```
In [1]:  import numpy as np
         import pandas as pd
```

```
In [4]:  df = pd.read_csv('datafile_02.csv')
         print(df.columns)
         df.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')
```

Out[4]:

	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

```
In [5]:  # Renaming the columns
         df.rename(columns = {'Traffic in Eleventh Plan (MT) (2011-12)Proj.':'Traffic_Projected'
         df
```

Out[5]:

	Port	Traffic_Projected	Traffic_Achieved	Traffic in Eleventh Plan (MT) (2011-12) %	Total_Capacity_Projected	Total_Capacity_
0	Kolkata	1343	1223	9100	3145	
1	Haldia	4450	3101	7000	6340	
2	Paradeep	7640	5425	7100	10640	
3	Visakhapatnam	8220	6742	8200	10810	
4	Ennore	4700	1496	3200	6420	
5	Chennai	5750	5571	9700	7230	
6	Tuticorin	3172	2810	8900	6398	
7	Cochin	3817	2010	5300	5475	

	Port	Traffic_Projected	Traffic_Achieved	Traffic in Eleventh Plan (MT) (2011-12) %	Total_Capacity_Projected	Total_Capacity_
8	NMPT	4881	3294	6800	6050	
9	Mormugao	4455	3900	8800	6690	
10	Mumbai	7105	5618	7900	9191	
11	JNPT	6604	6575	10000	9560	
12	Kandla	8672	8250	9500	12220	

In [6]:
```
# Perparing the Calculations:

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

Out[6]:
```
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
```

In [7]:
```
Total_Percent = round( (df.Total_Capacity_Achieved/df.Total_Capacity_Projected)*100,2)
Total_Percent
```

Out[7]:
```
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
```

In [8]:
```
# Replacing the existing columns with newly created columns
```

```
df.rename(columns = {'Traffic in Eleventh Plan (MT) (2011-12) %':'Traffic_Percent%','To
df.iloc[:,3:4] = Traffic_Percent
df.iloc[:,6:] = Total_Percent
df
```

Out[8]:

	Port	Traffic_Projected	Traffic_Achieved	Traffic_Percent%	Total_Capacity_Projected	Total_C
0	Kolkata	1343	1223	91.06	3145	
1	Haldia	4450	3101	69.69	6340	
2	Paradeep	7640	5425	71.01	10640	
3	Visakhapatnam	8220	6742	82.02	10810	
4	Ennore	4700	1496	31.83	6420	
5	Chennai	5750	5571	96.89	7230	
6	Tuticorin	3172	2810	88.59	6398	
7	Cochin	3817	2010	52.66	5475	
8	NMPT	4881	3294	67.49	6050	
9	Mormugao	4455	3900	87.54	6690	
10	Mumbai	7105	5618	79.07	9191	
11	JNPT	6604	6575	99.56	9560	
12	Kandla	8672	8250	95.13	12220	

In [9]:
```
df.shape
```

Out[9]: (13, 7)

In [10]:
```
# Checking for null values

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

Out[10]:
```
Port                        0
Traffic_Projected           0
Traffic_Achieved            0
Traffic_Percent%            0
Total_Capacity_Projected    0
Total_Capacity_Achieved     0
Total_Percent%              0
dtype: int64
```

In [11]:
```
df.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
```

In [12]:
```python
df.describe()
```

Out[12]:

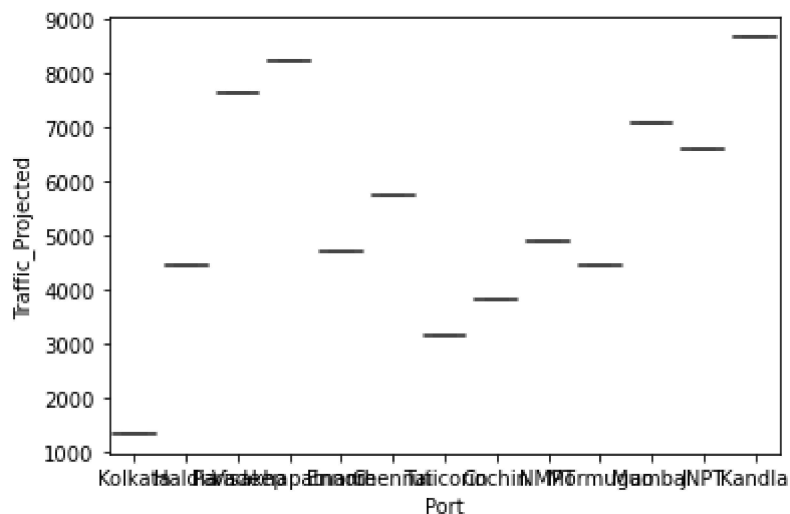|	Traffic_Projected	Traffic_Achieved	Traffic_Percent%	Total_Capacity_Projected	Total_Capacity_Achie
count	13.000000	13.000000	13.000000	13.000000	13.00
mean	5446.846154	4308.846154	77.887692	7705.307692	5306.38
std	2133.280019	2212.894855	19.382398	2570.242673	2140.25
min	1343.000000	1223.000000	31.830000	3145.000000	1635.00
25%	4450.000000	2810.000000	69.690000	6340.000000	4098.00
50%	4881.000000	3900.000000	82.020000	6690.000000	5070.00
75%	7105.000000	5618.000000	91.060000	9560.000000	7293.00
max	8672.000000	8250.000000	99.560000	12220.000000	8691.00

In [13]:
```python
#Finding Outliers anr replacing the outliers

import seaborn as sns
sns.boxplot(x='Port',y='Traffic_Projected',data=df)
```

Out[13]: <AxesSubplot:xlabel='Port', ylabel='Traffic_Projected'>

In [14]:
```python
# Check For Categorical Columns and do encoding

from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
print(df.Port.value_counts())
```

```python
df.Port = le.fit_transform(df.Port)
print(df.Port.value_counts())
```

```
Paradeep           1
Visakhapatnam      1
Kolkata            1
Mumbai             1
Kandla             1
Chennai            1
Ennore             1
Tuticorin          1
NMPT               1
Mormugao           1
Haldia             1
JNPT               1
Cochin             1
Name: Port, dtype: int64
0      1
1      1
2      1
3      1
4      1
5      1
6      1
7      1
8      1
9      1
10     1
11     1
12     1
Name: Port, dtype: int64
```

In [ ]: