CHRIST

(DEEMED TO BE UNIVERSITY)

DELHI-NCR, INDIA

AI/ML Programming

MCA-475

Assignment – 01

BY

HIMANSHU HEDA (24225013)

SUBMITTED TO

Dr. Manjula Shannhog

SCHOOL OF SCIENCES

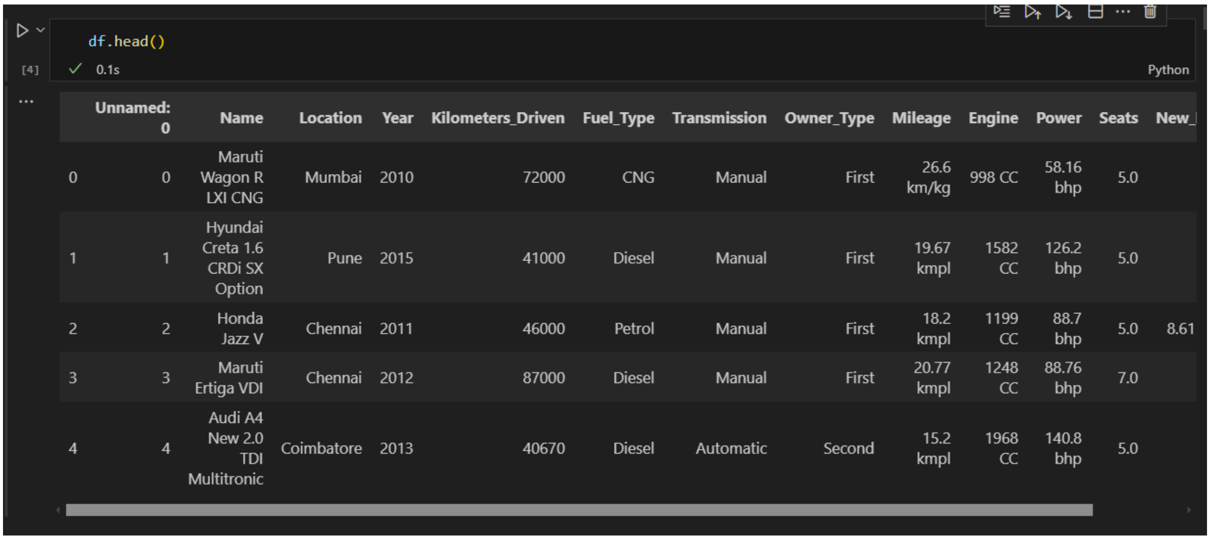
2024-25

Importing Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import date
```

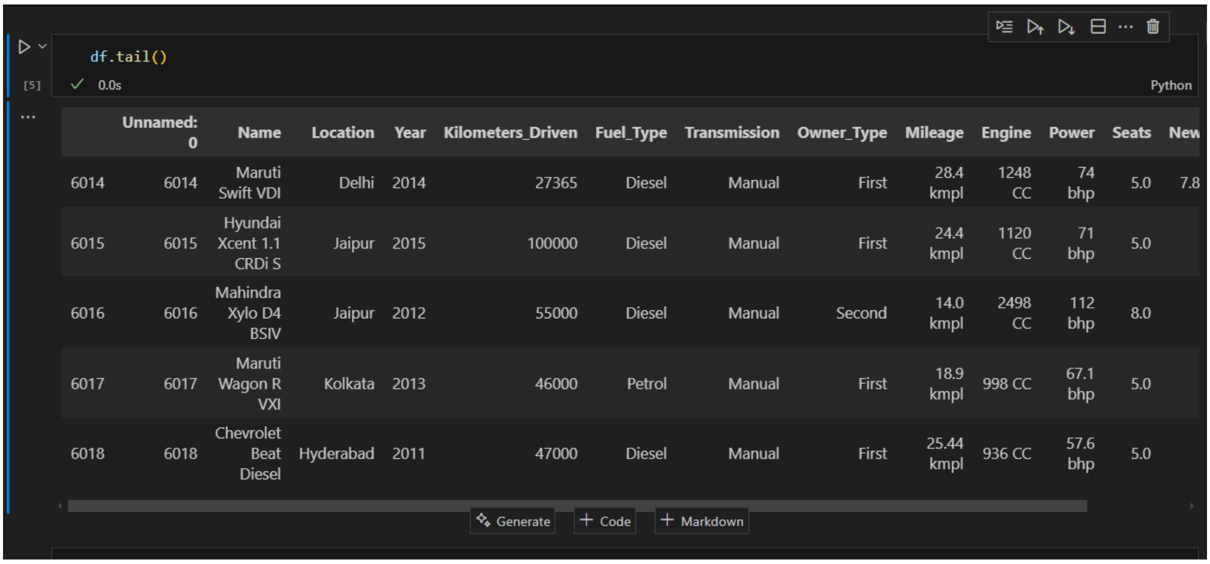
```
df = pd.read_csv('./Dataset/used_car_dataset.csv')
```

```
df.head()
```



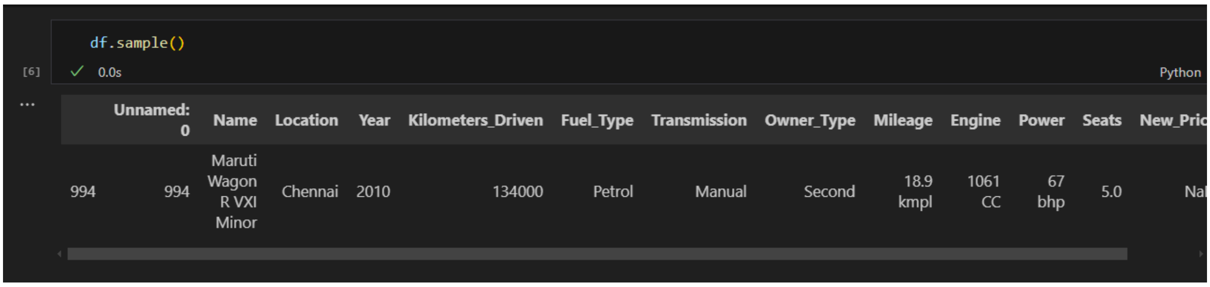
```
df.head()
[4] ✓ 0.1s                                                                 Python
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	



```
df.tail()
[5] ✓ 0.0s                                                                 Python
```

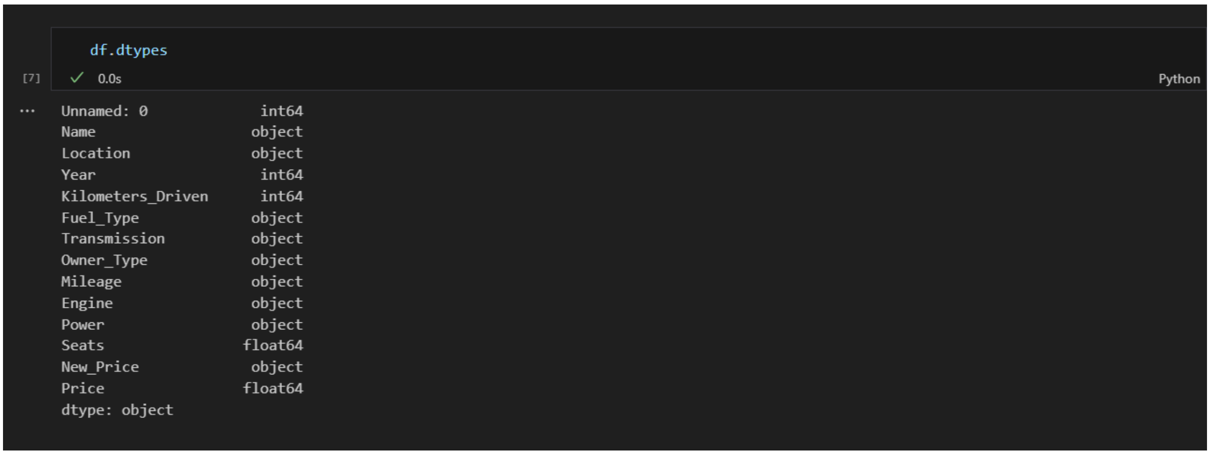
	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.4 kmpl	1248 CC	74 bhp	5.0	7.8
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	First	24.4 kmpl	1120 CC	71 bhp	5.0	
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual	Second	14.0 kmpl	2498 CC	112 bhp	8.0	
6017	6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual	First	18.9 kmpl	998 CC	67.1 bhp	5.0	
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual	First	25.44 kmpl	936 CC	57.6 bhp	5.0	



[6]

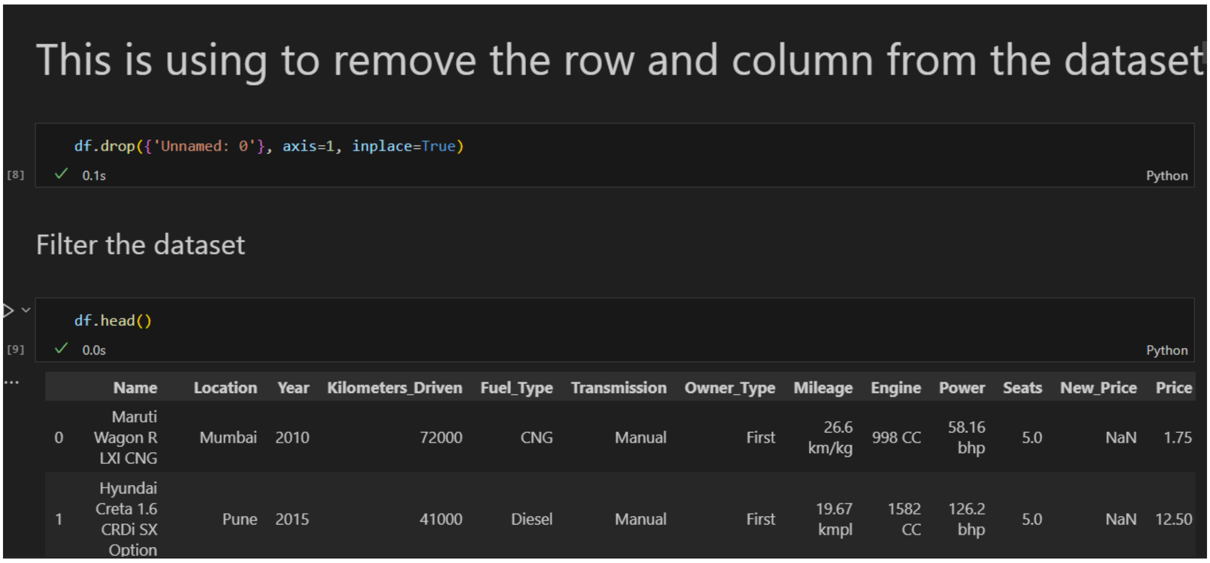
```
df.sample()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Pric
994	994	Maruti Wagon R VXI Minor	Chennai	2010	134000	Petrol	Manual	Second	18.9 kmpl	1061 CC	67 bhp	5.0	Na



```
df.dtypes
```

```
Unnamed: 0             int64
Name                  object
Location              object
Year                   int64
Kilometers_Driven      int64
Fuel_Type             object
Transmission          object
Owner_Type            object
Mileage               object
Engine                object
Power                 object
Seats                float64
New_Price             object
Price                float64
dtype: object
```



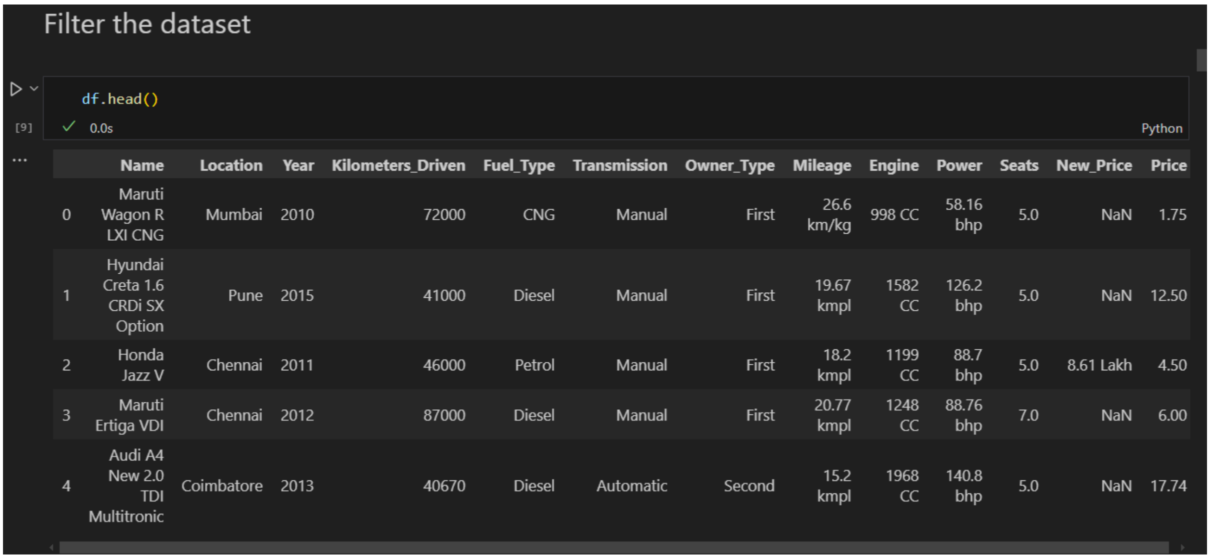
This is using to remove the row and column from the dataset

