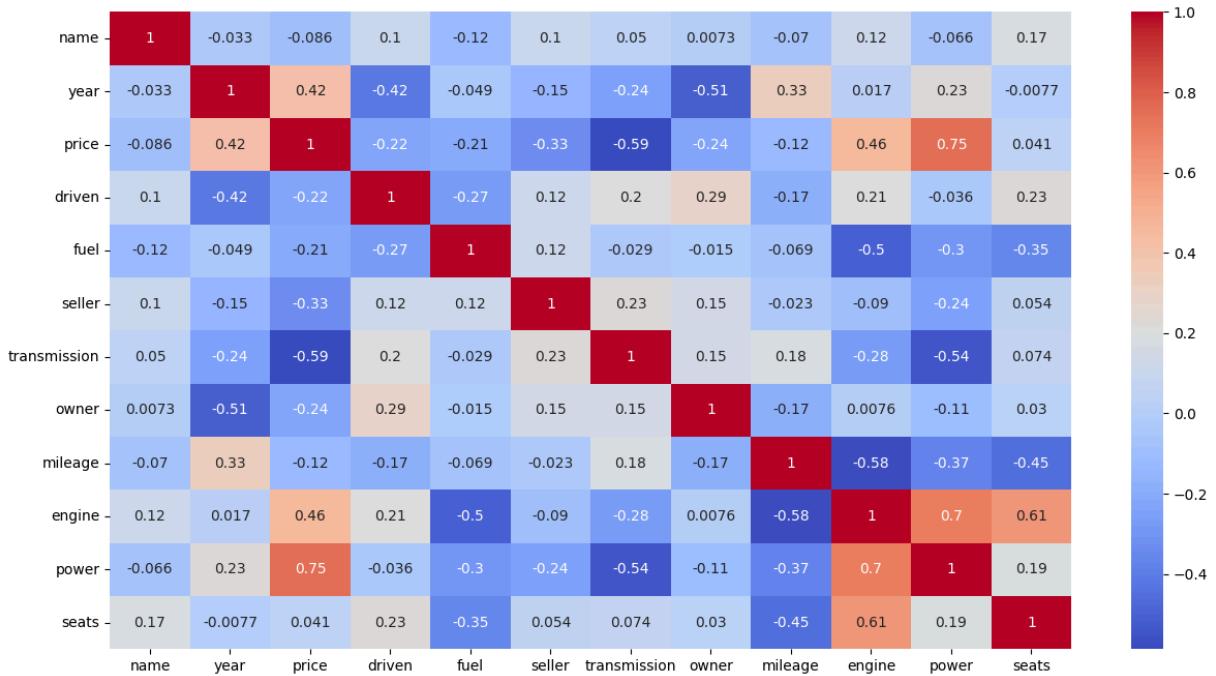


```
In [46]: le.classes_
```

```
Out[46]: array(['Automatic', 'Manual'], dtype=object)
```

```
In [47]: plt.figure(figsize = (15,8))
sns.heatmap(df.corr(), annot=True, cmap="coolwarm")
```

Out[47]: <Axes: >



In [48]: X=df[['engine','power','year']]

In [49]: y = np.log(df['price'])
print(y)

```

0      13.017003
1      12.821258
2      11.970350
3      12.323856
4      11.775290
       ...
8123    12.676076
8124    11.813030
8125    12.853176
8126    12.577636
8127    12.577636
Name: price, Length: 8028, dtype: float64

```

In [50]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_st

In [51]: X_train[['engine','power','year']].isna().sum()

Out[51]: engine 171
power 165
year 0
dtype: int64

In [52]: X_test[['engine','power','year']].isna().sum()

```
Out[52]: engine    43
          power     43
          year      0
          dtype: int64
```

```
In [53]: # ...existing code...
# Remove NaNs from X_train and y_train together
mask_train = ~np.isnan(X_train).any(axis=1)
X_train = X_train[mask_train]
y_train = y_train[mask_train]

# Remove NaNs from X_test and y_test together
mask_test = ~np.isnan(X_test).any(axis=1)
X_test = X_test[mask_test]
y_test = y_test[mask_test]
```

```
In [54]: X_train[['engine','power','year']].isna().sum()
```

```
Out[54]: engine    0
          power     0
          year      0
          dtype: int64
```

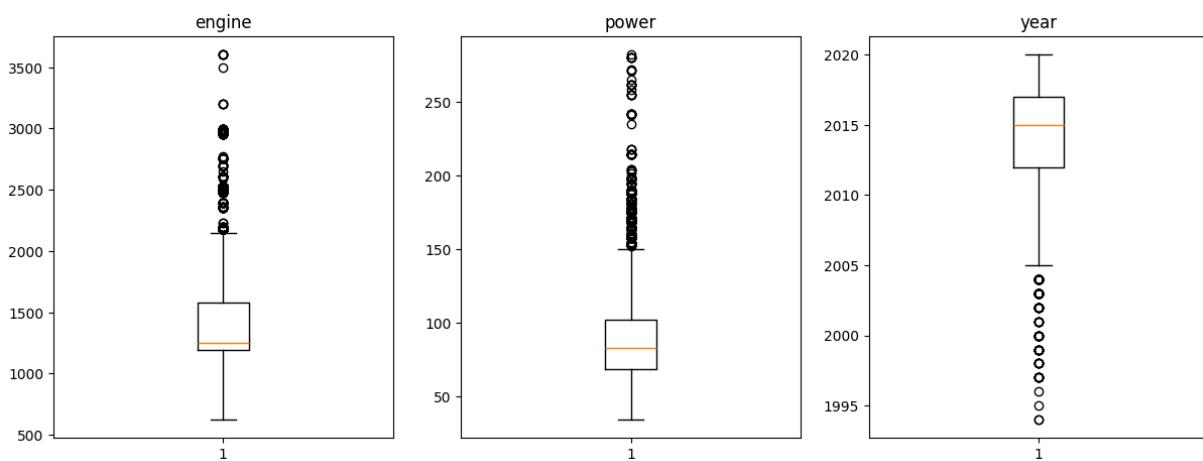
```
In [55]: X_test[['engine','power','year']].isna().sum()
```

```
Out[55]: engine    0
          power     0
          year      0
          dtype: int64
```

```
In [56]: y_train.isna().sum()
          y_test.isna().sum()
```

```
Out[56]: np.int64(0)
```

```
In [57]: col_dict = {'engine':1, 'power':2, 'year':3 }
plt.figure(figsize = (20,30))
for var, i in col_dict.items():
    plt.subplot(5,4,i)
    plt.boxplot(X_train[var])
    plt.title(var)
plt.show()
```



```
In [58]: def outlier_count(col, data = X_train):
    q75, q25 = np.percentile(data[col], [75, 25])
    iqr = q75 - q25
    min_val = q25 - (iqr*1.5)
    max_val = q75 + (iqr*1.5)
    outlier_count = len(np.where((data[col] > max_val) | (data[col] < min_val))[0])
    outlier_percent = round(outlier_count/len(data[col])*100, 2)

    if(outlier_count > 0):
        print("\n"+15*'- '+ col + 15*'-+'\n")
        print('Number of outliers: {}'.format(outlier_count))
        print('Percent of data that is outlier: {}%'.format(outlier_percent))
```

```
In [59]: for col in X_train.columns:
    outlier_count(col)
```

-----engine-----

Number of outliers: 960
Percent of data that is outlier: 15.36%

-----power-----

Number of outliers: 453
Percent of data that is outlier: 7.25%

-----year-----

Number of outliers: 135
Percent of data that is outlier: 2.16%

```
In [60]: from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

```
In [61]: print("Shape of X_train:", X_train.shape)
print("Shape of X_test:", X_test.shape)
print("Shape of y_train:", y_train.shape)
print("Shape of y_test:", y_test.shape)
```

```
Shape of X_train: (6251, 3)
Shape of X_test: (1563, 3)
Shape of y_train: (6251,)
Shape of y_train: (1563,)
```

```
In [62]: from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
lr=LinearRegression()
lr.fit(X_train,y_train)
yhat=lr.predict(X_test)
print("MSE: ", mean_squared_error(y_test,yhat))
print("R2 score: ", r2_score(y_test,yhat))
```

```
MSE:  0.09634624739397166
R2 score:  0.861596009065188
```

```
In [63]: from sklearn.svm import SVR
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
algorithm = [SVR(), KNeighborsRegressor(), DecisionTreeRegressor(random_state=0), R
algorithms_name = [ "SVR", "Kneighbors", "DecisionTreeRegressor", "RandomForestRegr
```

```
In [64]: from sklearn.model_selection import KFold, cross_val_score
train_mse = []
test_mse = []
kfold = KFold(n_splits=5, shuffle=True)
for i,model in enumerate(algorithm):
    scores = cross_val_score(model, X_train, y_train, cv=kfold, scoring='neg_mean_s
    print(f"{algorithms_name[i]} - Score: {scores}; Mean: {scores.mean()}")
```

```
SVR - Score: [-0.07708995 -0.07468777 -0.0782877 -0.08082459 -0.0776649 ]; Mean: -0.07771098264118467
Kneighbors - Score: [-0.06070371 -0.06136781 -0.05199902 -0.06084531 -0.06221756]; Mean: -0.05942668220334082
DecisionTreeRegressor - Score: [-0.05823046 -0.05790919 -0.06009509 -0.06304642 -0.05608061]; Mean: -0.05907235388071017
RandomForestRegressor - Score: [-0.04779032 -0.04569046 -0.05206744 -0.05408714 -0.04929919]; Mean: -0.049786911172275204
```

```
In [65]: from sklearn.model_selection import GridSearchCV
grid_params = {'bootstrap': [True], 'max_depth': [5, 10, None], 'n_estimators': [5, 10, 15, 20, 30, 50, 70, 100]}
rf=RandomForestRegressor(random_state=1)
grid = GridSearchCV(estimator=rf,
                    param_grid=grid_params,
                    cv=kfold,
                    n_jobs=1,
                    return_train_score=True,
                    refit=True,
                    scoring='r2')
grid.fit(X_train, y_train);
```

```
In [66]: grid.best_params_
```

```
Out[66]: {'bootstrap': True, 'max_depth': None, 'n_estimators': 15}
```

```
In [67]: best_mse=grid.best_score_
```

```
In [68]: best_mse
```

```
Out[68]: np.float64(0.9239973516526948)
```

```
In [69]: predict_y=grid.predict(X_test)
predict_y = np.exp(predict_y)
y_test = np.exp(y_test)
mean_squared_error(y_test,predict_y)
import numpy as np

mse = mean_squared_error(y_test, predict_y)
rmse = np.sqrt(mse)
print("Root Mean Squared Error:", rmse)

r2 = r2_score(y_test, predict_y)
print("R2 Score:", r2)
```

```
Root Mean Squared Error: 217043.32354388785
```

```
R2 Score: 0.9402985255700533
```

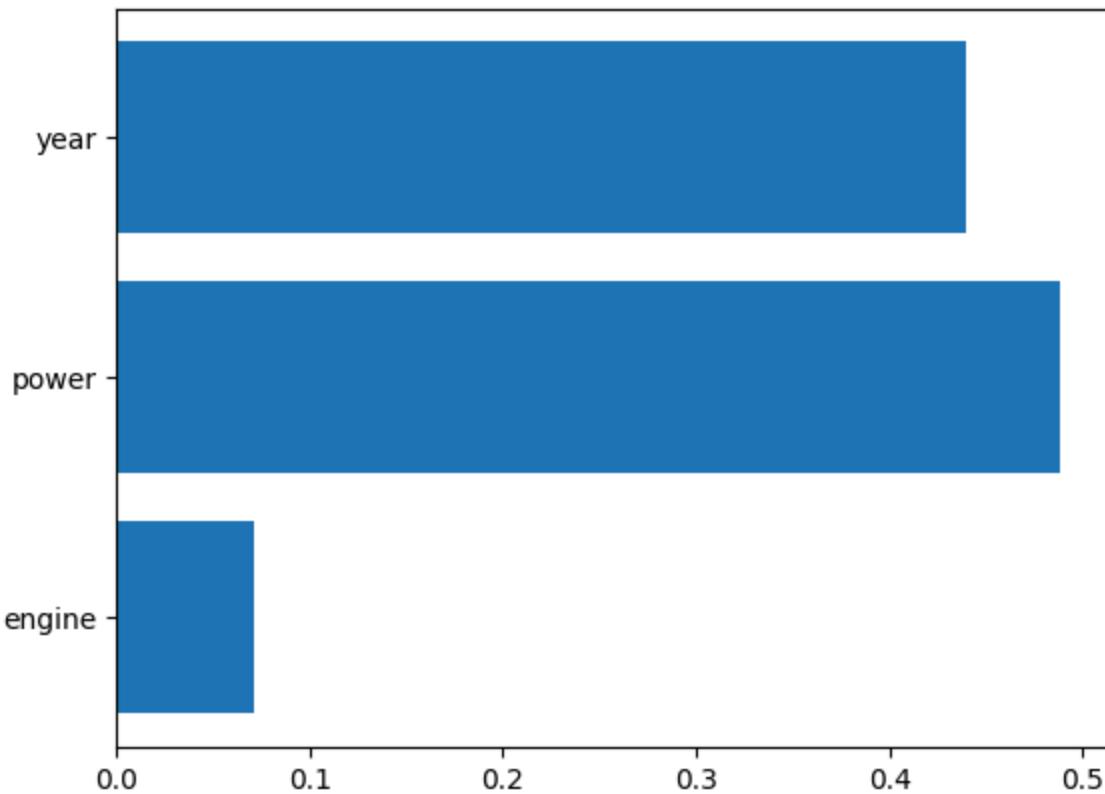
```
In [70]: rf = grid.best_estimator_
```

```
rf.feature_importances_
```

```
Out[70]: array([0.07145498, 0.48873843, 0.43980659])
```

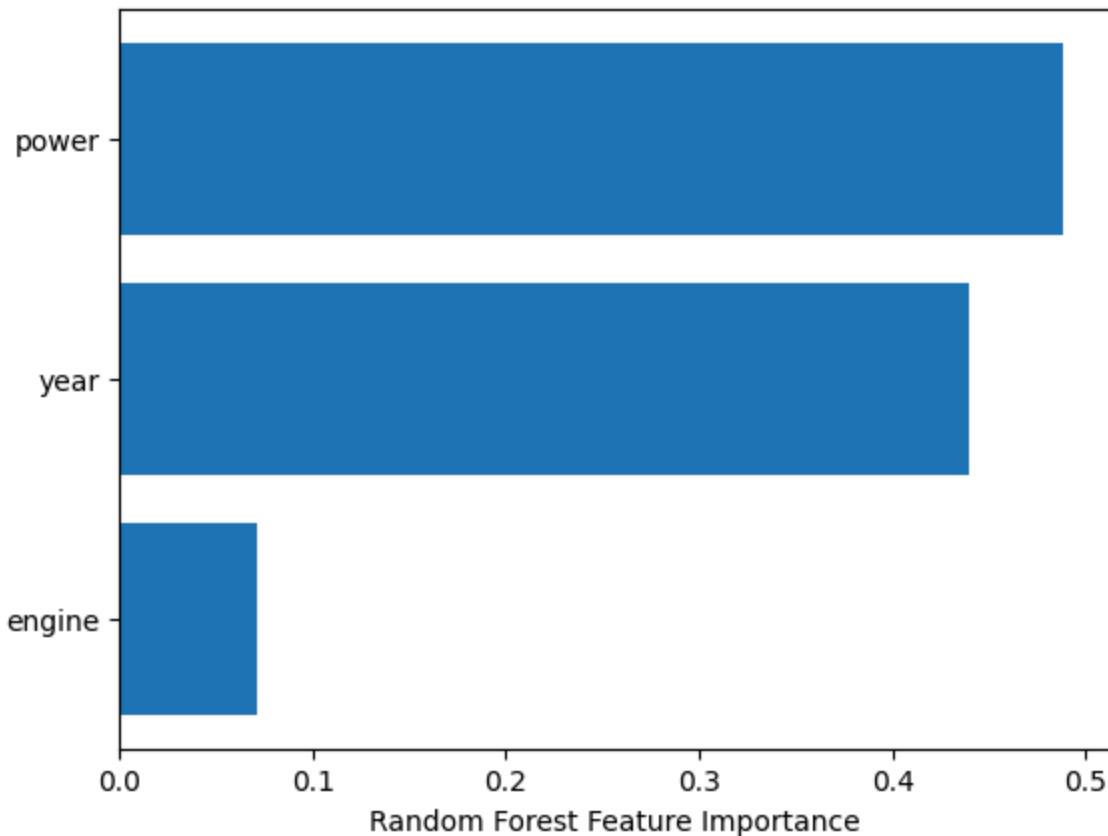
```
In [71]: plt.barh(X.columns, rf.feature_importances_)
```

```
Out[71]: <BarContainer object of 3 artists>
```



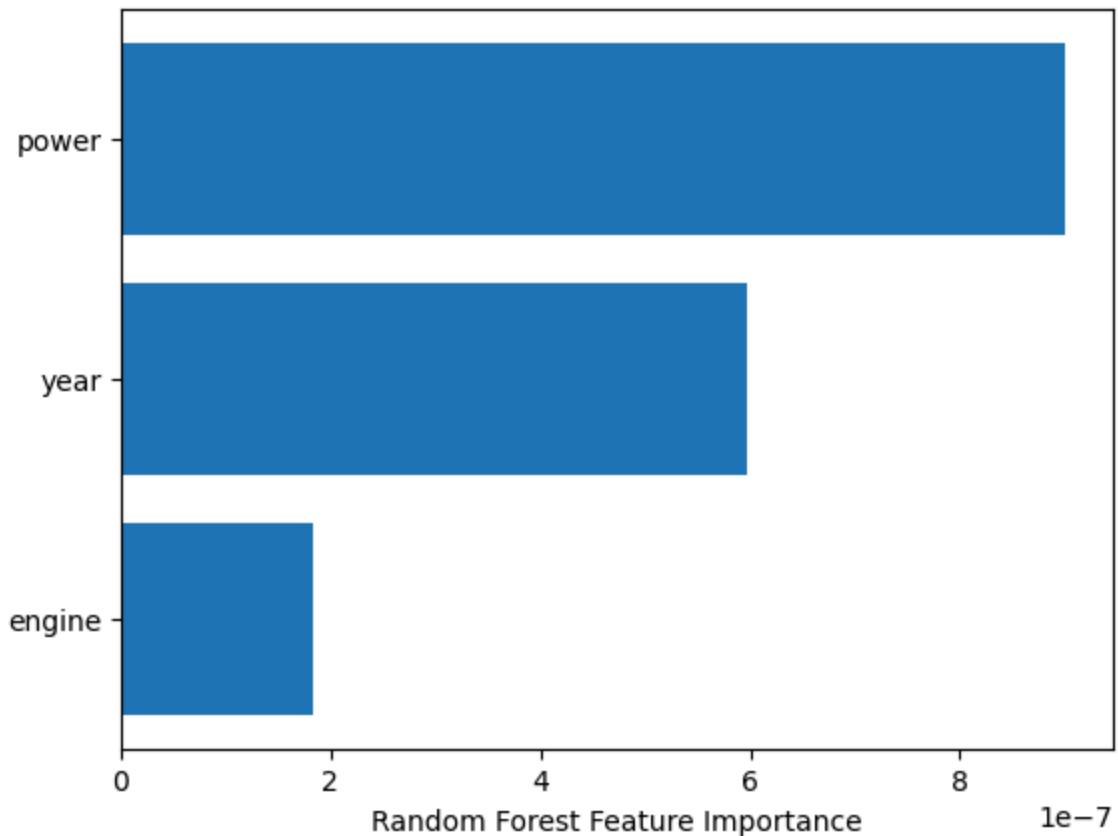
```
In [72]: sorted_idx = rf.feature_importances_.argsort()
plt.barh(X.columns[sorted_idx], rf.feature_importances_[sorted_idx])
plt.xlabel("Random Forest Feature Importance")
```

```
Out[72]: Text(0.5, 0, 'Random Forest Feature Importance')
```



```
In [73]: from sklearn.inspection import permutation_importance  
  
perm_importance = permutation_importance(rf, X_test, y_test)  
sorted_idx = perm_importance.importances_mean.argsort()  
plt.barh(X.columns[sorted_idx], perm_importance.importances_mean[sorted_idx])  
plt.xlabel("Random Forest Feature Importance")
```

```
Out[73]: Text(0.5, 0, 'Random Forest Feature Importance')
```



```
In [74]: pip install shap
```

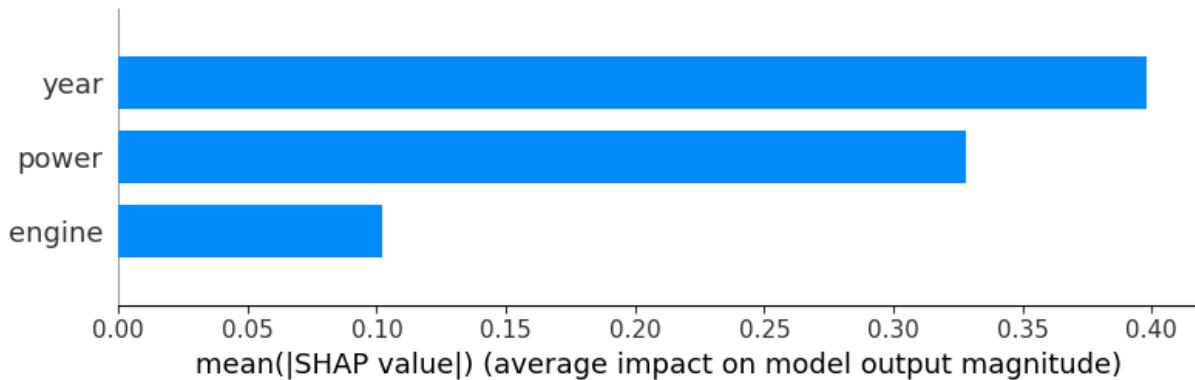
```
Requirement already satisfied: shap in d:\ml\.venv\lib\site-packages (0.48.0)
Requirement already satisfied: numpy in d:\ml\.venv\lib\site-packages (from shap)
(2.2.0)
Requirement already satisfied: scipy in d:\ml\.venv\lib\site-packages (from shap)
(1.16.1)
Requirement already satisfied: scikit-learn in d:\ml\.venv\lib\site-packages (from s
hap) (1.7.1)
Requirement already satisfied: pandas in d:\ml\.venv\lib\site-packages (from shap)
(2.3.2)
Requirement already satisfied: tqdm>=4.27.0 in d:\ml\.venv\lib\site-packages (from s
hap) (4.67.1)
Requirement already satisfied: packaging>20.9 in d:\ml\.venv\lib\site-packages (from
shap) (25.0)
Requirement already satisfied: slicer==0.0.8 in d:\ml\.venv\lib\site-packages (from
shap) (0.0.8)
Requirement already satisfied: numba>=0.54 in d:\ml\.venv\lib\site-packages (from sh
ap) (0.61.2)
Requirement already satisfied:云pickle in d:\ml\.venv\lib\site-packages (from sh
ap) (3.1.1)
Requirement already satisfied: typing-extensions in d:\ml\.venv\lib\site-packages (f
rom shap) (4.15.0)
Requirement already satisfied: llvmlite<0.45,>=0.44.0dev0 in d:\ml\.venv\lib\site-pa
ckages (from numba>=0.54->shap) (0.44.0)
Requirement already satisfied: colorama in d:\ml\.venv\lib\site-packages (from tqdm>
=4.27.0->shap) (0.4.6)
Requirement already satisfied: python-dateutil>=2.8.2 in d:\ml\.venv\lib\site-packag
es (from pandas->shap) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in d:\ml\.venv\lib\site-packages (from p
andas->shap) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in d:\ml\.venv\lib\site-packages (from
pandas->shap) (2025.2)
Requirement already satisfied: six>=1.5 in d:\ml\.venv\lib\site-packages (from pytho
n-dateutil>=2.8.2->pandas->shap) (1.17.0)
Requirement already satisfied: joblib>=1.2.0 in d:\ml\.venv\lib\site-packages (from
scikit-learn->shap) (1.5.1)
Requirement already satisfied: threadpoolctl>=3.1.0 in d:\ml\.venv\lib\site-packages
(from scikit-learn->shap) (3.6.0)
Note: you may need to restart the kernel to use updated packages.
```

In [75]: `import shap`

```
explainer = shap.TreeExplainer(rf)
shap_values = explainer.shap_values(X_test)
```

```
d:\ML\.venv\Lib\site-packages\tqdm\auto.py:21: TqdmWarning: IPProgress not found. Ple
ase update jupyter and ipywidgets. See https://ipywidgets.readthedocs.io/en/stable/u
ser_install.html
from .autonotebook import tqdm as notebook_tqdm
```

In [76]: `shap.summary_plot(shap_values, X_test, plot_type="bar", feature_names = X.columns)`



In [77]: `pip install pickle`

Note: you may need to restart the kernel to use updated packages.

ERROR: Could not find a version that satisfies the requirement pickle (from versions: none)

ERROR: No matching distribution found for pickle

In [78]: `import pickle`

```
# save the model to disk
filename = 'CarPriceModel/car_price.model'
pickle.dump(grid, open('car_price_file', 'wb'))
```

In [79]: `loaded_model = pickle.load(open('car_price_file', 'rb'))`

In [80]: `df[['engine', 'power', 'year']].loc[1]`

Out[80]:

| Feature | Value |
|---------|-------------------|
| engine | 1498.00 |
| power | 103.52 |
| year | 2014.00 |
| Name: | 1, dtype: float64 |

In [81]: `sample = np.array([[0.476, 271.000, 2015]])`

In [82]: `predicted_car_price = loaded_model.predict(sample)`
`predicted_car_price`

Out[82]: `array([15.13557638])`

Pipe Line

In []:

```
import json, pickle, numpy as np
from pathlib import Path

best = grid.best_estimator_

cartifacts = Path("cartifacts")
cartifacts.mkdir(exist_ok=True)
```

```
with open(cartifacts / "model.pkl", "wb") as f:
    pickle.dump(best, f)

numeric_cols = ['engine', 'power', 'year']
categorical_cols = []
schema = {
    "target": "price",
    "target_transform": "log",
    "feature_order": list(X.columns),
    "numeric_cols": numeric_cols,
    "categorical_cols": categorical_cols
}
with open(cartifacts / "schema.json", "w") as f:
    json.dump(schema, f, indent=2)

print("Saved cartifacts/model.pkl and cartifacts/schema.json")
```

Saved cartifacts/model.pkl and cartifacts/schema.json

In [1]: pip install nbconvert jupyter

