Experiment:- 2

Student Name:Jitesh Kumar **UID: 20BCS2334**
Branch: CSE **Section/Group: 20BCS_WM_903/A**
Semester: 5th **Date of Performance: Aug. 17, 2022**
Subject Code: 20CSP-317
Subject Name: MACHINE LEARNING LAB

Aim/Overview of the practical:

Implement Data Visualization.

Task to be done:

To perform Data Visualization on any standard dataset.

Apparatus/Simulator used:

- Jupyter Notebook/Google Collab
- Python
- pandas Library
- seaborn Library
- Standard Dataset

Code and Output:

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
cars_data=pd.read_csv('Toyota.csv',index_col=0,na_values=["??","????"])
```

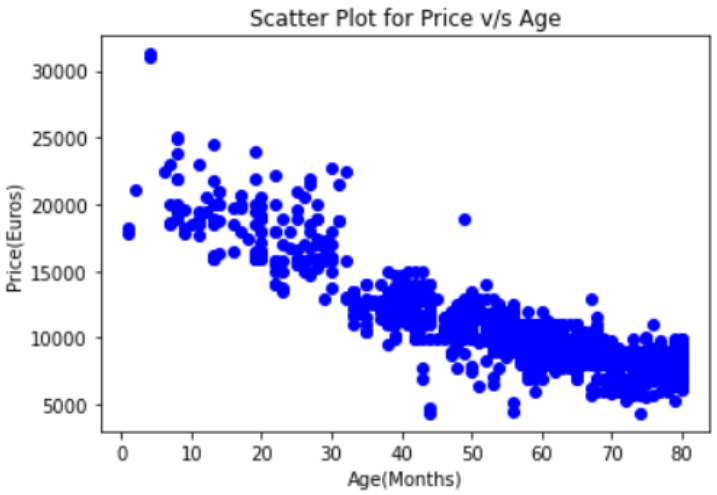
In [3]:
```
cars_data
```

Out[3]:

	Price	Age	KM	FuelType	HP	MetColor	Automatic	CC	Doors	Weight
0	13500	23.0	46986.0	Diesel	90.0	1.0	0	2000	three	1165
1	13750	23.0	72937.0	Diesel	90.0	1.0	0	2000	3	1165
2	13950	24.0	41711.0	Diesel	90.0	NaN	0	2000	3	1165
3	14950	26.0	48000.0	Diesel	90.0	0.0	0	2000	3	1165
4	13750	30.0	38500.0	Diesel	90.0	0.0	0	2000	3	1170
...	...	...	...	...	...	...	...	...	...	...
1431	7500	NaN	20544.0	Petrol	86.0	1.0	0	1300	3	1025
1432	10845	72.0	NaN	Petrol	86.0	0.0	0	1300	3	1015
1433	8500	NaN	17016.0	Petrol	86.0	0.0	0	1300	3	1015
1434	7250	70.0	NaN	NaN	86.0	1.0	0	1300	3	1015