```
df.drop({'Unnamed: 0'}, axis=1, inplace=True)
```

Filter the dataset

```
df.head()
```

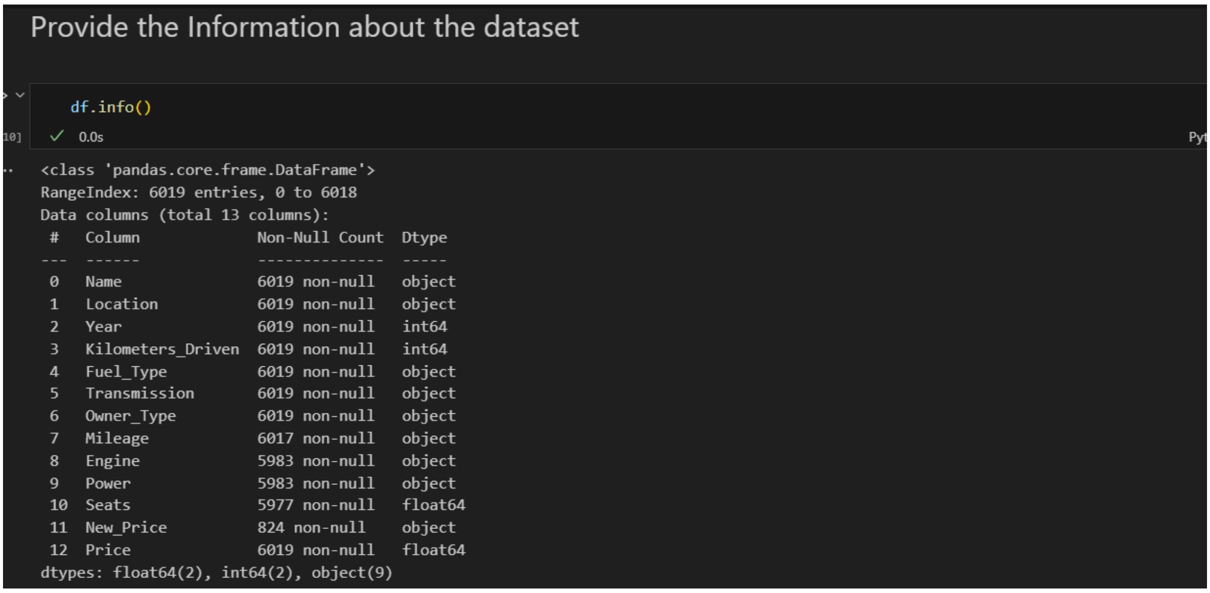
	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75
1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50



Filter the dataset

```
df.head()
```

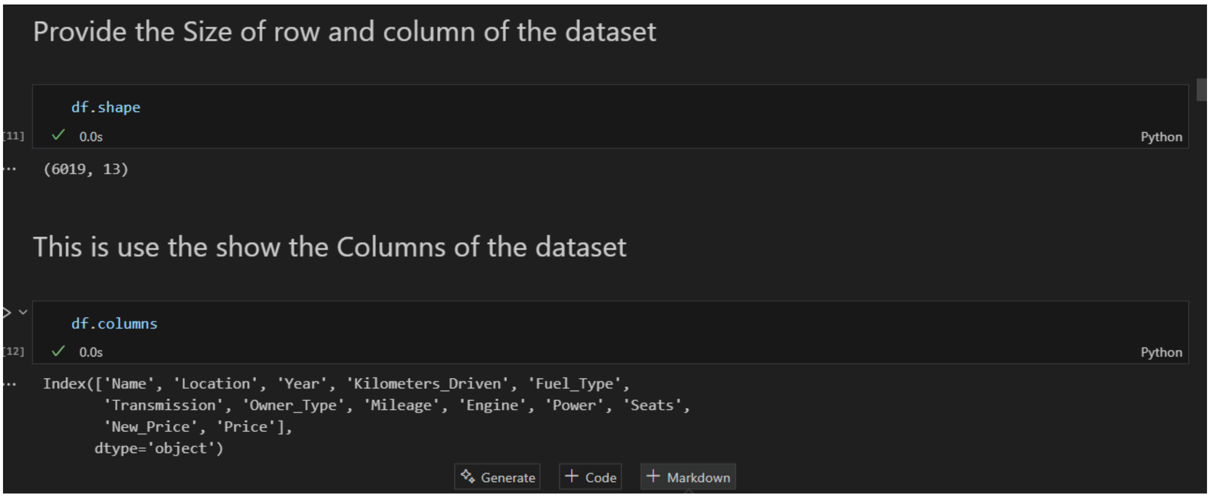
	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75
1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN	6.00
4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74



Provide the Information about the dataset

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6019 entries, 0 to 6018
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Name               6019 non-null   object 
 1   Location           6019 non-null   object 
 2   Year               6019 non-null   int64  
 3   Kilometers_Driven  6019 non-null   int64  
 4   Fuel_Type          6019 non-null   object 
 5   Transmission       6019 non-null   object 
 6   Owner_Type         6019 non-null   object 
 7   Mileage            6017 non-null   object 
 8   Engine             5983 non-null   object 
 9   Power              5983 non-null   object 
 10  Seats              5977 non-null   float64
 11  New_Price          824 non-null    object 
 12  Price              6019 non-null   float64
dtypes: float64(2), int64(2), object(9)
```



Provide the Size of row and column of the dataset

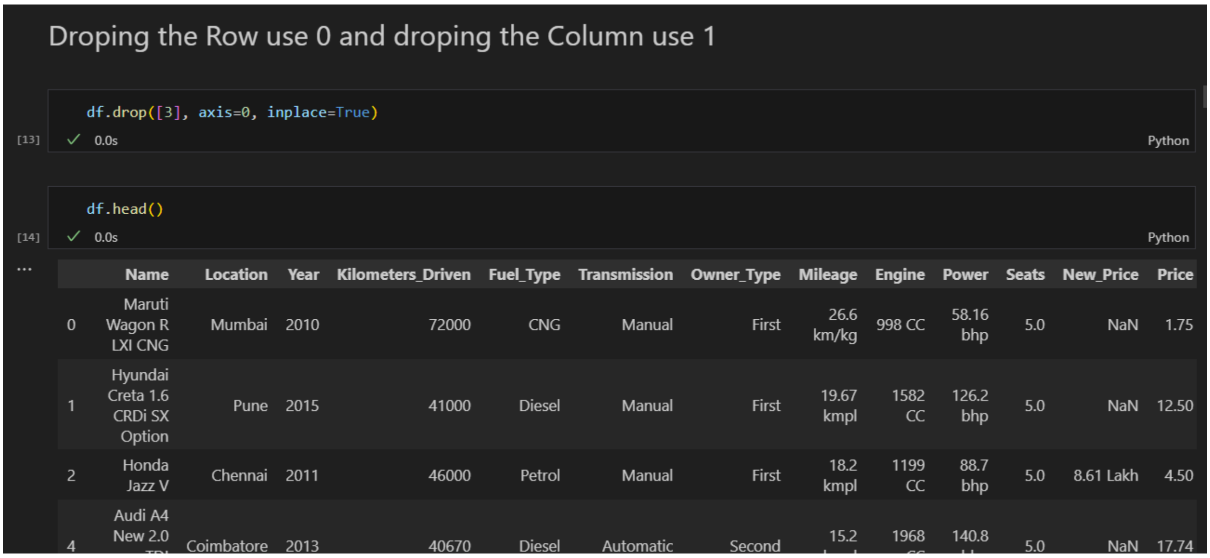
```
df.shape
```

```
(6019, 13)
```

This is use the show the Columns of the dataset

```
df.columns
```

```
Index(['Name', 'Location', 'Year', 'Kilometers_Driven', 'Fuel_Type',
       'Transmission', 'Owner_Type', 'Mileage', 'Engine', 'Power', 'Seats',
       'New_Price', 'Price'],
      dtype='object')
```

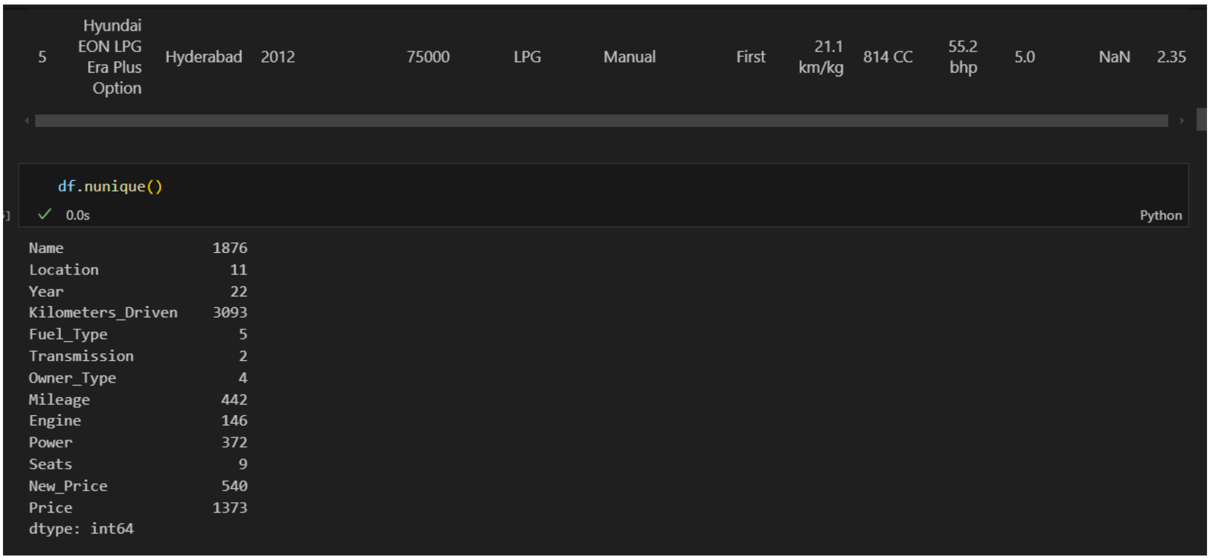


Droping the Row use 0 and droping the Column use 1

```
df.drop([3], axis=0, inplace=True)
```

```
df.head()
```

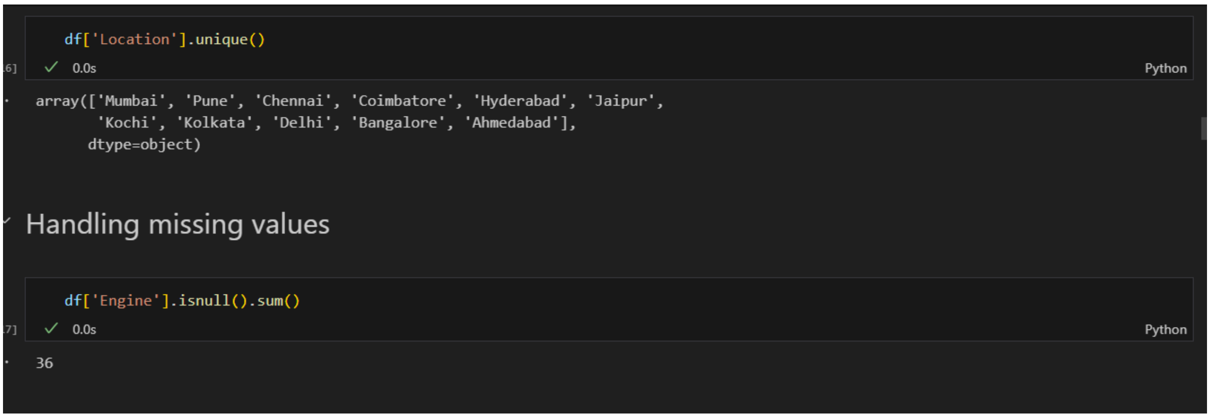
	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75
1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
4	Audi A4 New 2.0	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2	1968	140.8	5.0	NaN	17.74



5	Hyundai EON LPG Era Plus Option	Hyderabad	2012	75000	LPG	Manual	First	21.1 km/kg	814 CC	55.2 bhp	5.0	NaN	2.35

```python
df.nunique()
```

```
Name                 1876
Location               11
Year                   22
Kilometers_Driven    3093
Fuel_Type               5
Transmission            2
Owner_Type              4
Mileage               442
Engine                146
Power                 372
Seats                   9
New_Price             540
Price                1373
dtype: int64
```