```
Collecting nbconvert
  Downloading nbconvert-7.16.6-py3-none-any.whl.metadata (8.5 kB)
Collecting jupyter
  Downloading jupyter-1.1.1-py2.py3-none-any.whl.metadata (2.0 kB)
Collecting beautifulsoup4 (from nbconvert)
  Downloading beautifulsoup4-4.13.5-py3-none-any.whl.metadata (3.8 kB)
Collecting bleach!=5.0.0 (from bleach[css]!=5.0.0->nbconvert)
  Downloading bleach-6.2.0-py3-none-any.whl.metadata (30 kB)
Collecting defusedxml (from nbconvert)
  Downloading defusedxml-0.7.1-py2.py3-none-any.whl.metadata (32 kB)
Requirement already satisfied: jinja2>=3.0 in d:\ml\venv\lib\site-packages (from nbconvert) (3.1.6)
Requirement already satisfied: jupyter-core>=4.7 in d:\ml\venv\lib\site-packages (from nbconvert) (5.8.1)
Collecting jupyterlab-pygments (from nbconvert)
  Downloading jupyterlab_pygments-0.3.0-py3-none-any.whl.metadata (4.4 kB)
Requirement already satisfied: markupsafe>=2.0 in d:\ml\venv\lib\site-packages (from nbconvert) (3.0.2)
Collecting mistune<4,>=2.0.3 (from nbconvert)
  Downloading mistune-3.1.3-py3-none-any.whl.metadata (1.8 kB)
Collecting nbclient>=0.5.0 (from nbconvert)
  Downloading nbclient-0.10.2-py3-none-any.whl.metadata (8.3 kB)
Collecting nbformat>=5.7 (from nbconvert)
  Downloading nbformat-5.10.4-py3-none-any.whl.metadata (3.6 kB)
Requirement already satisfied: packaging in d:\ml\venv\lib\site-packages (from nbconvert) (25.0)
Collecting pandocfilters>=1.4.1 (from nbconvert)
  Downloading pandocfilters-1.5.1-py2.py3-none-any.whl.metadata (9.0 kB)
Requirement already satisfied: pygments>=2.4.1 in d:\ml\venv\lib\site-packages (from nbconvert) (2.19.2)
Requirement already satisfied: traitlets>=5.1 in d:\ml\venv\lib\site-packages (from nbconvert) (5.14.3)
Collecting notebook (from jupyter)
  Using cached notebook-7.4.5-py3-none-any.whl.metadata (10 kB)
Collecting jupyter-console (from jupyter)
  Downloading jupyter_console-6.6.3-py3-none-any.whl.metadata (5.8 kB)
Requirement already satisfied: ipykernel in d:\ml\venv\lib\site-packages (from jupyter) (6.30.1)
Collecting ipywidgets (from jupyter)
  Downloading ipywidgets-8.1.7-py3-none-any.whl.metadata (2.4 kB)
Collecting jupyterlab (from jupyter)
  Downloading jupyterlab-4.4.6-py3-none-any.whl.metadata (16 kB)
Collecting webencodings (from bleach!=5.0.0->bleach[css]!=5.0.0->nbconvert)
  Downloading webencodings-0.5.1-py2.py3-none-any.whl.metadata (2.1 kB)
Collecting tinyccs2<1.5,>=1.1.0 (from bleach[css]!=5.0.0->nbconvert)
  Downloading tinyccs2-1.4.0-py3-none-any.whl.metadata (3.0 kB)
Requirement already satisfied: platformdirs>=2.5 in d:\ml\venv\lib\site-packages (from jupyter-core>=4.7->nbconvert) (4.3.8)
Requirement already satisfied: pywin32>=300 in d:\ml\venv\lib\site-packages (from jupyter-core>=4.7->nbconvert) (311)
Requirement already satisfied: jupyter-client>=6.1.12 in d:\ml\venv\lib\site-packages (from nbclient>=0.5.0->nbconvert) (8.6.3)
Requirement already satisfied: python-dateutil>=2.8.2 in d:\ml\venv\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (2.9.0.post0)
Requirement already satisfied: pyzmq>=23.0 in d:\ml\venv\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (27.0.1)
```

```
Requirement already satisfied: tornado>=6.2 in d:\ml\.venv\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (6.5.2)
Collecting fastjsonschema>=2.15 (from nbformat>=5.7->nbconvert)
    Downloading fastjsonschema-2.21.2-py3-none-any.whl.metadata (2.3 kB)
Collecting jsonschema>=2.6 (from nbformat>=5.7->nbconvert)
    Downloading jsonschema-4.25.1-py3-none-any.whl.metadata (7.6 kB)
Collecting attrs>=22.2.0 (from jsonschema>=2.6->nbformat>=5.7->nbconvert)
    Using cached attrs-25.3.0-py3-none-any.whl.metadata (10 kB)
Collecting jsonschema-specifications>=2023.03.6 (from jsonschema>=2.6->nbformat>=5.7->nbconvert)
    Downloading jsonschema_specifications-2025.4.1-py3-none-any.whl.metadata (2.9 kB)
Collecting referencing>=0.28.4 (from jsonschema>=2.6->nbformat>=5.7->nbconvert)
    Downloading referencing-0.36.2-py3-none-any.whl.metadata (2.8 kB)
Collecting rpds-py>=0.7.1 (from jsonschema>=2.6->nbformat>=5.7->nbconvert)
    Downloading rpds_py-0.27.1-cp313-cp313-win_amd64.whl.metadata (4.3 kB)
Requirement already satisfied: six>=1.5 in d:\ml\.venv\lib\site-packages (from python-dateutil>=2.8.2->jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (1.17.0)
Collecting soupsieve>1.2 (from beautifulsoup4->nbconvert)
    Downloading soupsieve-2.8-py3-none-any.whl.metadata (4.6 kB)
Requirement already satisfied: typing-extensions>=4.0.0 in d:\ml\.venv\lib\site-packages (from beautifulsoup4->nbconvert) (4.15.0)
Requirement already satisfied: comm>=0.1.1 in d:\ml\.venv\lib\site-packages (from ipykernel->jupyter) (0.2.3)
Requirement already satisfied: debugpy>=1.6.5 in d:\ml\.venv\lib\site-packages (from ipykernel->jupyter) (1.8.16)
Requirement already satisfied: ipython>=7.23.1 in d:\ml\.venv\lib\site-packages (from ipykernel->jupyter) (9.4.0)
Requirement already satisfied: matplotlib-inline>=0.1 in d:\ml\.venv\lib\site-packages (from ipykernel->jupyter) (0.1.7)
Requirement already satisfied: nest-asyncio>=1.4 in d:\ml\.venv\lib\site-packages (from ipykernel->jupyter) (1.6.0)
Requirement already satisfied: psutil>=5.7 in d:\ml\.venv\lib\site-packages (from ipykernel->jupyter) (7.0.0)
Requirement already satisfied: colorama in d:\ml\.venv\lib\site-packages (from ipython>=7.23.1->ipykernel->jupyter) (0.4.6)
Requirement already satisfied: decorator in d:\ml\.venv\lib\site-packages (from ipython>=7.23.1->ipykernel->jupyter) (5.2.1)
Requirement already satisfied: ipython-pygments-lexers in d:\ml\.venv\lib\site-packages (from ipython>=7.23.1->ipykernel->jupyter) (1.1.1)
Requirement already satisfied: jedi>=0.16 in d:\ml\.venv\lib\site-packages (from ipython>=7.23.1->ipykernel->jupyter) (0.19.2)
Requirement already satisfied: prompt_toolkit<3.1.0,>=3.0.41 in d:\ml\.venv\lib\site-packages (from ipython>=7.23.1->ipykernel->jupyter) (3.0.51)
Requirement already satisfied: stack_data in d:\ml\.venv\lib\site-packages (from ipython>=7.23.1->ipykernel->jupyter) (0.6.3)
Requirement already satisfied: wcwidth in d:\ml\.venv\lib\site-packages (from prompt_toolkit<3.1.0,>=3.0.41->ipython>=7.23.1->ipykernel->jupyter) (0.2.13)
Requirement already satisfied: parso<0.9.0,>=0.8.4 in d:\ml\.venv\lib\site-packages (from jedi>=0.16->ipython>=7.23.1->ipykernel->jupyter) (0.8.4)
Collecting widgetsnbextension~=4.0.14 (from ipywidgets->jupyter)
    Downloading widgetsnbextension-4.0.14-py3-none-any.whl.metadata (1.6 kB)
Collecting jupyterlab_widgets~=3.0.15 (from ipywidgets->jupyter)
    Downloading jupyterlab_widgets-3.0.15-py3-none-any.whl.metadata (20 kB)
Collecting async-lru>=1.0.0 (from jupyterlab->jupyter)
    Downloading async_lru-2.0.5-py3-none-any.whl.metadata (4.5 kB)
Collecting httpx<1,>=0.25.0 (from jupyterlab->jupyter)
```

```
  Downloading httpx-0.28.1-py3-none-any.whl.metadata (7.1 kB)
Collecting jupyter-lsp>=2.0.0 (from jupyterlab->jupyter)
    Downloading jupyter_lsp-2.3.0-py3-none-any.whl.metadata (1.8 kB)
Collecting jupyter-server<3,>=2.4.0 (from jupyterlab->jupyter)
    Downloading jupyter_server-2.17.0-py3-none-any.whl.metadata (8.5 kB)
Collecting jupyterlab-server<3,>=2.27.1 (from jupyterlab->jupyter)
    Downloading jupyterlab_server-2.27.3-py3-none-any.whl.metadata (5.9 kB)
Collecting notebook-shim>=0.2 (from jupyterlab->jupyter)
    Downloading notebook_shim-0.2.4-py3-none-any.whl.metadata (4.0 kB)
Requirement already satisfied: setuptools>=41.1.0 in d:\ml\venv\lib\site-packages
(from jupyterlab->jupyter) (80.9.0)
Collecting anyio (from httpx<1,>=0.25.0->jupyterlab->jupyter)
    Downloading anyio-4.10.0-py3-none-any.whl.metadata (4.0 kB)
Requirement already satisfied: certifi in d:\ml\venv\lib\site-packages (from httpx<
1,>=0.25.0->jupyterlab->jupyter) (2025.8.3)
Collecting httpcore==1.* (from httpx<1,>=0.25.0->jupyterlab->jupyter)
    Downloading httpcore-1.0.9-py3-none-any.whl.metadata (21 kB)
Requirement already satisfied: idna in d:\ml\venv\lib\site-packages (from httpx<1,>
=0.25.0->jupyterlab->jupyter) (3.10)
Collecting h11>=0.16 (from httpcore==1.*->httpx<1,>=0.25.0->jupyterlab->jupyter)
    Downloading h11-0.16.0-py3-none-any.whl.metadata (8.3 kB)
Collecting argon2-cffi>=21.1 (from jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
    Downloading argon2_cffi-25.1.0-py3-none-any.whl.metadata (4.1 kB)
Collecting jupyter-events>=0.11.0 (from jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
    Downloading jupyter_events-0.12.0-py3-none-any.whl.metadata (5.8 kB)
Collecting jupyter-server-terminals>=0.4.4 (from jupyter-server<3,>=2.4.0->jupyterla
b->jupyter)
    Downloading jupyter_server_terminals-0.5.3-py3-none-any.whl.metadata (5.6 kB)
Collecting prometheus-client>=0.9 (from jupyter-server<3,>=2.4.0->jupyterlab->jupyter
)
    Downloading prometheus_client-0.22.1-py3-none-any.whl.metadata (1.9 kB)
Collecting pywinpty>=2.0.1 (from jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
    Downloading pywinpty-3.0.0-cp313-cp313-win_amd64.whl.metadata (101 bytes)
Collecting send2trash>=1.8.2 (from jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
    Downloading Send2Trash-1.8.3-py3-none-any.whl.metadata (4.0 kB)
Collecting terminado>=0.8.3 (from jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
    Downloading terminado-0.18.1-py3-none-any.whl.metadata (5.8 kB)
Collecting websocket-client>=1.7 (from jupyter-server<3,>=2.4.0->jupyterlab->jupyter
)
    Downloading websocket_client-1.8.0-py3-none-any.whl.metadata (8.0 kB)
Collecting babel>=2.10 (from jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter)
    Downloading babel-2.17.0-py3-none-any.whl.metadata (2.0 kB)
Collecting json5>=0.9.0 (from jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter)
    Downloading json5-0.12.1-py3-none-any.whl.metadata (36 kB)
Requirement already satisfied: requests>=2.31 in d:\ml\venv\lib\site-packages (from
jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (2.32.5)
Collecting sniffio>=1.1 (from anyio->httpx<1,>=0.25.0->jupyterlab->jupyter)
    Downloading sniffio-1.3.1-py3-none-any.whl.metadata (3.9 kB)
Collecting argon2-cffi-bindings (from argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->j
upyterlab->jupyter)
    Downloading argon2_cffi_bindings-25.1.0-cp39-abi3-win_amd64.whl.metadata (7.5 kB)
Collecting python-json-logger>=2.0.4 (from jupyter-events>=0.11.0->jupyter-server<3,>
=2.4.0->jupyterlab->jupyter)
    Downloading python_json_logger-3.3.0-py3-none-any.whl.metadata (4.0 kB)
Collecting pyyaml>=5.3 (from jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupy
```

```
erlab->jupyter)
  Using cached PyYAML-6.0.2-cp313-cp313-win_amd64.whl.metadata (2.1 kB)
Collecting rfc3339-validator (from jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading rfc3339_validator-0.1.4-py2.py3-none-any.whl.metadata (1.5 kB)
Collecting rfc3986-validator>=0.1.1 (from jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading rfc3986_validator-0.1.1-py2.py3-none-any.whl.metadata (1.7 kB)
Collecting fqdn (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading fqdn-1.5.1-py3-none-any.whl.metadata (1.4 kB)
Collecting isoduration (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading isoduration-20.11.0-py3-none-any.whl.metadata (5.7 kB)
Collecting jsonpointer>1.13 (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading jsonpointer-3.0.0-py2.py3-none-any.whl.metadata (2.3 kB)
Collecting rfc3987-syntax>=1.1.0 (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading rfc3987_syntax-1.1.0-py3-none-any.whl.metadata (7.7 kB)
Collecting uri-template (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading uri_template-1.3.0-py3-none-any.whl.metadata (8.8 kB)
Collecting webcolors>=24.6.0 (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading webcolors-24.11.1-py3-none-any.whl.metadata (2.2 kB)
Requirement already satisfied: charset_normalizer<4,>=2 in d:\ml\.venv\lib\site-packages (from requests>=2.31->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (3.4.3)
Requirement already satisfied: urllib3<3,>=1.21.1 in d:\ml\.venv\lib\site-packages (from requests>=2.31->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (2.5.0)
Collecting lark>=1.2.2 (from rfc3987-syntax>=1.1.0->jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading lark-1.2.2-py3-none-any.whl.metadata (1.8 kB)
Collecting cffi>=1.0.1 (from argon2-cffi-bindings->argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading cffi-1.17.1-cp313-cp313-win_amd64.whl.metadata (1.6 kB)
Collecting pycparser (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading pycparser-2.22-py3-none-any.whl.metadata (943 bytes)
Collecting arrow>=0.15.0 (from isoduration->jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading arrow-1.3.0-py3-none-any.whl.metadata (7.5 kB)
Collecting types-python-dateutil>=2.8.10 (from arrow>=0.15.0->isoduration->jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
  Downloading types_python_dateutil-2.9.0.20250822-py3-none-any.whl.metadata (1.8 kB)
Requirement already satisfied: executing>=1.2.0 in d:\ml\.venv\lib\site-packages (from stack_data->ipython>=7.23.1->ipykernel->jupyter) (2.2.0)
Requirement already satisfied: asttokens>=2.1.0 in d:\ml\.venv\lib\site-packages (from stack_data->ipython>=7.23.1->ipykernel->jupyter) (3.0.0)
Requirement already satisfied: pure-eval in d:\ml\.venv\lib\site-packages (from stack_data->ipython>=7.23.1->ipykernel->jupyter) (0.2.3)
Downloading nbconvert-7.16.6-py3-none-any.whl (258 kB)
Downloading mistune-3.1.3-py3-none-any.whl (53 kB)
```

```
Downloading jupyter-1.1.1-py2.py3-none-any.whl (2.7 kB)
Downloading bleach-6.2.0-py3-none-any.whl (163 kB)
Downloading tiny.css-2.1.4.0-py3-none-any.whl (26 kB)
Downloading nbclient-0.10.2-py3-none-any.whl (25 kB)
Downloading nbformat-5.10.4-py3-none-any.whl (78 kB)
Downloading fastjsonschema-2.21.2-py3-none-any.whl (24 kB)
Downloading jsonschema-4.25.1-py3-none-any.whl (90 kB)
Using cached attrs-25.3.0-py3-none-any.whl (63 kB)
Downloading jsonschema_specifications-2025.4.1-py3-none-any.whl (18 kB)
Downloading pandocfilters-1.5.1-py2.py3-none-any.whl (8.7 kB)
Downloading referencing-0.36.2-py3-none-any.whl (26 kB)
Downloading rpds_py-0.27.1-cp313-cp313-win_amd64.whl (232 kB)
Downloading webencodings-0.5.1-py2.py3-none-any.whl (11 kB)
Downloading beautifulsoup4-4.13.5-py3-none-any.whl (105 kB)
Downloading soupsieve-2.8-py3-none-any.whl (36 kB)
Downloading defusedxml-0.7.1-py2.py3-none-any.whl (25 kB)
Downloading ipywidgets-8.1.7-py3-none-any.whl (139 kB)
Downloading jupyterlab_widgets-3.0.15-py3-none-any.whl (216 kB)
Downloading widgetsnbextension-4.0.14-py3-none-any.whl (2.2 MB)
----- 0.0/2.2 MB ? eta -:-:-
----- 2.2/2.2 MB 41.1 MB/s 0:00:00
Downloading jupyter_console-6.6.3-py3-none-any.whl (24 kB)
Downloading jupyterlab-4.4.6-py3-none-any.whl (12.3 MB)
----- 0.0/12.3 MB ? eta -:-:-
----- 6.3/12.3 MB 33.6 MB/s eta 0:00:01
----- 7.9/12.3 MB 19.4 MB/s eta 0:00:01
----- 9.2/12.3 MB 15.8 MB/s eta 0:00:01
----- 10.7/12.3 MB 13.2 MB/s eta 0:00:01
----- 12.1/12.3 MB 12.1 MB/s eta 0:00:01
----- 12.3/12.3 MB 11.6 MB/s 0:00:01
Downloading httpx-0.28.1-py3-none-any.whl (73 kB)
Downloading httpcore-1.0.9-py3-none-any.whl (78 kB)
Downloading jupyter_server-2.17.0-py3-none-any.whl (388 kB)
Downloading jupyterlab_server-2.27.3-py3-none-any.whl (59 kB)
Downloading anyio-4.10.0-py3-none-any.whl (107 kB)
Downloading argon2_cffi-25.1.0-py3-none-any.whl (14 kB)
Downloading async_lru-2.0.5-py3-none-any.whl (6.1 kB)
Downloading babel-2.17.0-py3-none-any.whl (10.2 MB)
----- 0.0/10.2 MB ? eta -:-:-
----- 1.3/10.2 MB 6.8 MB/s eta 0:00:02
----- 2.6/10.2 MB 7.0 MB/s eta 0:00:02
----- 3.9/10.2 MB 6.6 MB/s eta 0:00:01
----- 6.0/10.2 MB 7.5 MB/s eta 0:00:01
----- 7.6/10.2 MB 7.8 MB/s eta 0:00:01
----- 9.7/10.2 MB 8.1 MB/s eta 0:00:01
----- 10.2/10.2 MB 8.0 MB/s 0:00:01
Downloading h11-0.16.0-py3-none-any.whl (37 kB)
Downloading json5-0.12.1-py3-none-any.whl (36 kB)
Downloading jupyter_events-0.12.0-py3-none-any.whl (19 kB)
Downloading jsonpointer-3.0.0-py2.py3-none-any.whl (7.6 kB)
Downloading jupyter_lsp-2.3.0-py3-none-any.whl (76 kB)
Downloading jupyter_server_terminals-0.5.3-py3-none-any.whl (13 kB)
Downloading notebook_shim-0.2.4-py3-none-any.whl (13 kB)
Downloading prometheus_client-0.22.1-py3-none-any.whl (58 kB)
Downloading python_json_logger-3.3.0-py3-none-any.whl (15 kB)
Downloading pywinpty-3.0.0-cp313-cp313-win_amd64.whl (2.1 MB)
```