1435	6950	76.0	1.0	Petrol	110.0	0.0	0	1600	5	1114

1436 rows × 10 columns

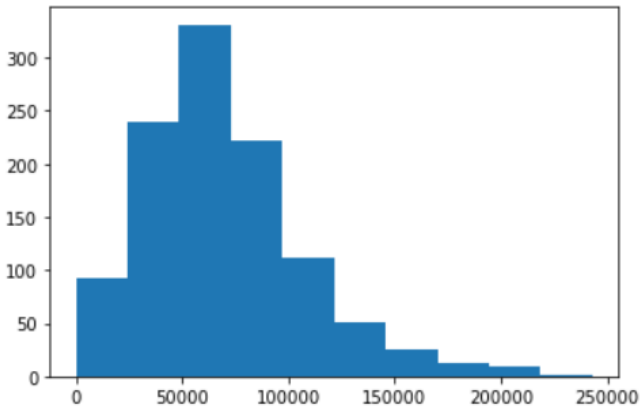
In [6]:
```
plt.scatter(cars_data['Age'],cars_data['Price'],c='blue')
plt.title('Scatter Plot for Price v/s Age')
plt.xlabel("Age(Months)")
plt.ylabel("Price(Euros)")
plt.show()
```



Plotted Histogram

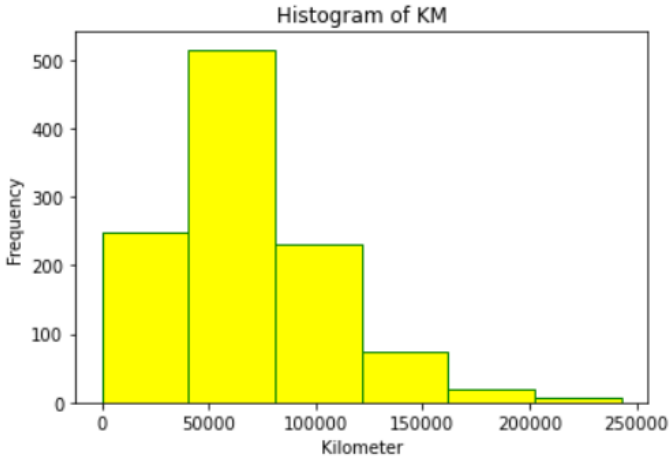
In [7]:
```
plt.hist(cars_data['KM'])
```

Out[7]:
```
(array([ 92., 239., 331., 222., 111.,  51.,  25.,  13.,  10.,   2.]),
 array([1.000000e+00, 2.430090e+04, 4.860080e+04, 7.290070e+04,
        9.720060e+04, 1.215005e+05, 1.458004e+05, 1.701003e+05,
        1.944002e+05, 2.187001e+05, 2.430000e+05]),
 <BarContainer object of 10 artists>)
```

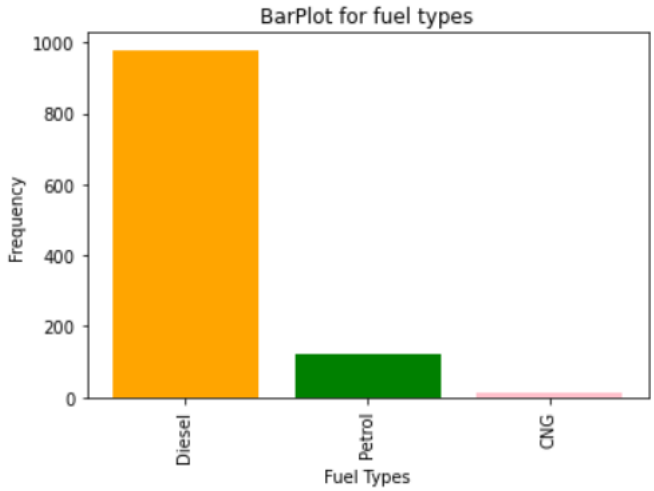


In [9]:
```
plt.hist(cars_data['KM'],color='yellow',edgecolor='green',bins=6)
plt.title('Histogram of KM')
plt.xlabel('Kilometer')
plt.ylabel('Frequency')
```

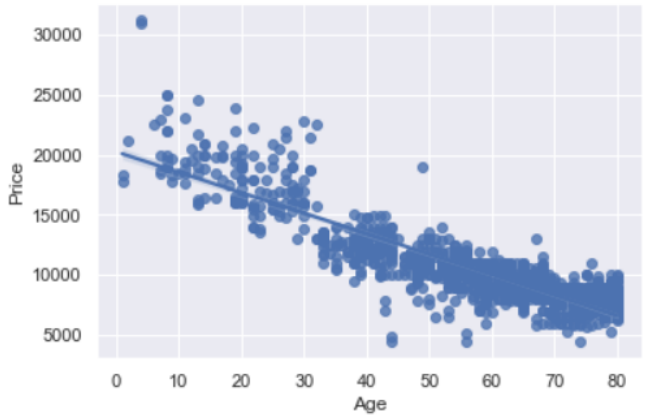
Out[9]:
```
Text(0, 0.5, 'Frequency')
```



In [15]:
```
plt.bar(index,counts,color=['orange','green','pink'])
plt.title('BarPlot for fuel types')
plt.xlabel('Fuel Types')
plt.ylabel('Frequency')
plt.xticks(index,FuelType,rotation=90)
plt.show()
```

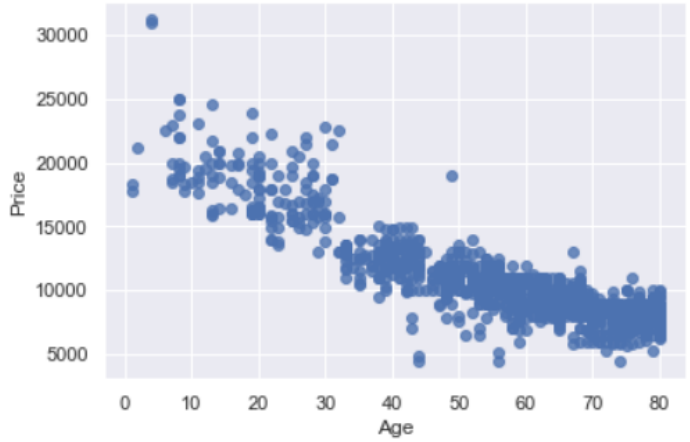


```
In [16]: import seaborn as sns

In [17]: sns.set(style='darkgrid')

In [18]: sns.regplot(x=cars_data['Age'],y=cars_data['Price'])
Out[18]: <AxesSubplot:xlabel='Age', ylabel='Price'>
```



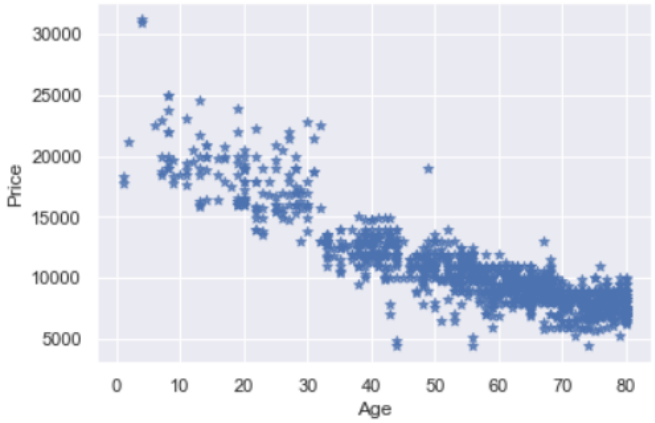
To remove the regression line we have to make it false as its default value is true.