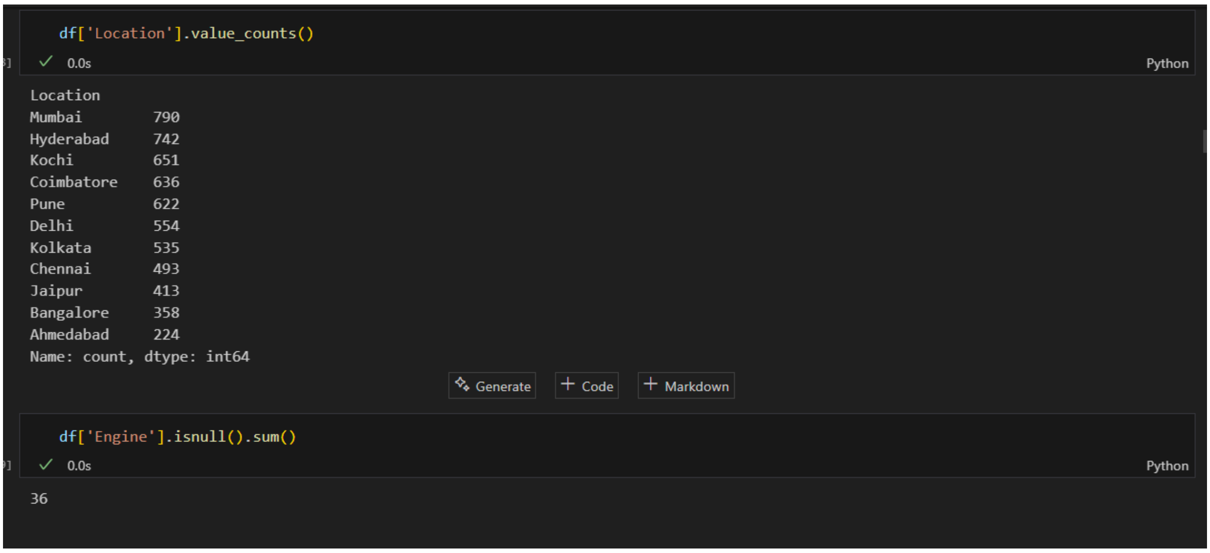
```python
df['Location'].unique()
```

```
array(['Mumbai', 'Pune', 'Chennai', 'Coimbatore', 'Hyderabad', 'Jaipur',
       'Kochi', 'Kolkata', 'Delhi', 'Bangalore', 'Ahmedabad'],
      dtype=object)
```

Handling missing values

```python
df['Engine'].isnull().sum()
```

```
36
```

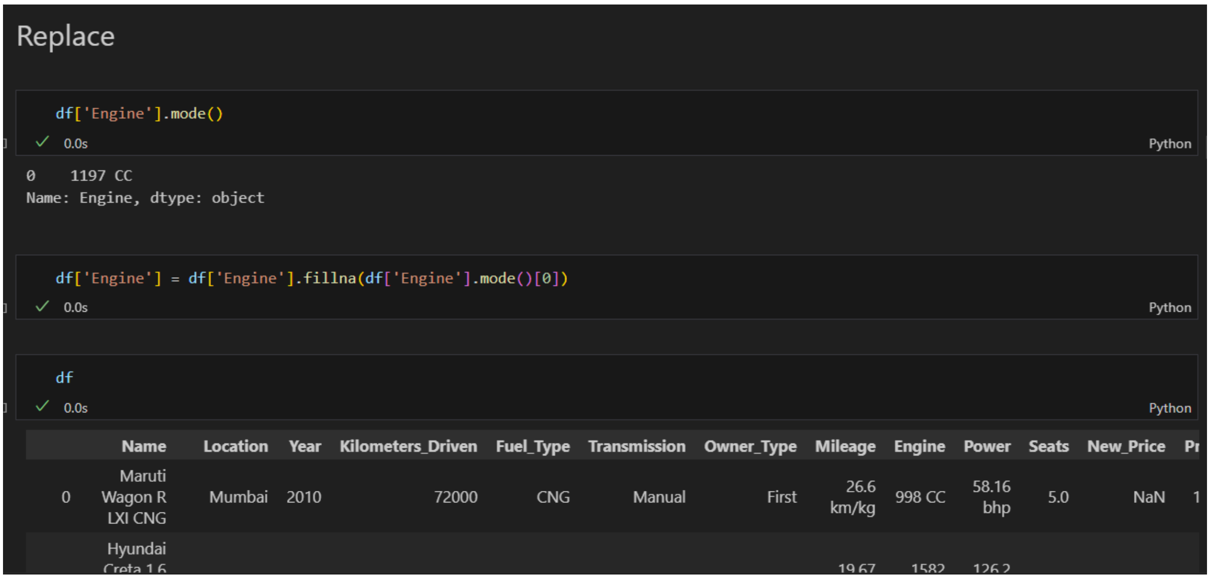


```python
df['Location'].value_counts()
```

```
Location
Mumbai        790
Hyderabad     742
Kochi         651
Coimbatore    636
Pune          622
Delhi         554
Kolkata       535
Chennai       493
Jaipur        413
Bangalore     358
Ahmedabad     224
Name: count, dtype: int64
```

```python
df['Engine'].isnull().sum()
```

```
36
```



Replace

```
df['Engine'].mode()
```

✓ 0.0s Python

```
0    1197 CC
Name: Engine, dtype: object
```

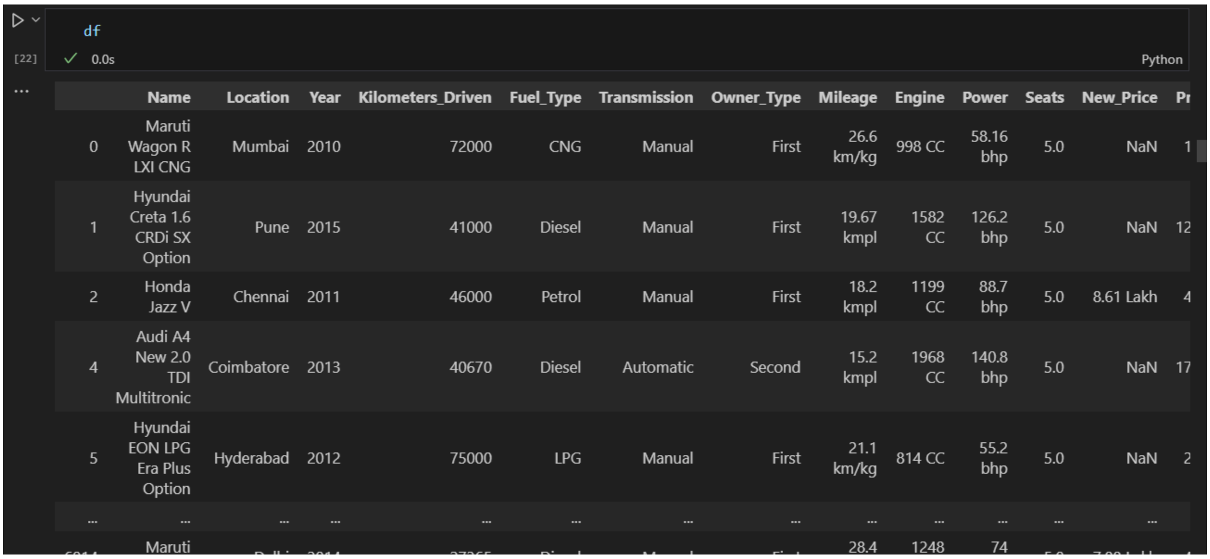
```
df['Engine'] = df['Engine'].fillna(df['Engine'].mode()[0])
```

✓ 0.0s Python

```
df
```

✓ 0.0s Python

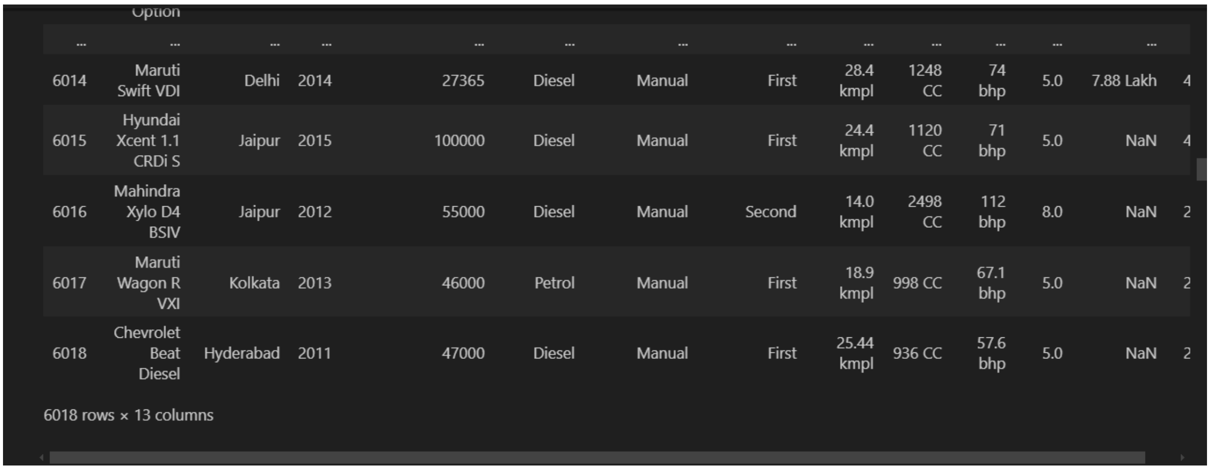
	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Pr
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1
	Hyundai Creta 1.6							19.67	1582	126.2			



```
df
```

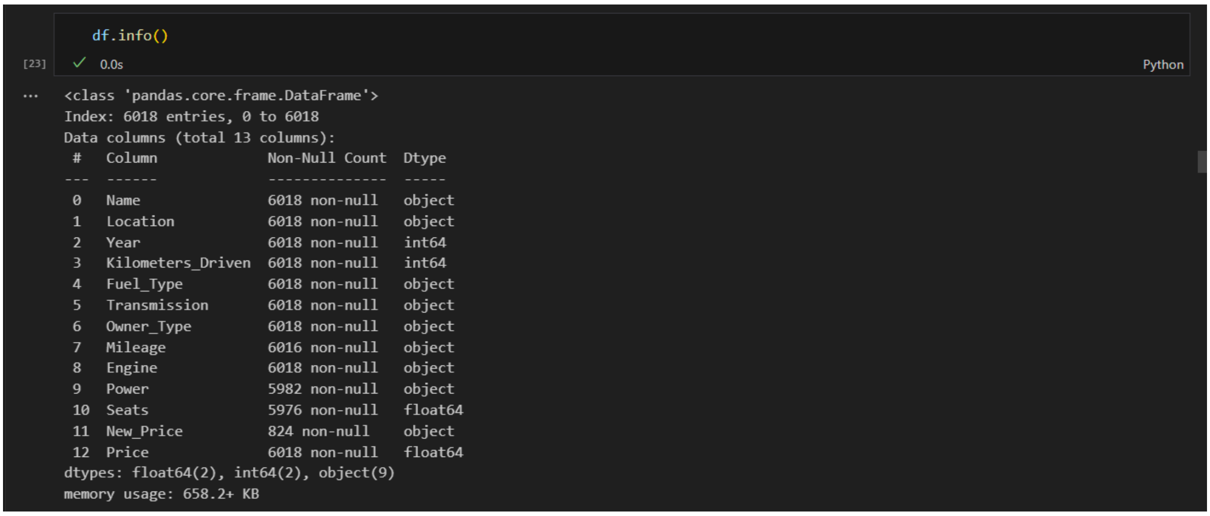
[22] ✓ 0.0s Python

	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Pr
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1
1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12
2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4
4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17
5	Hyundai EON LPG Era Plus Option	Hyderabad	2012	75000	LPG	Manual	First	21.1 km/kg	814 CC	55.2 bhp	5.0	NaN	2
...	...	...	...	...	...	...	...	...	...	...	...	...	
	Maruti							28.4	1248	74			



