Experiment:- 1

Student Name: Jitesh Kumar
UID: 20BCS2334
Branch: CSE
Section/Group: 20BCS_WM_903/A
Semester: 5th
Date of Performance: Aug. 10,2022

Subject Code: 20CSP-317
Subject Name: MACHINE LEARNING LAB

1.1 Aim/Overview of the practical: Exploratory Data Analysis on any data set.

1.2. Task to be done: Perform EDA on any given data set.

1.3 Apparatus/Simulator used:

- Jupyter Notebook
- Python
- Pandas Library
- Data Set

1.3. Code:

```
import pandas as pd
import numpy as np
cars_data=pd.read_csv('Toyota.csv',index_col=0,na_values=["??","????"])
```

a. cars_data

```
In [8]: cars_data
```

Out[8]:

	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

b. pd.crosstab(index=cars_data['FuelType'],columns='count',dropna=True)

```
In [10]: pd.crosstab(index=cars_data['FuelType'],columns='count',dropna=True)
```

Out[10]:

FuelType	col_0 count
CNG	15
Diesel	144
Petrol	1177

c. pd.crosstab(index=cars_data['Automatic'],columns=cars_data['FuelType'],dropna=True)

```
In [12]: pd.crosstab(index=cars_data['Automatic'],columns=cars_data['FuelType'],dropna=True)
```

Out[12]:

Automatic	FuelType CNG	Diesel	Petrol
0	15	144	1104
1	0	0	73

d. pd.crosstab(index=cars_data['Automatic'],columns=cars_data['FuelType'],normalize=True,dropna=True)

```
In [13]: pd.crosstab(index=cars_data['Automatic'],columns=cars_data['FuelType'],normalize=True,dropna=True)
```

Out[13]:

Automatic	FuelType CNG	Diesel	Petrol
0	0.011228	0.107784	0.826347
1	0.000000	0.000000	0.054641

```
In [14]: num_data=cars_data.select_dtypes(exclude=[object])

In [15]: corr_matrix=num_data.corr()

In [16]: corr_matrix
```

Out[16]:

	Price	Age	KM	HP	MetColor	Automatic	CC	Weight
Price	1.000000	-0.878407	-0.574720	0.309902	0.112041	0.033081	0.165067	0.581198
Age	-0.878407	1.000000	0.512735	-0.157904	-0.099659	0.032573	-0.120706	-0.464299
KM	-0.574720	0.512735	1.000000	-0.335285	-0.093825	-0.081248	0.299993	-0.026271
HP	0.309902	-0.157904	-0.335285	1.000000	0.064749	0.013755	0.053758	0.086737
MetColor	0.112041	-0.099659	-0.093825	0.064749	1.000000	-0.013973	0.029189	0.057142
Automatic	0.033081	0.032573	-0.081248	0.013755	-0.013973	1.000000	-0.069321	0.057249
CC	0.165067	-0.120706	0.299993	0.053758	0.029189	-0.069321	1.000000	0.651450
Weight	0.581198	-0.464299	-0.026271	0.086737	0.057142	0.057249	0.651450	1.000000

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.