```
In [19]: sns.regplot(x=cars_data['Age'],y=cars_data['Price'],fit_reg=False)
Out[19]: <AxesSubplot:xlabel='Age', ylabel='Price'>
```



In [20]:
```
sns.regplot(x=cars_data['Age'],y=cars_data['Price'],fit_reg=False,marker="*")
```

Out[20]:
```
<AxesSubplot:xlabel='Age', ylabel='Price'>
```

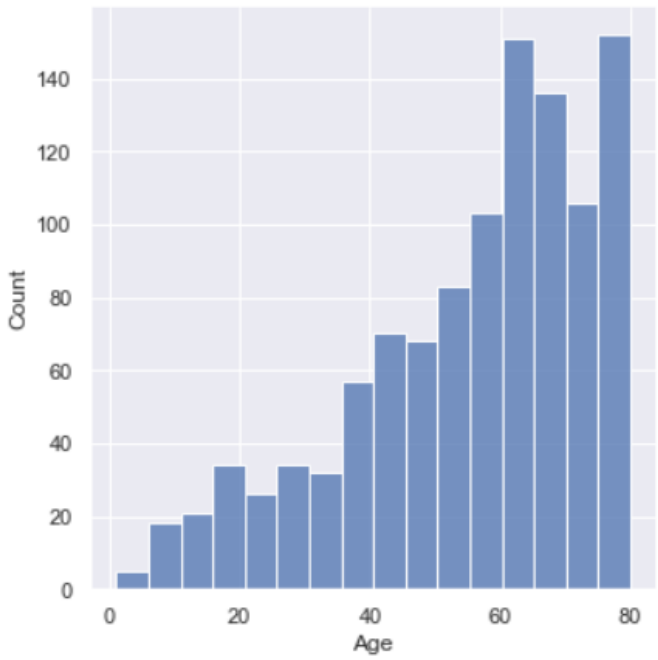


In [22]:
```
sns.lmplot(x='Age',y='Price',data=cars_data,fit_reg=False,hue='FuelType',legend=True,palette="Set1")
```

Out[22]:
```
<seaborn.axisgrid.FacetGrid at 0x21e37793b20>
```

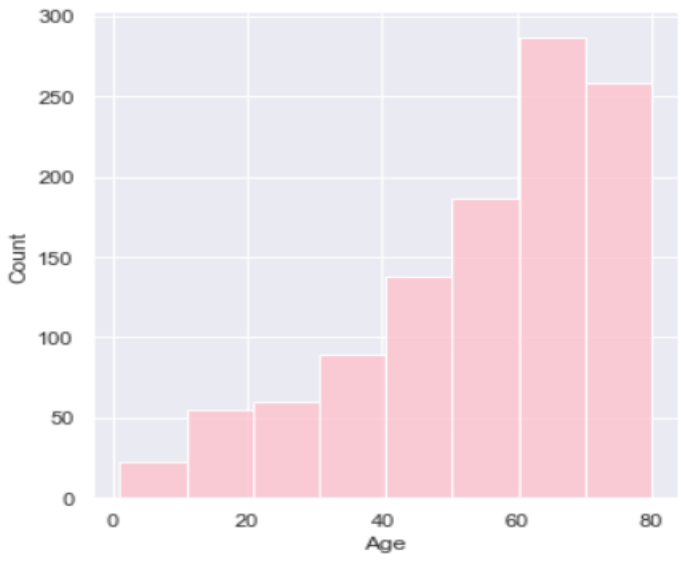
In [24]:
```
sns.displot(cars_data['Age'])
```

Out[24]:
```
<seaborn.axisgrid.FacetGrid at 0x21e3a4481f0>
