

```
import numpy as num
import pandas as pan
```

```
# Loading the dataset
data = pan.read_csv('datafile_02.csv')
print(data.columns)
data.head()
```

```
Index(['Port', 'Traffic in Eleventh Plan (MT) (2011-12)Proj.',
      'Traffic in Eleventh Plan (MT) (2011-12) Ach.',
      'Traffic in Eleventh Plan (MT) (2011-12) %',
      'Total Capacity in Eleventh Plan (MT) (2011-12) Proj.',
      'Total Capacity in Eleventh Plan (MT) (2011-12) Ach.',
      'Total Capacity in Eleventh Plan (MT) (2011-12) %'],
      dtype='object')
```

	Port	Traffic in Eleventh Plan (MT) (2011- 12)Proj.	Traffic in Eleventh Plan (MT) (2011- 12) Ach.	Traffic in Eleventh Plan (MT) (2011- 12) %	Total Capacity in Eleventh Plan (MT) (2011-12) Proj.	Total Capacity in Eleventh Plan (MT) (2011-12) Ach.	Total Capacity in Eleventh Plan (MT) (2011-12) %
0	Kolkata	1343	1223	9100	3145	1635	5100
1	Haldia	4450	3101	7000	6340	5070	7900
2	Paradeep	7640	5425	7100	10640	7650	7100
3	Visakhapatnam	8220	6742	8200	10810	7293	6700
4	Ennore	4700	1496	3200	6420	3100	4800



```
# Preprocessing the dataset
```

```
# Renaming the columns
data.rename(columns = {'Traffic in Eleventh Plan (MT) (2011-12)Proj.': 'Traffic_Pr  
data
```

	Port	Traffic_Projected	Traffic_Achieved	Traffic in Eleventh Plan (MT) (2011- 12) %	Total_Capacity_Pro
0	Kolkata	1343	1223	9100	
1	Haldia	4450	3101	7000	
2	Paradeep	7640	5425	7100	
3	Visakhapatnam	8220	6742	8200	
4	Ennore	4700	1496	3200	
5	Chennai	5750	5571	9700	
6	Tuticorin	3172	2810	8900	
7	Cochin	3817	2010	5300	
8	NMPT	4881	3294	6800	
9	..	1155	8888	8888	

Perparing the Calculations:

```
Traffic_Percent = round((data.Traffic_Achieved/data.Traffic_Projected)*100,2)
```

```
Traffic_Percent
```

```
0    91.06
1    69.69
2    71.01
3    82.02
4    31.83
5    96.89
6    88.59
7    52.66
8    67.49
9    87.54
10   79.07
11   99.56
12   95.13
dtype: float64
```

```
Total_Percent = round( (data.Total_Capacity_Achieved/data.Total_Capacity_Projected)*100,2)
Total_Percent
```

```
0    51.99
1    79.97
2    71.90
3    67.47
4    48.29
5   110.26
6    52.11
```

```
7      74.85
8      84.25
9      62.63
10     48.45
11     66.95
12     71.12
dtype: float64
```

```
Total_Percent = round( (data.Total_Capacity_Achieved/data.Total_Capacity_Projected)*100,2)
Total_Percent
```

```
0      51.99
1      79.97
2      71.90
3      67.47
4      48.29
5     110.26
6      52.11
7      74.85
8      84.25
9      62.63
10     48.45
11     66.95
12     71.12
dtype: float64
```

```
# Replacing the existing columns with newly created columns
data.rename(columns = {'Traffic in Eleventh Plan (MT) (2011-12) %': 'Traffic_Percent'})
data.iloc[:,3:4] = Traffic_Percent
data.iloc[:,6:] = Total_Percent
data
```

	Port	Traffic_Projected	Traffic_Achieved	Traffic_Percent%	Total_Capa
0	Kolkata	1343	1223	91.06	

```
data.shape
```

```
(13, 7)
```

```
# Checking for null values
```

```
data.isnull().sum()
```

```
Port      0
Traffic_Projected    0
Traffic_Achieved     0
Traffic_Percent%     0
Total_Capacity_Projected  0
Total_Capacity_Achieved  0
Total_Percent%       0
dtype: int64
```

```
# Summary of Dataset
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13 entries, 0 to 12
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Port                  13 non-null    object
1   Traffic_Projected     13 non-null    int64
2   Traffic_Achieved      13 non-null    int64
3   Traffic_Percent%      13 non-null    float64
4   Total_Capacity_Projected 13 non-null    int64
5   Total_Capacity_Achieved 13 non-null    int64
6   Total_Percent%        13 non-null    float64
dtypes: float64(2), int64(4), object(1)
memory usage: 856.0+ bytes
```

```
data.describe()
```

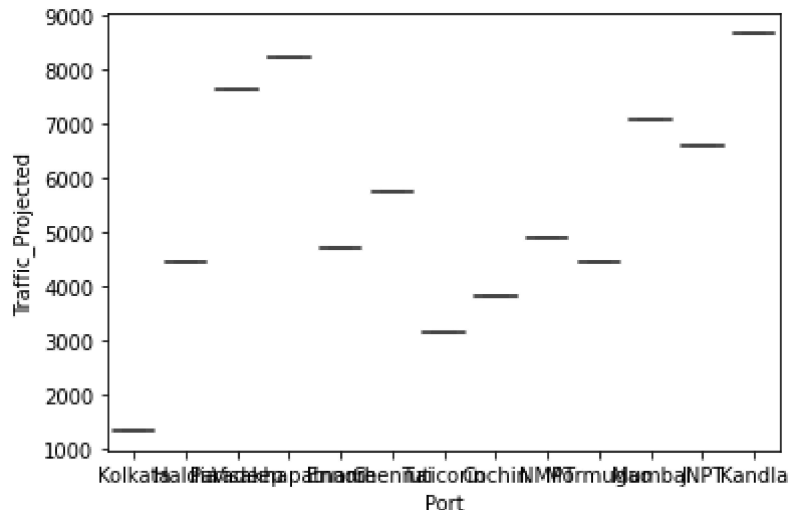
	Traffic_Projected	Traffic_Achieved	Traffic_Percent%	Total_Capacity_Projected
count	13.000000	13.000000	13.000000	13.000000
mean	5446.846154	4308.846154	77.887692	7705.307692

#Finding Outliers and replacing the outliers

```
import seaborn as sea
```

```
sea.boxplot(x='Port',y='Traffic_Projected',data=data)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fa28bcbf910>



Check For Categorical Columns and do encoding

```
from sklearn.preprocessing import LabelEncoder
```

```
la = LabelEncoder()
```

```
print(data.Port.value_counts())
```

```
data.Port = la.fit_transform(data.Port)
```

```
print(data.Port.value_counts())
```

```

Kolkata      1
Haldia       1
Paradeep     1
Visakhapatnam 1
Ennore       1
Chennai      1
Tuticorin    1
Cochin       1
NMPT         1
Mormugao     1
Mumbai       1
JNPT         1
Kandla       1
Name: Port, dtype: int64
6      1
3      1
10     1
12     1

```

```
2      1
0      1
11     1
1      1
9      1
7      1
8      1
4      1
5      1
```

Name: Port, dtype: int64

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 11:42

