

Sem : MSc ICT 3rd Sem

Name :- Reshamwala Vivek Yogeshkumar

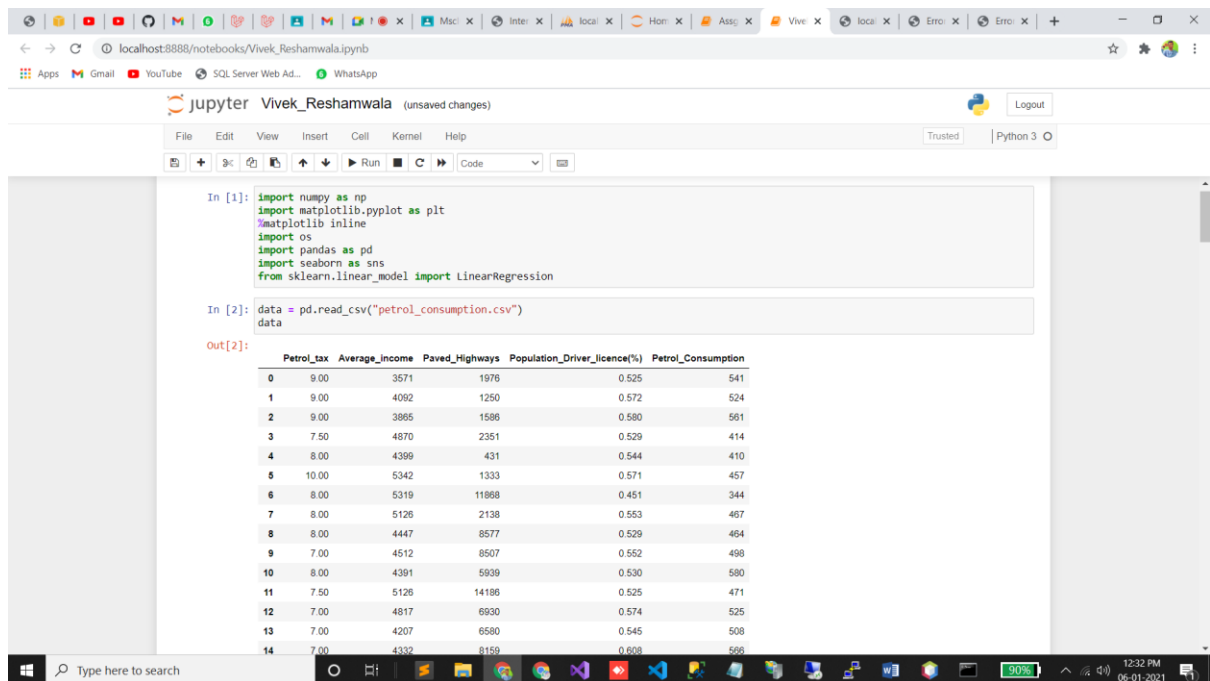
Roll No :- 52

Q1. Data Science

30 marks

Predict the the petrol consumption , when given the details average income, paved highways, petrol tax and population data in the given dataset.

Perform exploratory data analysis (EDA) and linear regression. Print the intercep and coeffients. Plot any two graphs.



The screenshot shows a Jupyter Notebook titled 'Vivek_Reshamwala' with two input cells. The first cell contains import statements for numpy, matplotlib, os, pandas, seaborn, and sklearn. The second cell loads a CSV file named 'petrol_consumption.csv'. The output of the second cell is a preview of the dataset, showing columns: Petrol_tax, Average_income, Paved_Highways, Population_Driver_licence(%), and Petrol_Consumption. The data is displayed as a table with 15 rows (index 0 to 14).

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import matplotlib inline
import os
import pandas as pd
import seaborn as sns
from sklearn.linear_model import LinearRegression

In [2]: data = pd.read_csv("petrol_consumption.csv")
data
```

	Petrol_tax	Average_income	Paved_Highways	Population_Driver_licence(%)	Petrol_Consumption
0	9.00	3571	1976	0.525	541
1	9.00	4092	1250	0.572	524
2	9.00	3865	1586	0.580	561
3	7.50	4870	2351	0.529	414
4	8.00	4399	431	0.544	410
5	10.00	5342	1333	0.571	457
6	8.00	5319	11868	0.451	344
7	8.00	5126	2138	0.553	487
8	8.00	4447	8577	0.529	464
9	7.00	4512	8507	0.552	498
10	8.00	4391	5939	0.530	580
11	7.50	5126	14186	0.525	471
12	7.00	4817	6930	0.574	525
13	7.00	4207	6580	0.545	508
14	7.00	4332	8159	0.608	566