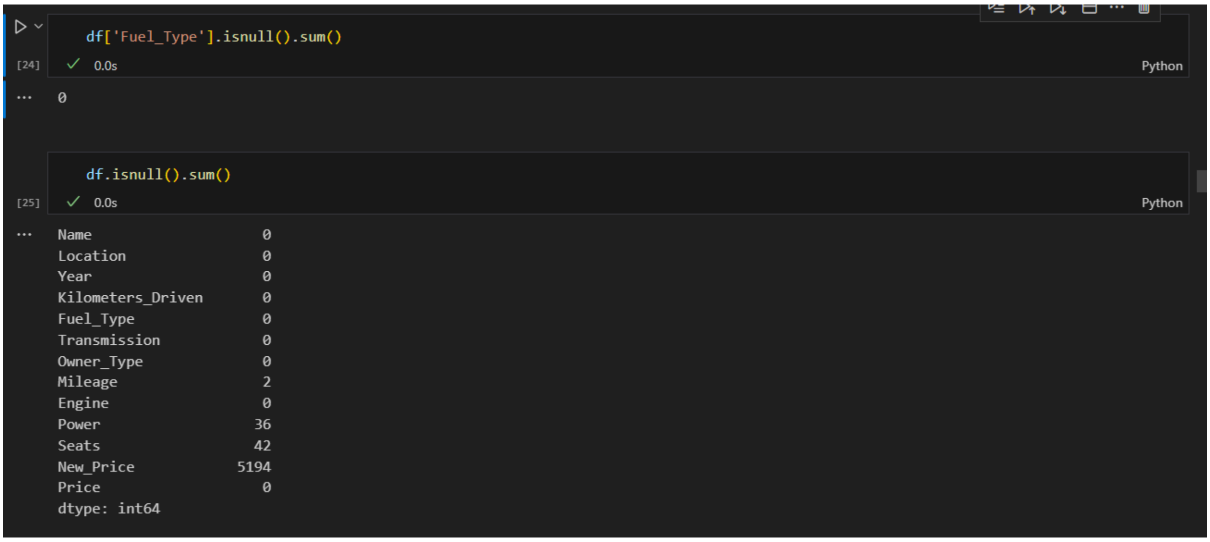
	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Pr
	Option												
...	...	...	...	...	...	...	...	...	...	...	...	...	
6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.4 kmpl	1248 CC	74 bhp	5.0	7.88 Lakh	4
6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	First	24.4 kmpl	1120 CC	71 bhp	5.0	NaN	4
6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual	Second	14.0 kmpl	2498 CC	112 bhp	8.0	NaN	2
6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual	First	18.9 kmpl	998 CC	67.1 bhp	5.0	NaN	2
6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual	First	25.44 kmpl	936 CC	57.6 bhp	5.0	NaN	2

6018 rows × 13 columns



```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 6018 entries, 0 to 6018
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Name               6018 non-null   object 
 1   Location           6018 non-null   object 
 2   Year               6018 non-null   int64  
 3   Kilometers_Driven  6018 non-null   int64  
 4   Fuel_Type          6018 non-null   object 
 5   Transmission       6018 non-null   object 
 6   Owner_Type         6018 non-null   object 
 7   Mileage            6016 non-null   object 
 8   Engine             6018 non-null   object 
 9   Power              5982 non-null   object 
 10  Seats              5976 non-null   float64
 11  New_Price          824 non-null    object 
 12  Price              6018 non-null   float64
dtypes: float64(2), int64(2), object(9)
memory usage: 658.2+ KB
```

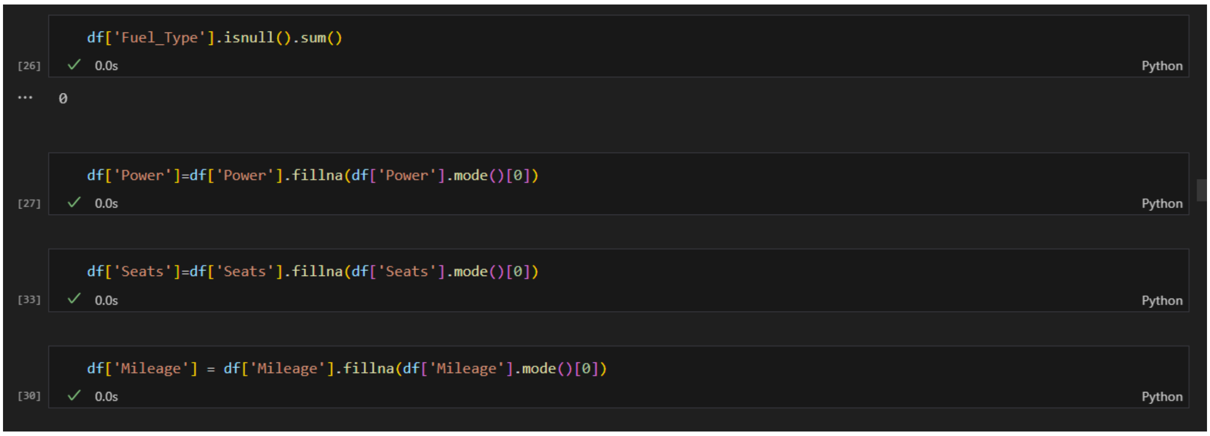


```python
df['Fuel_Type'].isnull().sum()
```

```
0
```

```python
df.isnull().sum()
```

```
Name                    0
Location                0
Year                    0
Kilometers_Driven       0
Fuel_Type               0
Transmission            0
Owner_Type              0
Mileage                 2
Engine                  0
Power                  36
Seats                  42
New_Price            5194
Price                   0
dtype: int64
```



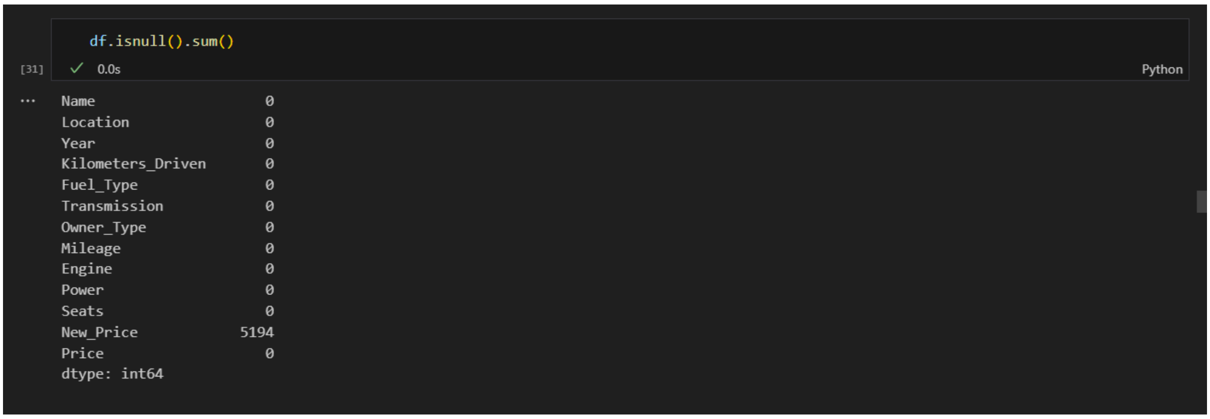
```python
df['Fuel_Type'].isnull().sum()
```

```
0
```

```python
df['Power']=df['Power'].fillna(df['Power'].mode()[0])
```

```python
df['Seats']=df['Seats'].fillna(df['Seats'].mode()[0])
```

```python
df['Mileage'] = df['Mileage'].fillna(df['Mileage'].mode()[0])
```

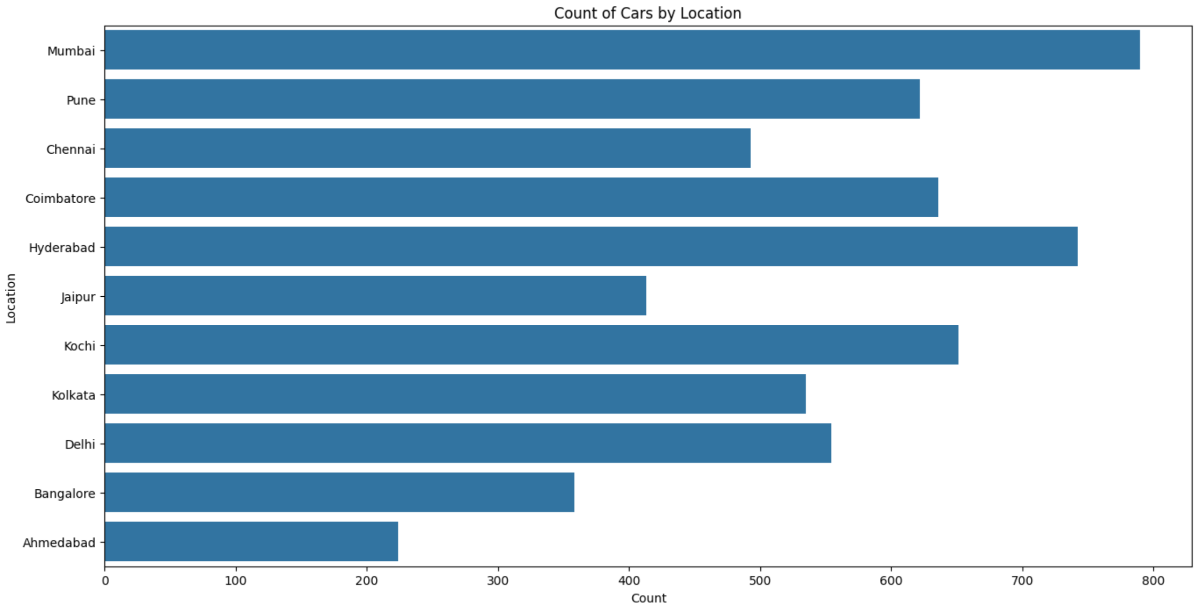


```
df.isnull().sum()
```

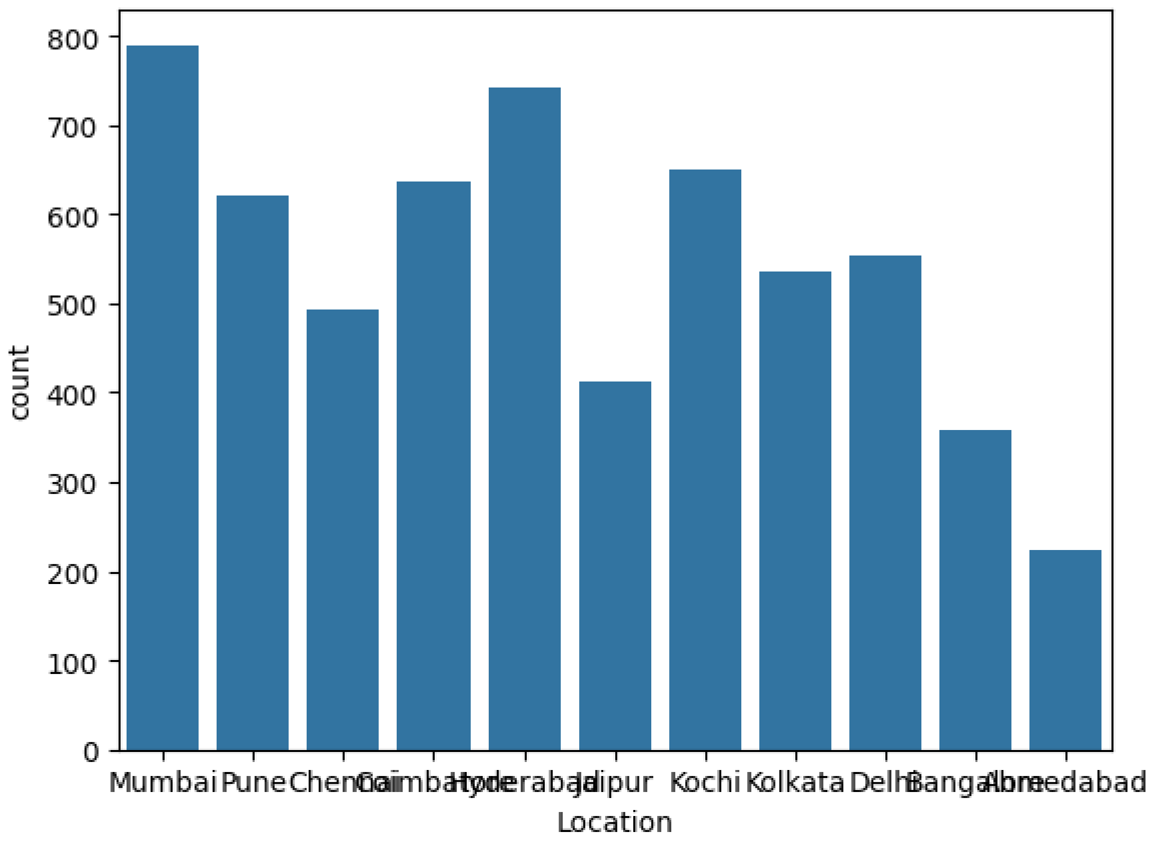
```
Name                    0
Location                0
Year                    0
Kilometers_Driven       0
Fuel_Type               0
Transmission            0
Owner_Type              0
Mileage                 0
Engine                  0
Power                   0
Seats                   0
New_Price            5194
Price                   0
dtype: int64
```