```



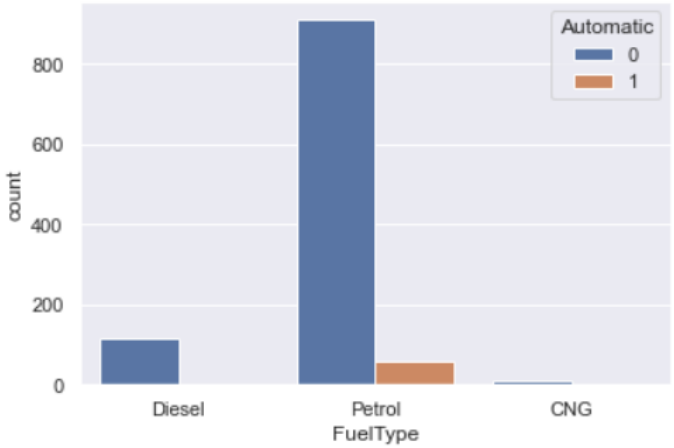
In [25]:
```
sns.displot(cars_data['Age'],color='pink',bins=8)
```

Out[25]:
```
<seaborn.axisgrid.FacetGrid at 0x21e3778e670>
```



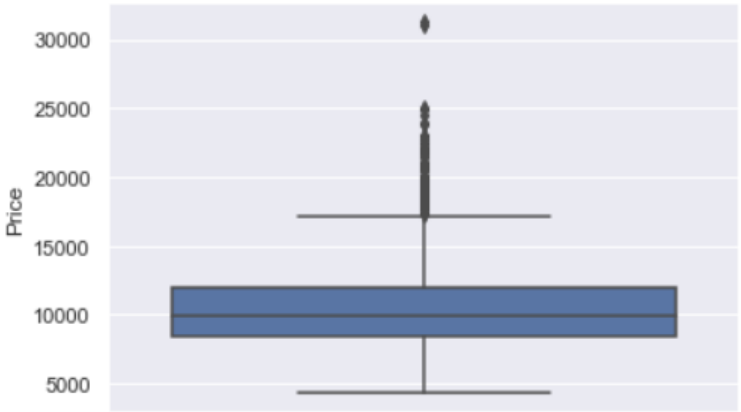
In [26]:
```
sns.countplot(x='FuelType',data=cars_data,hue="Automatic")
```

Out[26]: <AxesSubplot:xlabel='FuelType', ylabel='count'>



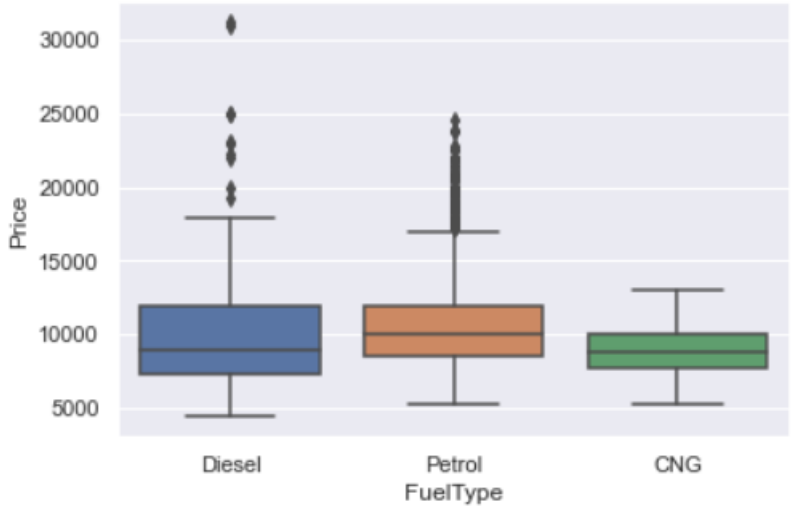
In [27]:
```
sns.boxplot(y=cars_data['Price'])
```

Out[27]: <AxesSubplot:ylabel='Price'>



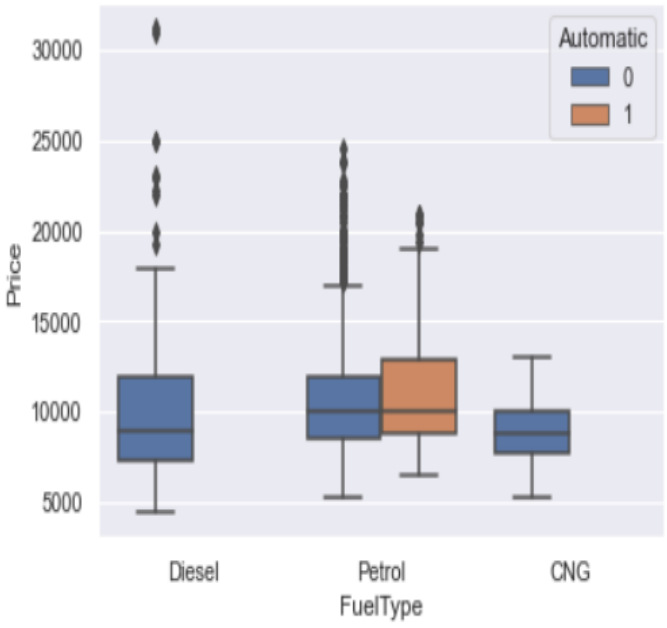
In [28]:
```
sns.boxplot(x=cars_data['FuelType'],y=cars_data['Price'])
```

Out[28]:
```
<AxesSubplot:xlabel='FuelType', ylabel='Price'>