```
----- 0.0/2.1 MB ? eta -:-:--  
----- 1.0/2.1 MB 4.9 MB/s eta 0:00:01  
----- 2.1/2.1 MB 5.4 MB/s 0:00:00  
Using cached PyYAML-6.0.2-cp313-cp313-win_amd64.whl (156 kB)  
Downloading rfc3986_validator-0.1.1-py2.py3-none-any.whl (4.2 kB)  
Downloading rfc3987_syntax-1.1.0-py3-none-any.whl (8.0 kB)  
Downloading lark-1.2.2-py3-none-any.whl (111 kB)  
Downloading Send2Trash-1.8.3-py3-none-any.whl (18 kB)  
Downloading sniffio-1.3.1-py3-none-any.whl (10 kB)  
Downloading terminado-0.18.1-py3-none-any.whl (14 kB)  
Downloading webcolors-24.11.1-py3-none-any.whl (14 kB)  
Downloading websocket_client-1.8.0-py3-none-any.whl (58 kB)  
Downloading argon2_cffi_bindings-25.1.0-cp39-abi3-win_amd64.whl (31 kB)  
Downloading cffi-1.17.1-cp313-cp313-win_amd64.whl (182 kB)  
Downloading fqdn-1.5.1-py3-none-any.whl (9.1 kB)  
Downloading isoduration-20.11.0-py3-none-any.whl (11 kB)  
Downloading arrow-1.3.0-py3-none-any.whl (66 kB)  
Downloading types_python_dateutil-2.9.0.20250822-py3-none-any.whl (17 kB)  
Downloading jupyterlab_pygments-0.3.0-py3-none-any.whl (15 kB)  
Downloading notebook-7.4.5-py3-none-any.whl (14.3 MB)  
----- 0.0/14.3 MB ? eta -:-:--  
----- 1.3/14.3 MB 7.2 MB/s eta 0:00:02  
----- 3.1/14.3 MB 8.0 MB/s eta 0:00:02  
----- 5.8/14.3 MB 9.5 MB/s eta 0:00:01  
----- 7.9/14.3 MB 9.6 MB/s eta 0:00:01  
----- 10.2/14.3 MB 9.9 MB/s eta 0:00:01  
----- 12.8/14.3 MB 10.4 MB/s eta 0:00:01  
----- 14.3/14.3 MB 10.0 MB/s 0:00:01  
-----  
Downloading pycparser-2.22-py3-none-any.whl (117 kB)  
Downloading rfc3339_validator-0.1.4-py2.py3-none-any.whl (3.5 kB)  
Downloading uri_template-1.3.0-py3-none-any.whl (11 kB)  
Installing collected packages: webencodings, fastjsonschema, widgetsnbextension, web  
socket-client, webcolors, uri-template, types-python-dateutil, tinycss2, soupsieve,  
sniffio, send2trash, rpds-py, rfc3986-validator, rfc3339-validator, pyyaml, pywinpty,  
python-json-logger, pycparser, prometheus-client, pandocfilters, mistune, lark, j  
upyterlab_widgets, jupyterlab_pygments, jsonpointer, json5, h11, fqdn, defusedxml, b  
leach, babel, attrs, async-lru, terminado, rfc3987-syntax, referencing, httpcore, cf  
fi, beautifulsoup4, arrow, anyio, jupyter-server-terminals, jsonschema-specification  
s, isoduration, ipywidgets, httpx, argon2-cffi-bindings, jupyter-console, jsonschema  
a, argon2-cffi, nbformat, nbclient, jupyter-events, nbconvert, jupyter-server, noteb  
ook-shim, jupyterlab-server, jupyter-lsp, jupyterlab, notebook, jupyter  
----- 1/61 [fastjsonschema]  
----- 2/61 [widgetsnbextension]  
----- 3/61 [websocket-client]  
----- 3/61 [websocket-client]  
----- 3/61 [websocket-client]  
----- 4/61 [webcolors]  
----- 7/61 [tinycss2]  
----- 8/61 [soupsieve]  
----- 10/61 [send2trash]  
----- 10/61 [send2trash]  
----- 11/61 [rpds-py]  
----- 14/61 [pyyaml]  
----- 14/61 [pyyaml]  
----- 15/61 [pywinpty]
```

```
----- 17/61 [pycparser]
----- 17/61 [pycparser]
----- 18/61 [prometheus-client]
----- 18/61 [prometheus-client]
----- 19/61 [pandocfilters]
----- 20/61 [mistune]
----- 20/61 [mistune]
----- 20/61 [mistune]
----- 21/61 [lark]
----- 21/61 [lark]
----- 21/61 [lark]
----- 24/61 [jsonpointer]
----- 25/61 [json5]
----- 26/61 [h11]
----- 26/61 [h11]
----- 28/61 [defusedxml]
----- 29/61 [bleach]
----- 29/61 [bleach]
----- 29/61 [bleach]
----- 29/61 [bleach]
----- 30/61 [babel]
----- 31/61 [attrs]
----- 31/61 [attrs]
----- 35/61 [referencing]
----- 35/61 [referencing]
----- 36/61 [httpcore]
----- 36/61 [httpcore]
----- 36/61 [httpcore]
----- 37/61 [cffi]
----- 37/61 [cffi]
----- 38/61 [beautifulsoup4]
----- 38/61 [beautifulsoup4]
----- 39/61 [arrow]
----- 40/61 [anyio]
----- 40/61 [anyio]
----- 40/61 [anyio]
----- 41/61 [jupyter-server-terminals]
----- 43/61 [isoduration]
----- 44/61 [ipywidgets]
----- 45/61 [httpx]
----- 45/61 [httpx]
----- 45/61 [httpx]
----- 46/61 [argon2-cffi-bindings]
----- 47/61 [jupyter-console]
```

```
----- 48/61 [jsonschema]
----- 48/61 [jsonschema]
----- 48/61 [jsonschema]
----- 48/61 [jsonschema]
----- 49/61 [argon2-cffi]
----- 50/61 [nbformat]
----- 50/61 [nbformat]
----- 50/61 [nbformat]
----- 50/61 [nbformat]
----- 51/61 [nbclient]
----- 51/61 [nbclient]
----- 52/61 [jupyter-events]
----- 53/61 [nbconvert]
----- 54/61 [jupyter-server]
----- 55/61 [notebook-shim]
----- 56/61 [jupyterlab-server]
----- 56/61 [jupyterlab-server]
----- 57/61 [jupyter-lsp]
----- 57/61 [jupyter-lsp]
----- 57/61 [jupyter-lsp]
----- 57/61 [jupyter-lsp]
----- 58/61 [jupyterlab]
----- 59/61 [notebook]
----- 59/61 [notebook]
----- 59/61 [notebook]
----- 59/61 [notebook]
----- 61/61 [jupyter]
```

Successfully installed anyio-4.10.0 argon2-cffi-25.1.0 argon2-cffi-bindings-25.1.0 arrow-1.3.0 async-lru-2.0.5 attrs-25.3.0 babel-2.17.0 beautifulsoup4-4.4.13.5 bleach-6.2.0 cffi-1.17.1 defusedxml-0.7.1 fastjsonschema-2.21.2 fqdn-1.5.1 h11-0.16.0 httpcore-1.0.9 httpx-0.28.1 ipywidgets-8.1.7 isoduration-20.11.0 json5-0.12.1 jsonpointer-3.0.0 jsonschema-4.25.1 jsonschema-specifications-2025.4.1 jupyter-1.1.1 jupyter-con

```
sole-6.6.3 jupyter-events-0.12.0 jupyter-lsp-2.3.0 jupyter-server-2.17.0 jupyter-server-terminals-0.5.3 jupyterlab-4.4.6 jupyterlab-pygments-0.3.0 jupyterlab-server-2.27.3 jupyterlab_widgets-3.0.15 lark-1.2.2 mistune-3.1.3 nbclient-0.10.2 nbconvert-7.16.6 nbformat-5.10.4 notebook-7.4.5 notebook-shim-0.2.4 pandocfilters-1.5.1 prometheus-client-0.22.1 pyparser-2.22 python-json-logger-3.3.0 pywinpty-3.0.0 pyyaml-6.0.2referencing-0.36.2 rfc3339-validator-0.1.4 rfc3986-validator-0.1.1 rfc3987-syntax-1.1.0 rpds-py-0.27.1 send2trash-1.8.3 sniffio-1.3.1 soupsieve-2.8 terminado-0.18.1 tinyrss-1.4.0 types-python-dateutil-2.9.0.20250822 uri-template-1.3.0 webcolors-24.11.1 webencodings-0.5.1 websocket-client-1.8.0 widgetsnbextension-4.0.14
```

Note: you may need to restart the kernel to use updated packages.

After I find features that effects the price with plot and heatmap. After that I found engine, power and manufacture year are the one and I choose to use them to train. I thought driven km would be the one since plot show the less the km the pricier the car. However after getting heatmap, I ignore that. There are a lot of outliers in engine and power. However, I think sport cars with high engine power are pricy and realistic in real life and a lot of outliers mean this is not wrong data since its got market.

I tried with different algorithm and I found R square and MSE result of random forest is the best and use to test with it. And I tried to find which features effect the most with shap and manufacture year is the one on top.