Univarte Analysis

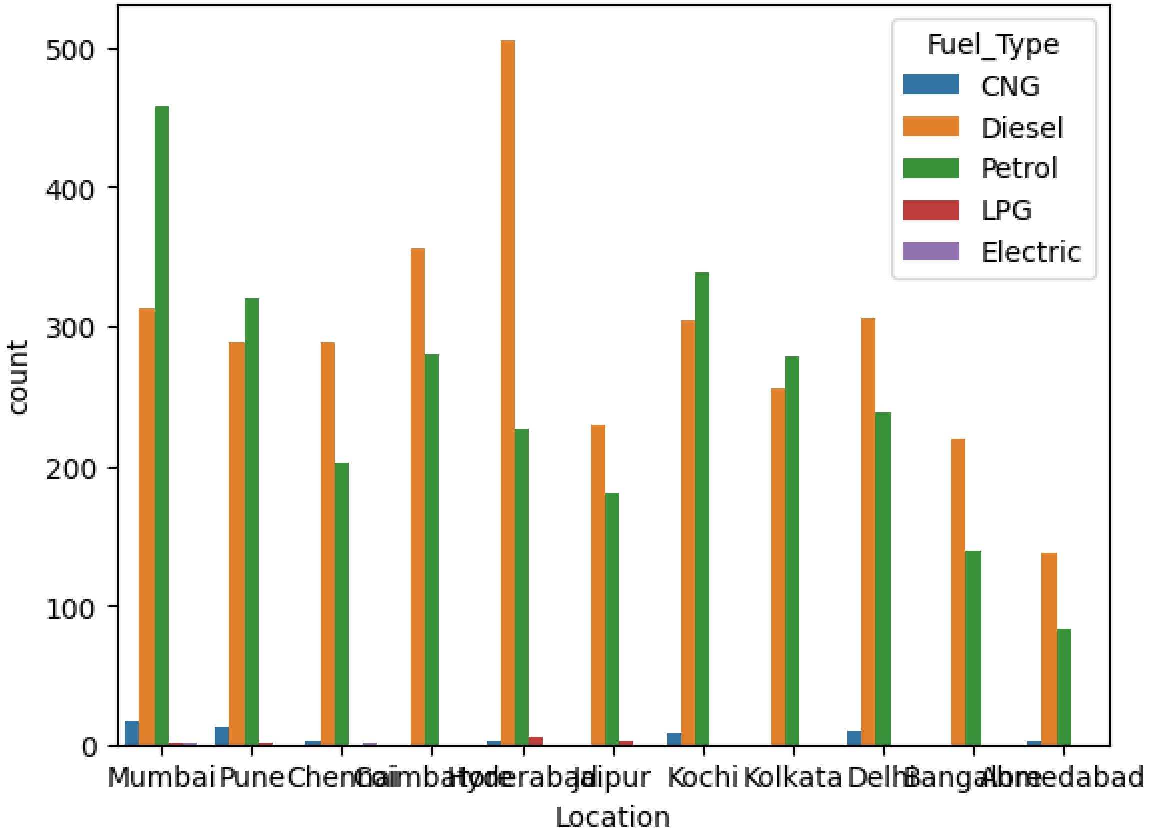
```
plt.figure(figsize=(16, 8))
sns.countplot(data=df, y='Location')
plt.title('Count of Cars by Location')
plt.xlabel('Count')
plt.ylabel('Location')
plt.show()
```

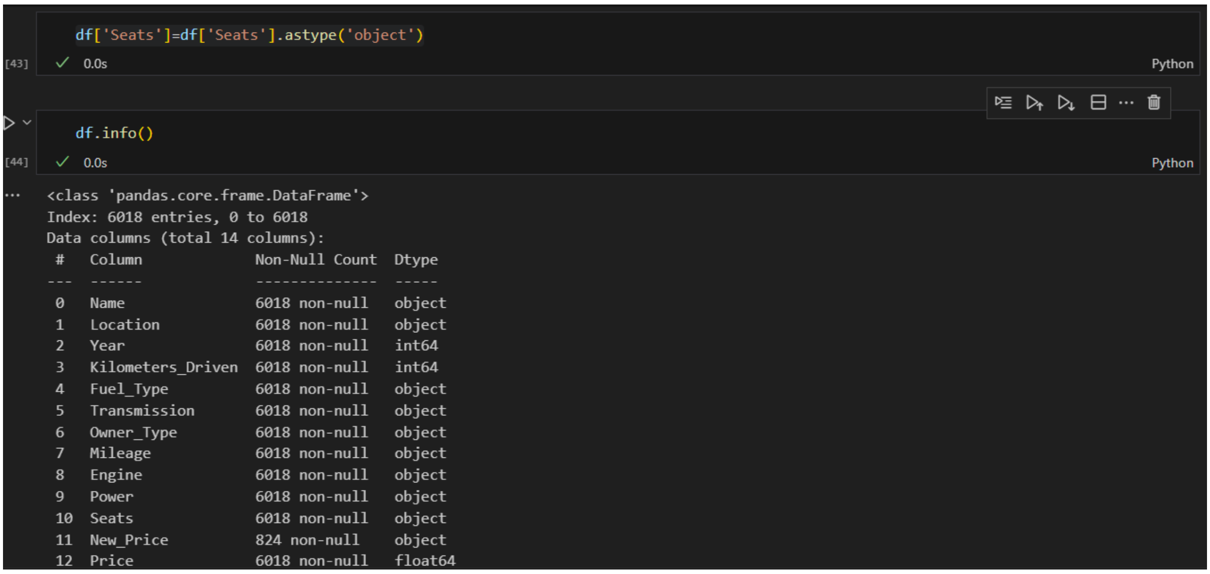


```
sns.countplot(data=df, x='Location')
plt.figure(figsize=(16, 8))
```



```
sns.countplot(data=df, x='Location', hue='Fuel_Type')
```



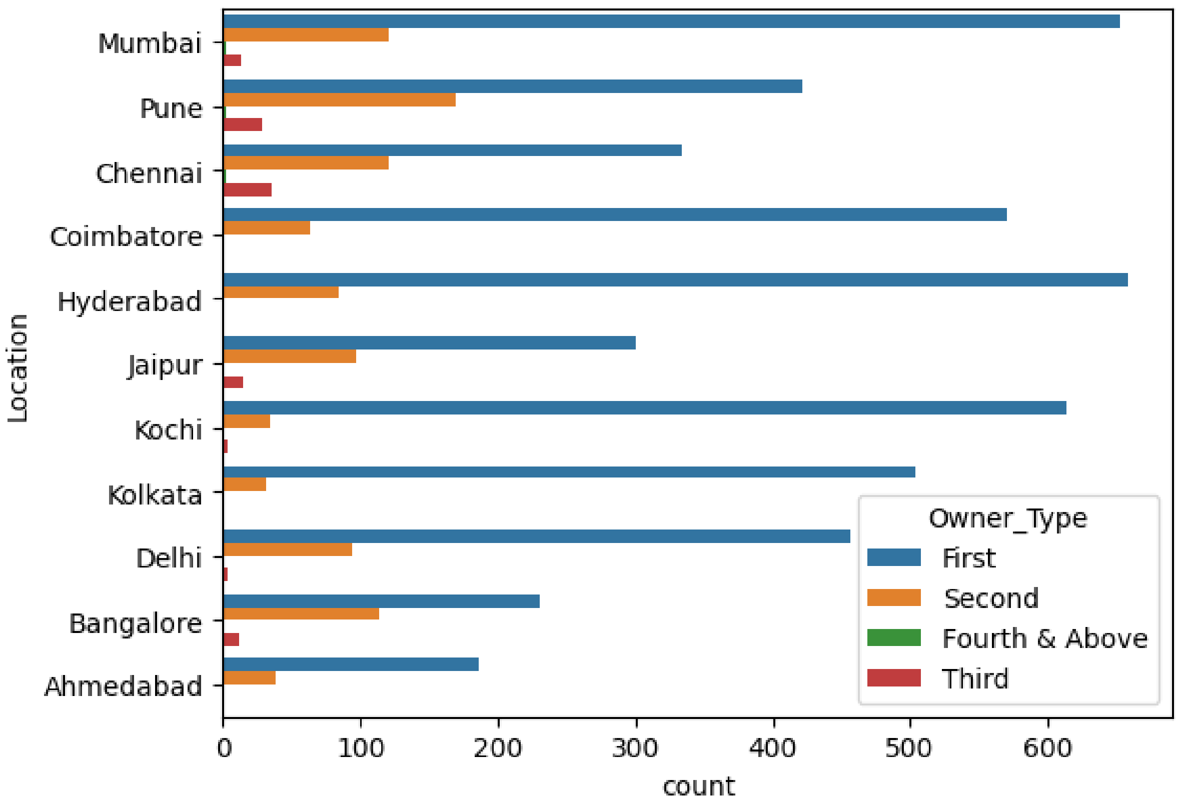


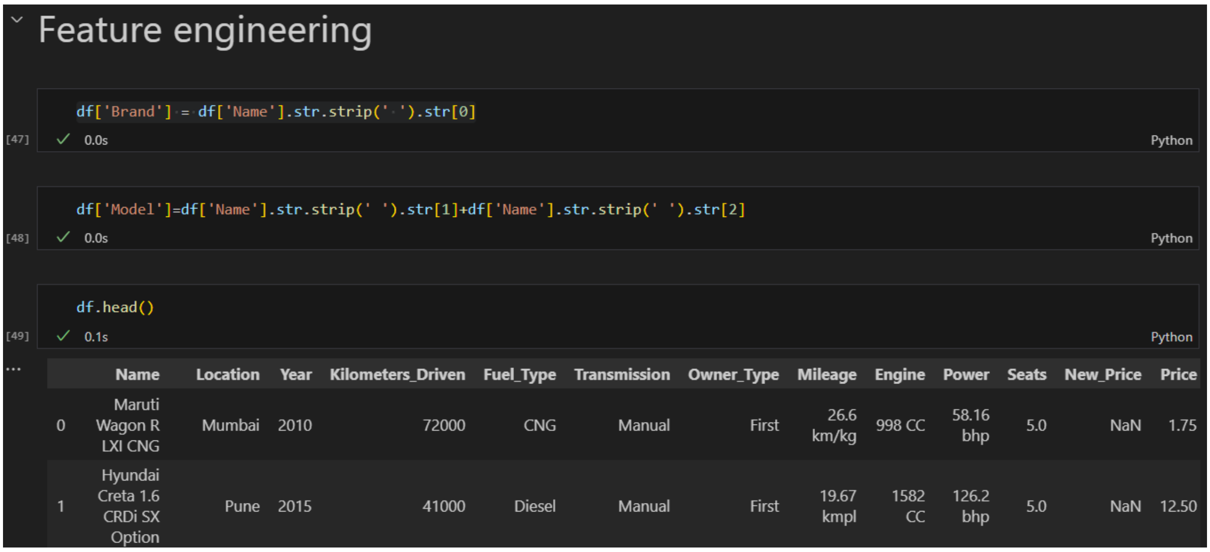
```
df['Seats']=df['Seats'].astype('object')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 6018 entries, 0 to 6018
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Name               6018 non-null   object 
 1   Location           6018 non-null   object 
 2   Year               6018 non-null   int64  
 3   Kilometers_Driven  6018 non-null   int64  
 4   Fuel_Type          6018 non-null   object 
 5   Transmission       6018 non-null   object 
 6   Owner_Type         6018 non-null   object 
 7   Mileage            6018 non-null   object 
 8   Engine             6018 non-null   object 
 9   Power              6018 non-null   object 
 10  Seats              6018 non-null   object 
 11  New_Price          824 non-null    object 
 12  Price              6018 non-null   float64
```

```
sns.countplot(data=df, y='Location', hue='Owner_Type')
```





Feature engineering

```
df['Brand'] = df['Name'].str.strip(' ').str[0]
```

[47] ✓ 0.0s Python

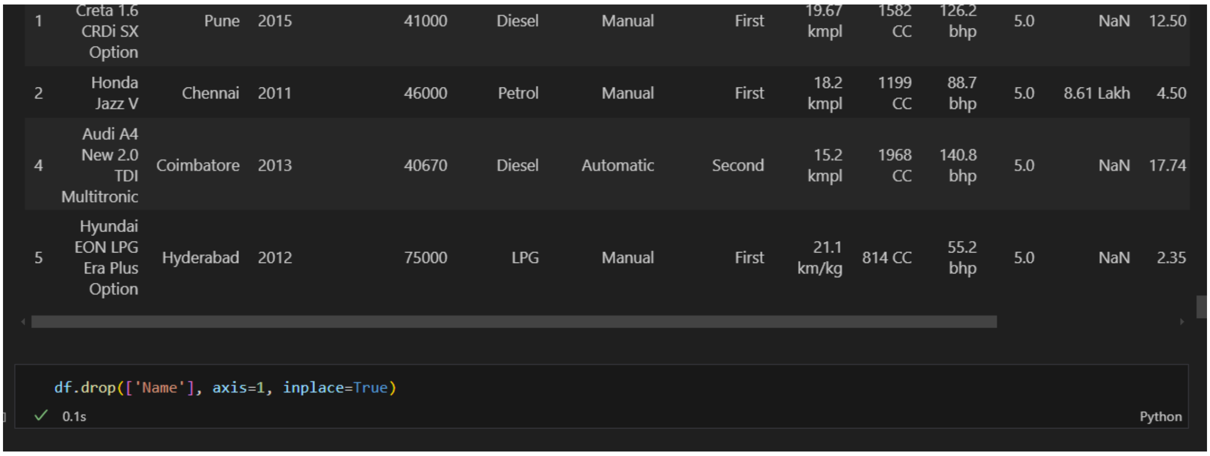
```
df['Model']=df['Name'].str.strip(' ').str[1]+df['Name'].str.strip(' ').str[2]
```

[48] ✓ 0.0s Python

```
df.head()
```

[49] ✓ 0.1s Python

	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75
1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50