```



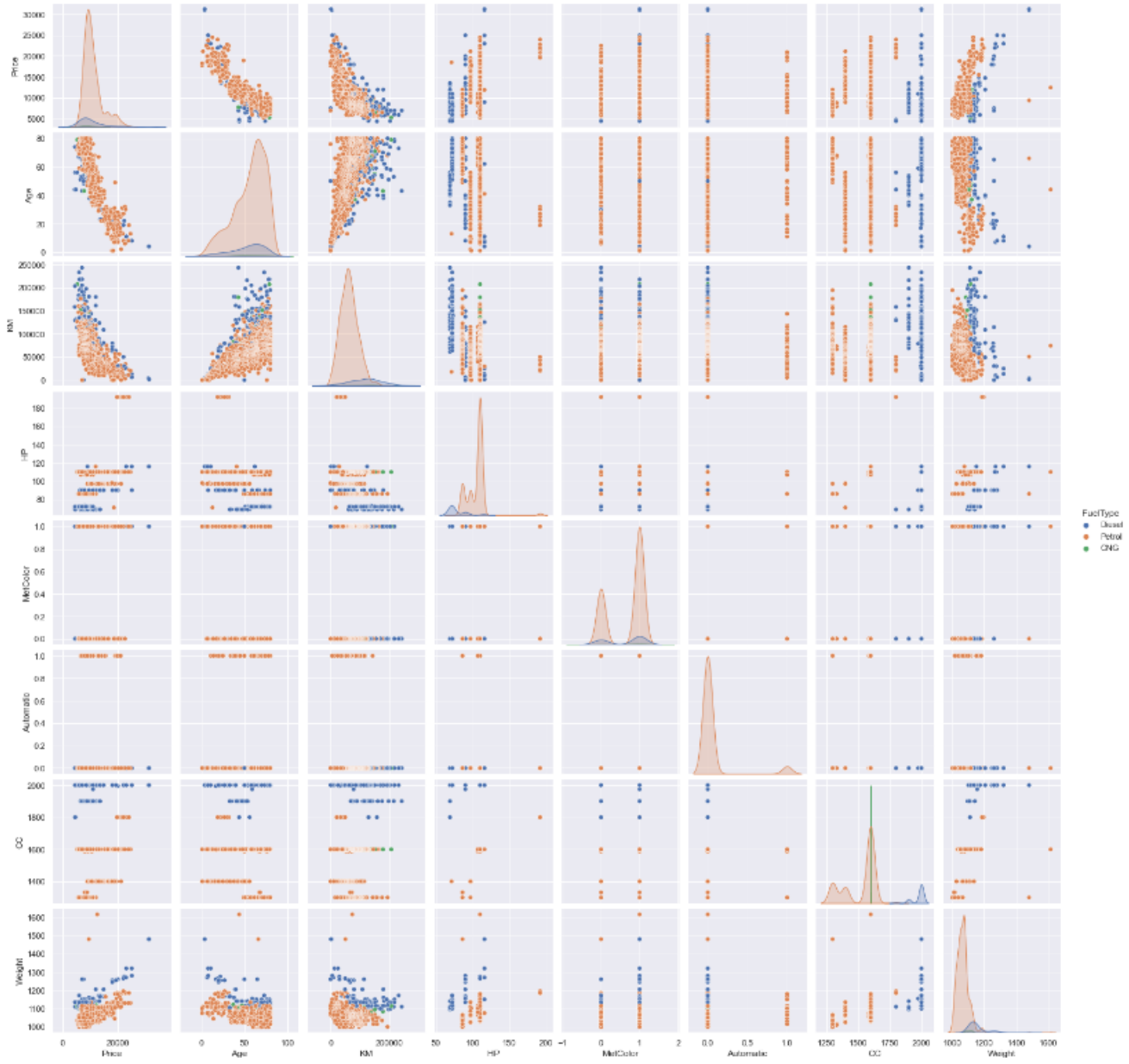
In [29]:
```
sns.boxplot(x='FuelType',y=cars_data['Price'],hue="Automatic",data=cars_data)
```

Out[29]:
```
<AxesSubplot:xlabel='FuelType', ylabel='Price'>
```



In [30]:
```
sns.pairplot(cars_data,kind="scatter",hue="FuelType")
```

Out[30]: <seaborn.axisgrid.PairGrid at 0x21e3a6cdd60>



Learning outcomes (What I have learnt):

1. To understand Data Visualization.
2. Learn about pandas', matplotlib and seaborn library/package ofpython.
3. Learn about the different methods/functions that are needed togenerate different types of graphs, charts and plots of the given dataset.
4. Leaned about regression line, KDE.