```
localhost:8888/notebooks/Vivek_Reshamwala.ipynb
jupyter Vivek_Reshamwala (unsaved changes)
Python 3

In [3]: print(data.head())
print(data.tail())

    Petrol_tax  Average_income  Paved_Highways  Population_Driver_licence(%) \
0           9.0           3571           1976                0.525
1           9.0           4092           1250                0.572
2           9.0           3865           1586                0.580
3           7.5           4870           2351                0.529
4           8.0           4399           431                0.544

    Petrol_consumption
0           541
1           524
2           561
3           414
4           410

    Petrol_tax  Average_income  Paved_Highways  Population_Driver_licence(%) \
43          7.0           3745           2611                0.508
44          6.0           5215           2302                0.672
45          9.0           4476           3942                0.571
46          7.0           4296           4083                0.623
47          7.0           5002           9794                0.593

    Petrol_consumption
43          591
44          782
45          510
46          610
47          524

In [4]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48 entries, 0 to 47
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---
```

```
localhost:8888/notebooks/Vivek_Reshamwala.ipynb
jupyter Vivek_Reshamwala (unsaved changes)
Python 3

In [4]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48 entries, 0 to 47
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---

```
0 Petrol_tax 48 non-null float64
1 Average_income 48 non-null int64
2 Paved_Highways 48 non-null int64
3 Population_Driver_licence(%) 48 non-null float64
4 Petrol_consumption 48 non-null int64
dtypes: float64(2), int64(3)
memory usage: 2.0 KB

In [5]: print(data.count())

 Petrol_tax 48
 Average_income 48
 Paved_Highways 48
 Population_Driver_licence(%) 48
 Petrol_consumption 48
 dtype: int64

In [6]: data.isna().sum()

Out[6]: Petrol_tax 0
 Average_income 0
 Paved_Highways 0
 Population_Driver_licence(%) 0
 Petrol_consumption 0
 dtype: int64
```


```

localhost:8888/notebooks/Vivek_Reshamwala.ipynb

jupyter Vivek_Reshamwala (unsaved changes)

File Edit View Insert Cell Kernel Help

Python 3

```
In [4]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48 entries, 0 to 47
Data columns (total 5 columns):
 #   Column              Non-Null Count  Dtype  
---  --
 0   Petrol_tax           48 non-null     float64
 1   Average_income       48 non-null     int64  
 2   Paved_Highways       48 non-null     int64  
 3   Population_Driver_licence(%) 48 non-null     float64
 4   Petrol_Consumption   48 non-null     int64  
dtypes: float64(2), int64(3)
memory usage: 2.0 KB

In [5]: print(data.count())

Petrol_tax           48
Average_income       48
Paved_Highways       48
Population_Driver_licence(%) 48
Petrol_Consumption   48
dtype: int64

In [6]: data.isna().sum()

Out[6]: Petrol_tax           0
Average_income       0
Paved_Highways       0
Population_Driver_licence(%) 0
Petrol_Consumption   0
dtype: int64
```

localhost:8888/notebooks/Vivek_Reshamwala.ipynb

jupyter Vivek_Reshamwala (unsaved changes)

File Edit View Insert Cell Kernel Help

Python 3

```
In [13]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48 entries, 0 to 47
Data columns (total 5 columns):
 #   Column              Non-Null Count  Dtype  
---  --
 0   Petrol_tax           48 non-null     float64
 1   Average_income       48 non-null     int64  
 2   Paved_Highways       48 non-null     int64  
 3   Population_Driver_licence(%) 48 non-null     float64
 4   Petrol_Consumption   48 non-null     int64  
dtypes: float64(2), int64(3)
memory usage: 2.0 KB

In [14]: data.isna().sum()

Out[14]: Petrol_tax           0
Average_income       0
Paved_Highways       0
Population_Driver_licence(%) 0
Petrol_Consumption   0
dtype: int64

In [15]: data.describe()

Out[15]:
```

	Petrol_tax	Average_income	Paved_Highways	Population_Driver_licence(%)	Petrol_Consumption
count	48.000000	48.000000	48.000000	48.000000	48.000000
mean	7.668333	4241.833333	5565.416667	0.570333	576.770833
std	0.950770	573.623788	3491.507166	0.055470	111.885816
min	5.000000	3063.000000	431.000000	0.451000	344.000000
25%	7.000000	3739.000000	3110.250000	0.529750	509.500000