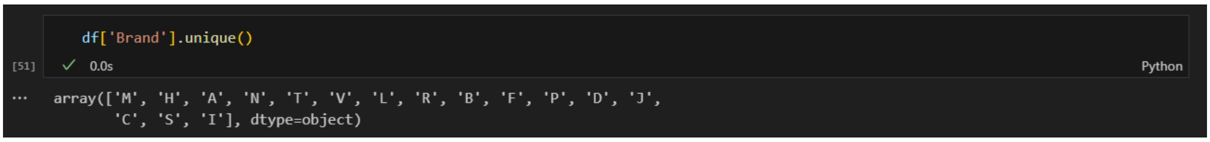


1	Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74
5	Hyundai EON LPG Era Plus Option	Hyderabad	2012	75000	LPG	Manual	First	21.1 km/kg	814 CC	55.2 bhp	5.0	NaN	2.35

```
df.drop(['Name'], axis=1, inplace=True)
```

✓ 0.1s

Python



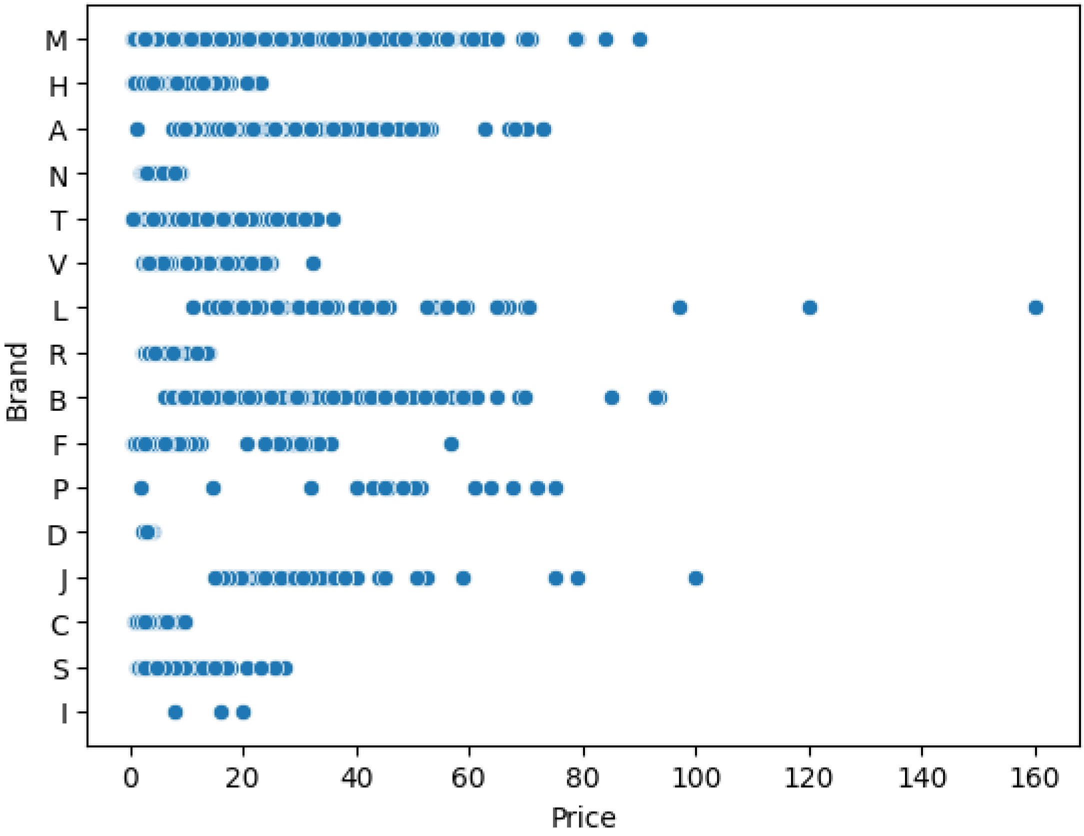
```
df['Brand'].unique()
```

[51] ✓ 0.0s

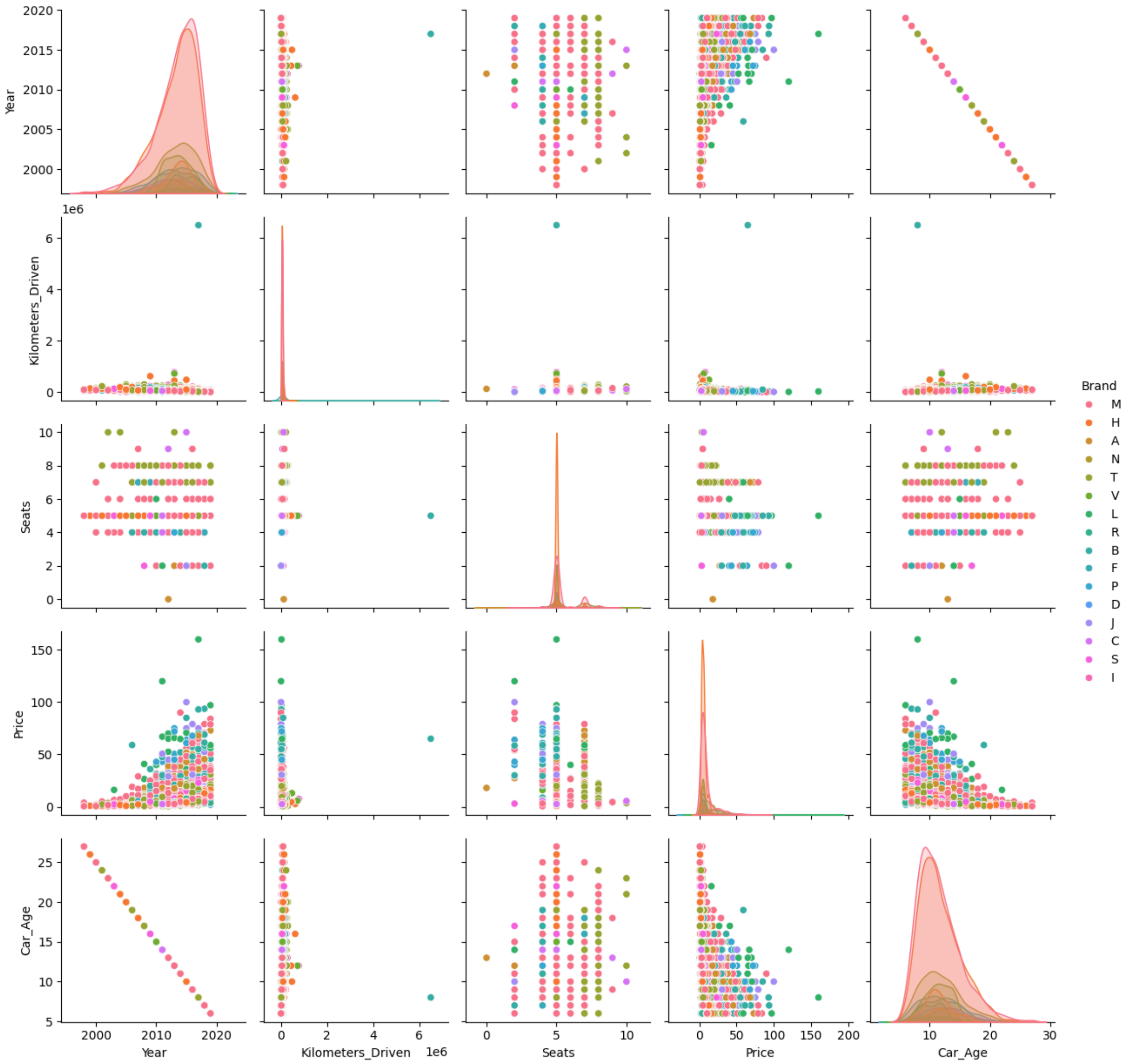
Python

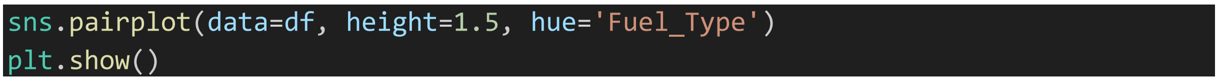
```
array(['M', 'H', 'A', 'N', 'T', 'V', 'L', 'R', 'B', 'F', 'P', 'D', 'J',
       'C', 'S', 'I'], dtype=object)
```

```
sns.scatterplot(data=df, y='Brand' ,x='Price')
```

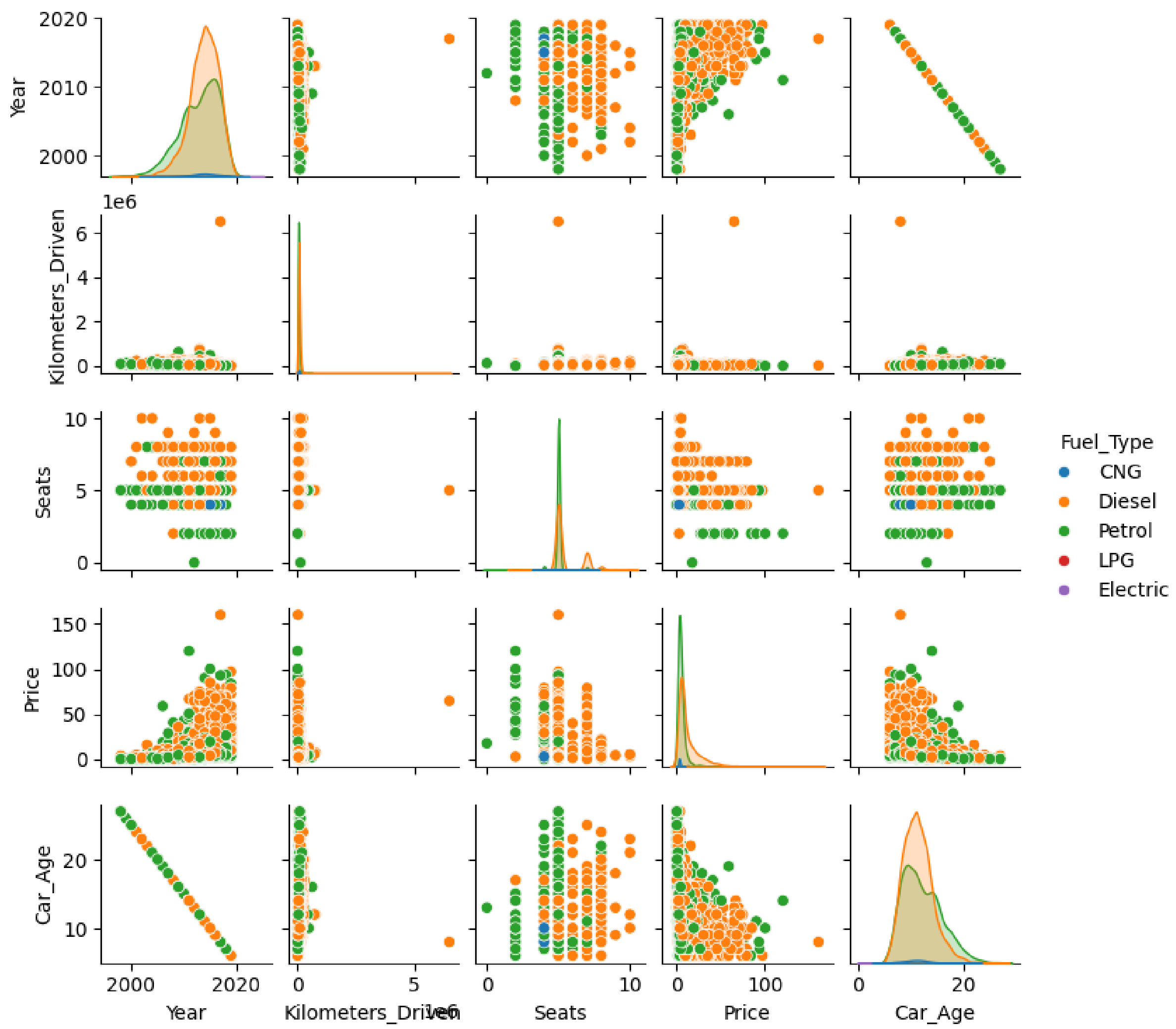


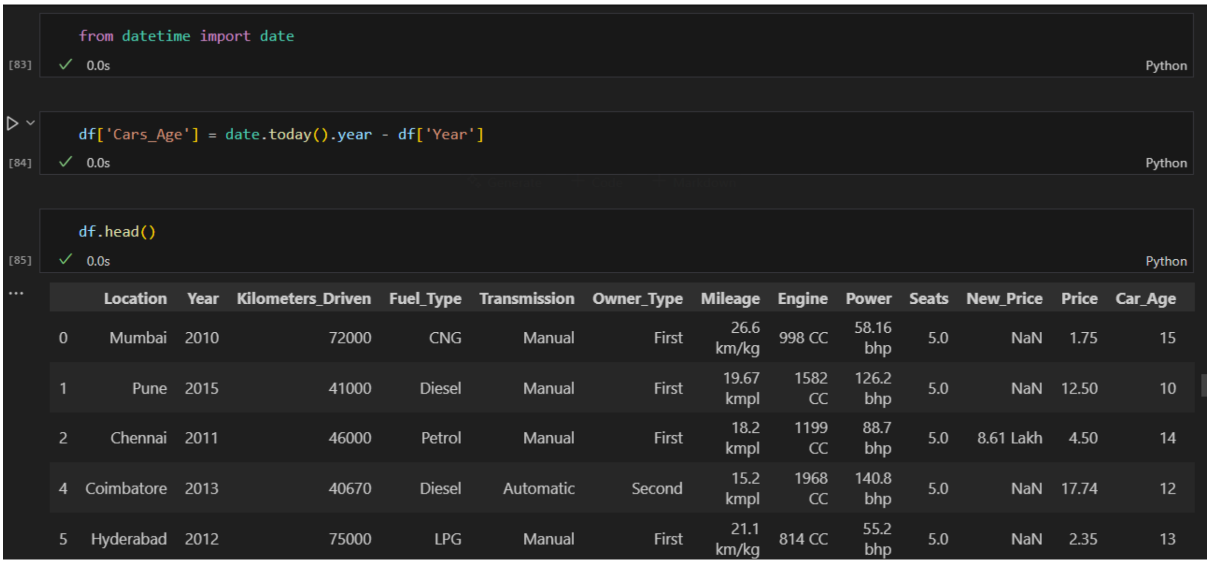
```
sns.pairplot(df, hue='Brand', diag_kind='kde')
```





```
sns.pairplot(data=df, height=1.5, hue='Fuel_Type')
plt.show()
```



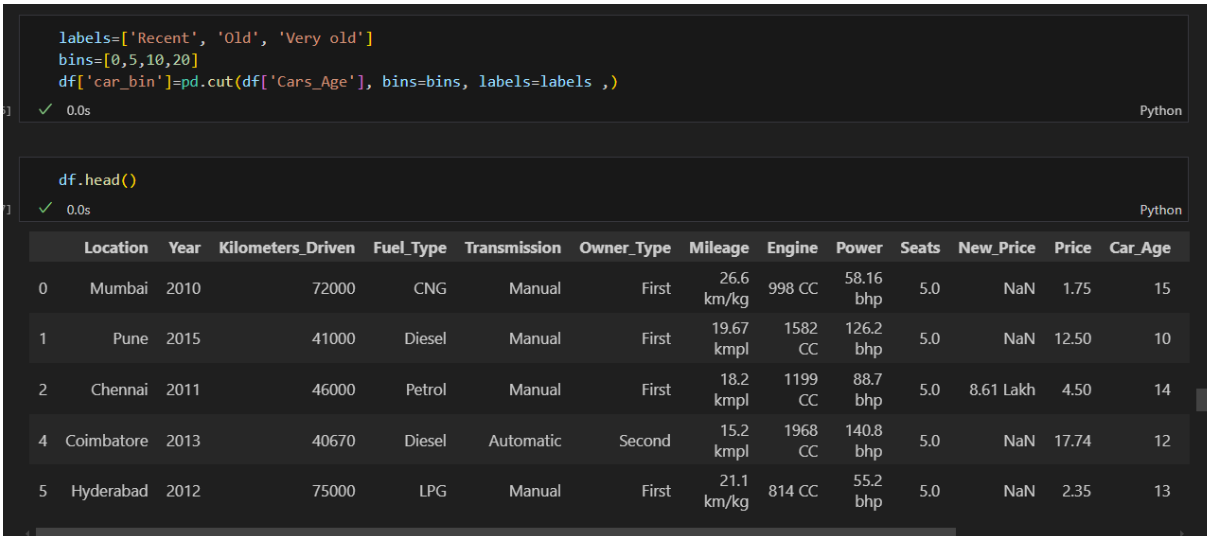


```
[83]
from datetime import date
✓ 0.0s                                                                 Python
```

```
[84]
df['Cars_Age'] = date.today().year - df['Year']
✓ 0.0s                                                                 Python
```

```
[85]
df.head()
✓ 0.0s                                                                 Python
```

	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Car_Age
0	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75	15
1	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50	10
2	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50	14
4	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74	12
5	Hyderabad	2012	75000	LPG	Manual	First	21.1 km/kg	814 CC	55.2 bhp	5.0	NaN	2.35	13



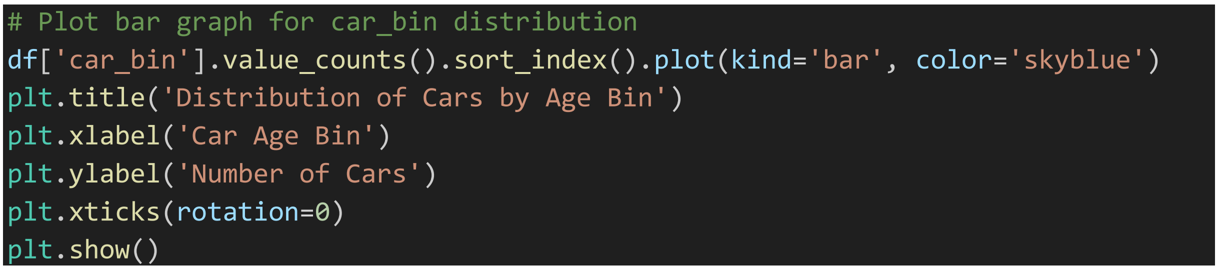
```
labels=['Recent', 'Old', 'Very old']
bins=[0,5,10,20]
df['car_bin']=pd.cut(df['Cars_Age'], bins=bins, labels=labels ,)
```

✓ 0.0s Python

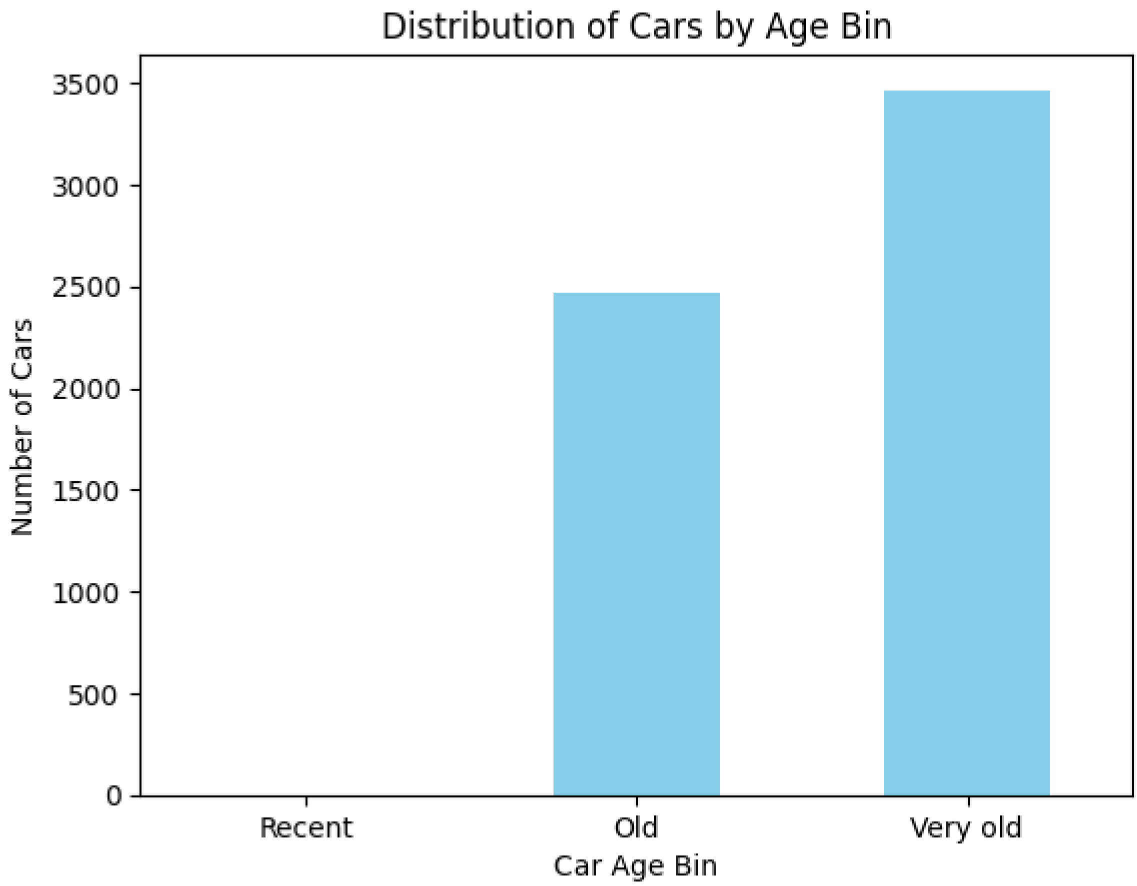
```
df.head()
```

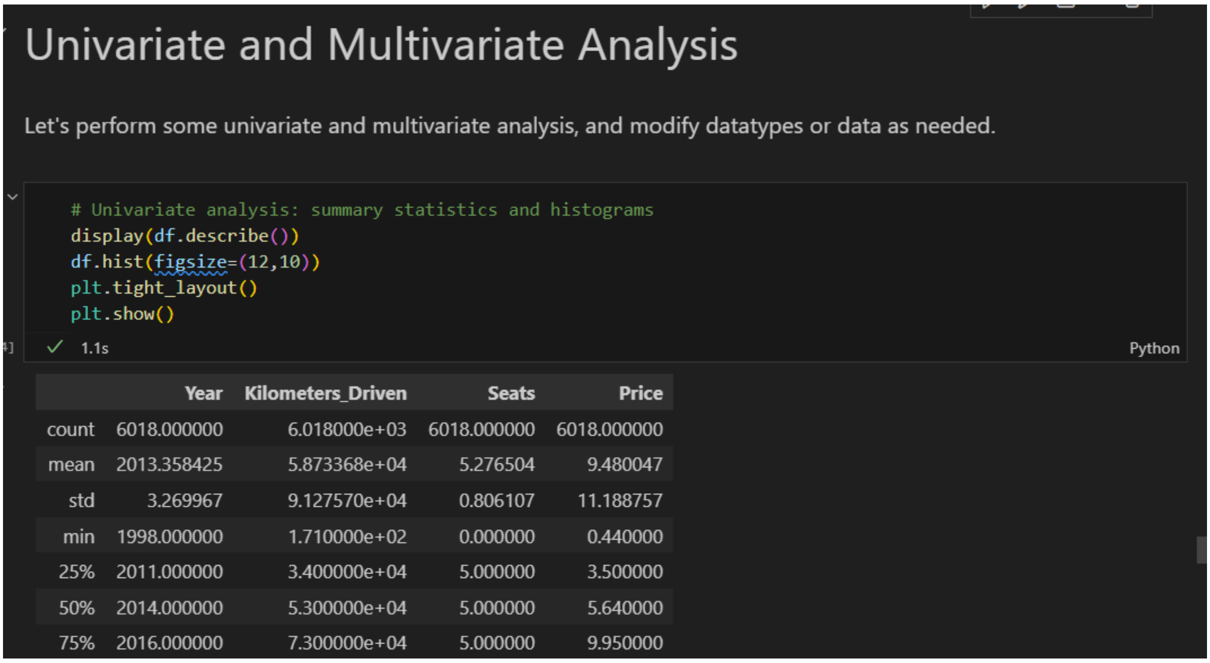
✓ 0.0s Python

	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Car_Age
0	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75	15
1	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50	10
2	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50	14
4	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74	12
5	Hyderabad	2012	75000	LPG	Manual	First	21.1 km/kg	814 CC	55.2 bhp	5.0	NaN	2.35	13



```
# Plot bar graph for car_bin distribution
df['car_bin'].value_counts().sort_index().plot(kind='bar', color='skyblue')
plt.title('Distribution of Cars by Age Bin')
plt.xlabel('Car Age Bin')
plt.ylabel('Number of Cars')
plt.xticks(rotation=0)
plt.show()
```





Univariate and Multivariate Analysis

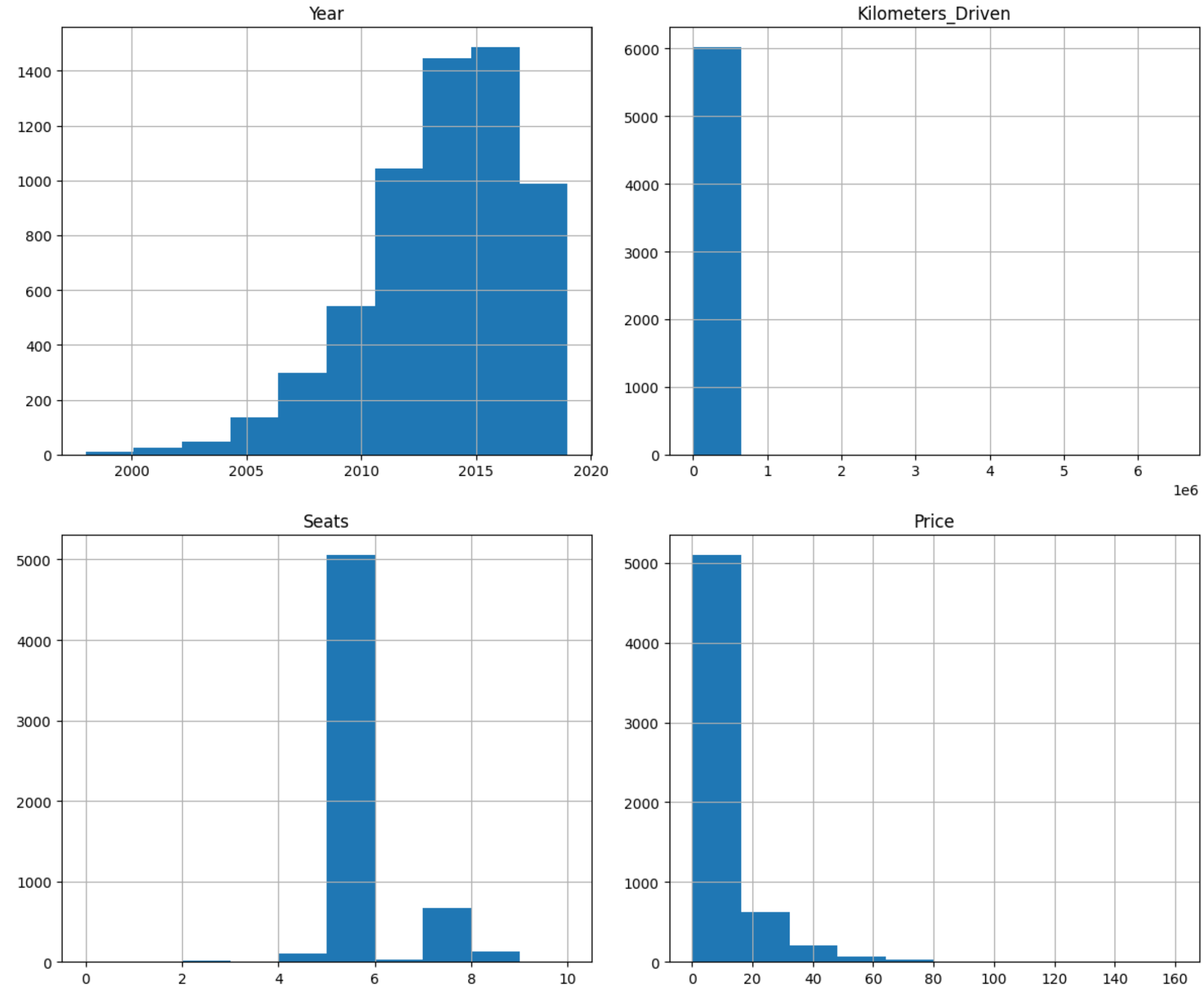
Let's perform some univariate and multivariate analysis, and modify datatypes or data as needed.

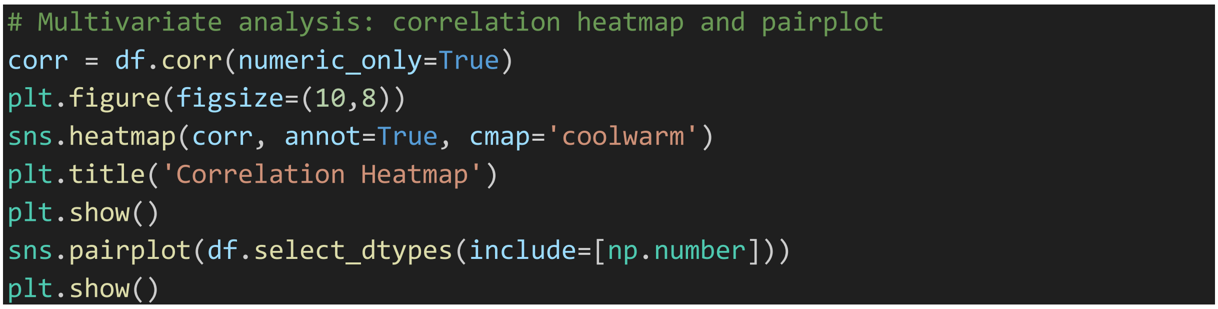
```python
# Univariate analysis: summary statistics and histograms
display(df.describe())
df.hist(figsize=(12,10))
plt.tight_layout()
plt.show()
```

✓ 1.1s

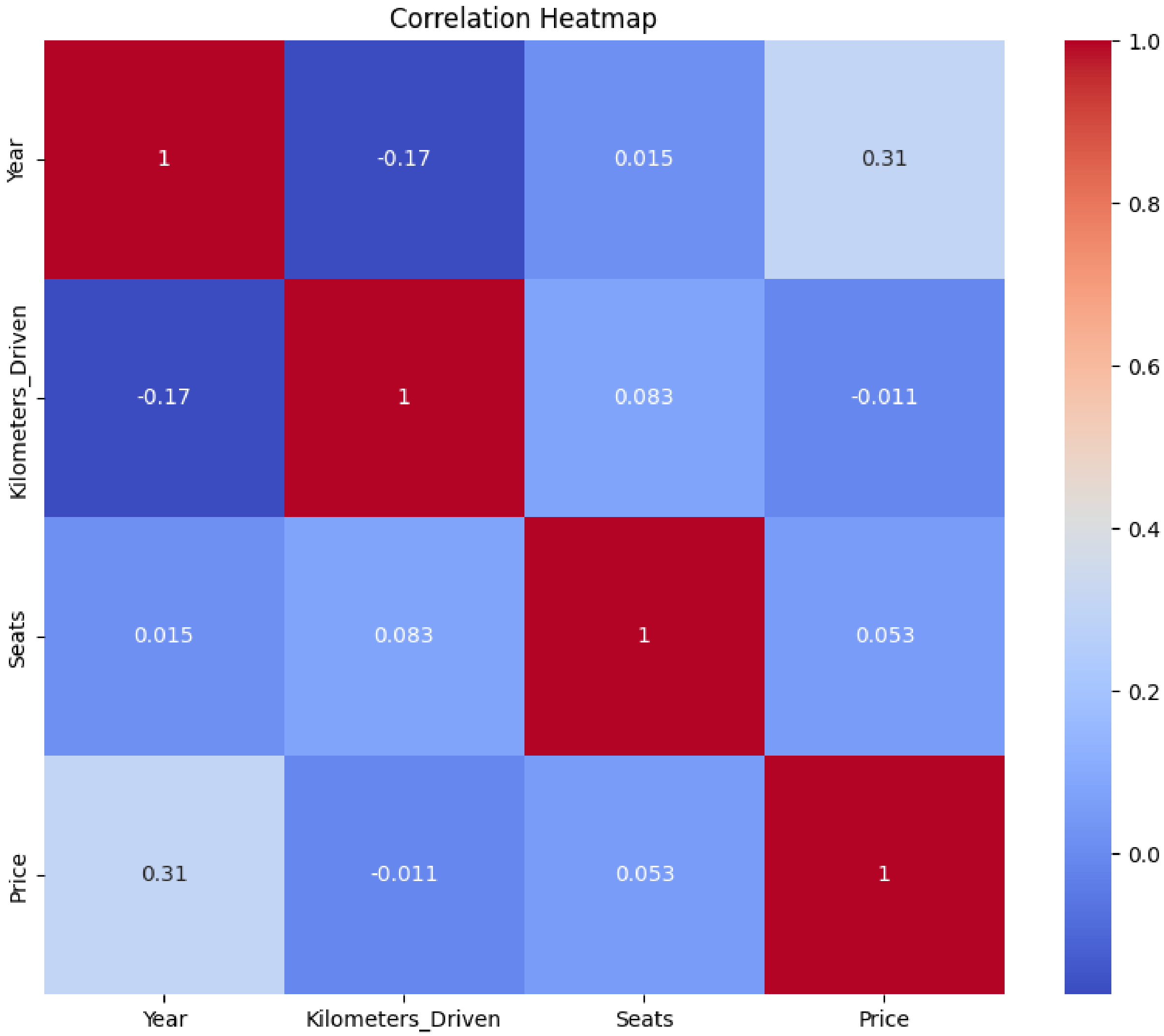
Python

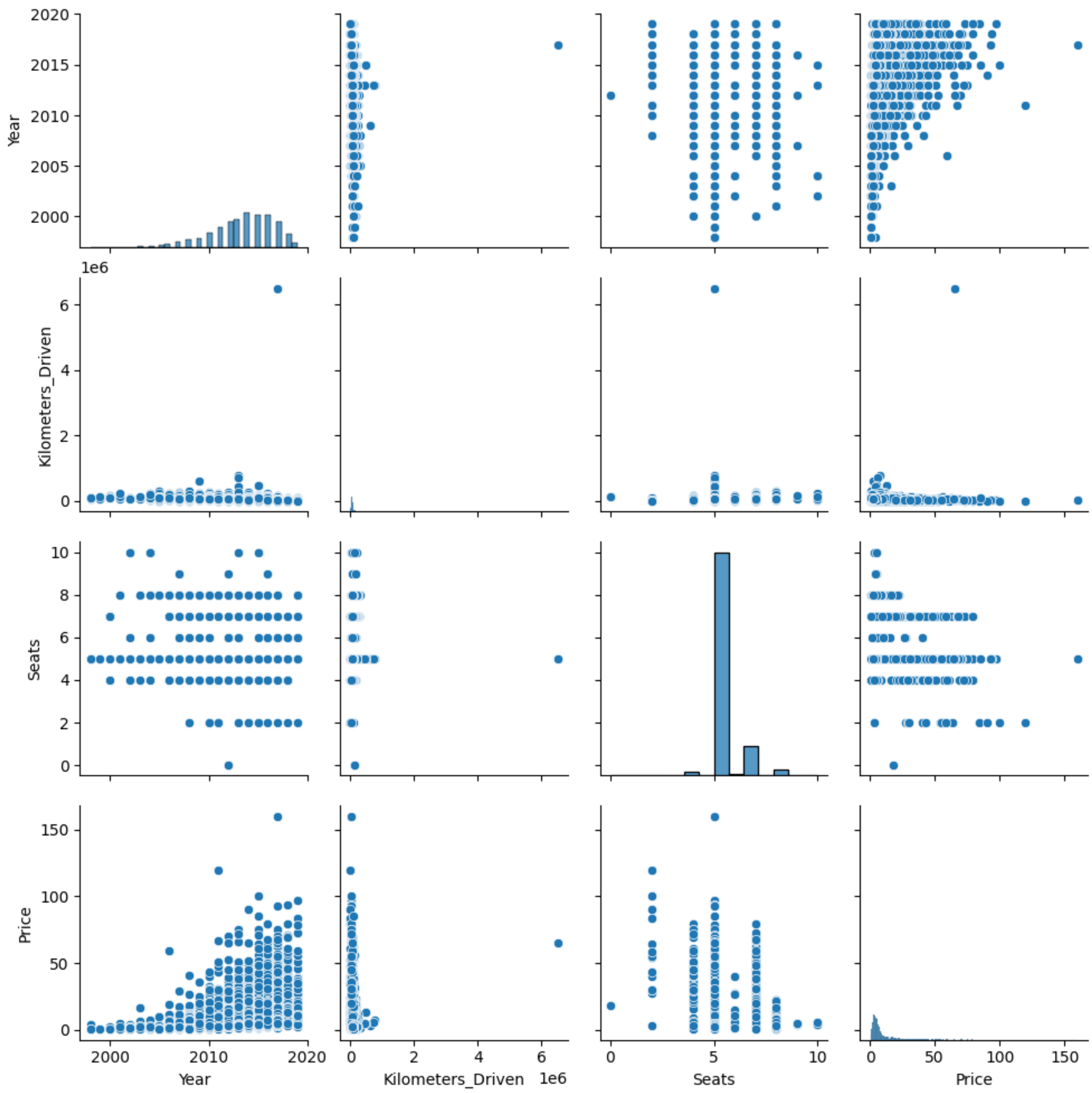
```
              Year  Kilometers_Driven         Seats        Price
count  6018.000000       6.018000e+03   6018.000000  6018.000000
mean   2013.358425       5.873368e+04      5.276504     9.480047
std       3.269967       9.127570e+04      0.806107    11.188757
min    1998.000000       1.710000e+02      0.000000     0.440000
25%    2011.000000       3.400000e+04      5.000000     3.500000
50%    2014.000000       5.300000e+04      5.000000     5.640000
75%    2016.000000       7.300000e+04      5.000000     9.950000
```





```
# Multivariate analysis: correlation heatmap and pairplot
corr = df.corr(numeric_only=True)
plt.figure(figsize=(10,8))
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
sns.pairplot(df.select_dtypes(include=[np.number]))
plt.show()
```




Year
Kilometers_Driven
Seats
Price
1e6
2020
2015
2010
2005
2000
0
2
4
6
8
10
50
100
150
2000
2010
2020
Year
Kilometers_Driven
Seats
Price