COURSE NAME :DATA ANALYTICS WITH COGNOS

PROJECT: AIR QUALITY ANALYSIS IN TAMIL NADU.

In this phase let us see how to analysis the air quality with the help of air quality index in Tamil Nadu using Python. Air quality is analysis based on chemical pollutant quantity. By using machine learning, we can AQ(Air quality).
AQ: The air quality is an index for reporting air quality on a regular interval of time period. In other words, it is a measure of how air pollution affects one's health within a time period. The AQ is calculated based on the average concentration of a particular pollutant measured over a standard time interval. We can see how air pollution is by looking at the AQI.

Data Set Description

In This dataset file contains 11 attributes, of which 4 are chemical pollution quantities. The chemical quantites such as PM2.5 NO2, SO2 and RSPM/PM10 are independent attributes. This dataset shows the air qualities in various location of Tamil Nadu. Since air quality is calculated based on the attributes in the file. You can download the dataset https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014

The data is numeric and there are no missing values in the data, so no preprocessing is required. Our goal is to predict the air quality in Tamil Nadu, so this task is either Classification or regression. So as our class label is continuous, **regression** technique is required.

Regression is **supervised learning technique** that fits the data in a given range. Example Regression techniques in Python:

- Random Forest Regressor
- Ada Boost Regressor
- Bagging Regressor.
- Linear Regression etc.

Data preprocessing

```
In [3]: # Load the dataset
        df = pd.read_csv('cpcb_dly_aq_tamil_nadu-2014.csv')
        print(df.head())
```

```
   Stn Code Sampling Date       State City/Town/Village/Area  \
0        38      01-02-14  Tamil Nadu                Chennai
1        38      01-07-14  Tamil Nadu                Chennai
2        38      21-01-14  Tamil Nadu                Chennai
3        38      23-01-14  Tamil Nadu                Chennai
4        38      28-01-14  Tamil Nadu                Chennai

                      Location of Monitoring Station  \
0  Kathivakkam, Municipal Kalyana Mandapam, Chennai
1  Kathivakkam, Municipal Kalyana Mandapam, Chennai
2  Kathivakkam, Municipal Kalyana Mandapam, Chennai
3  Kathivakkam, Municipal Kalyana Mandapam, Chennai
4  Kathivakkam, Municipal Kalyana Mandapam, Chennai

                                  Agency Type of Location   SO2   NO2  \
0  Tamilnadu State Pollution Control Board  Industrial Area  11.0  17.0
1  Tamilnadu State Pollution Control Board  Industrial Area  13.0  17.0
2  Tamilnadu State Pollution Control Board  Industrial Area  12.0  18.0
3  Tamilnadu State Pollution Control Board  Industrial Area  15.0  16.0
4  Tamilnadu State Pollution Control Board  Industrial Area  13.0  14.0

   RSPM/PM10  PM 2.5
0       55.0     NaN
1       45.0     NaN
2       50.0     NaN
3       46.0     NaN
4       42.0     NaN
```

Load the dataset

Dataset link:https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014

```
In [1]: # importing pandas module for data frame
        import pandas as pd

        # loading dataset and storing in train variable
        train=pd.read_csv('cpcb_dly_aq_tamil_nadu-2014')

        # display top 5 data
        train.head()
```

Out[1]:

	Stn Code	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
0	38	01-02-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0	NaN
1	38	01-07-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0	NaN
2	38	21-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0	NaN
3	38	23-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	16.0	46.0	NaN
4	38	28-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0	NaN

DATA ANAYLSIS

1. STATISTICAL

```
In [4]: df.describe()
```

Out[4]:

	Stn Code	SO2	NO2	RSPM/PM10	PM 2.5
count	2879.000000	2868.000000	2866.000000	2875.000000	0.0
mean	475.750261	11.503138	22.136776	62.494261	NaN
std	277.675577	5.051702	7.128694	31.368745	NaN
min	38.000000	2.000000	5.000000	12.000000	NaN
25%	238.000000	8.000000	17.000000	41.000000	NaN
50%	366.000000	12.000000	22.000000	55.000000	NaN
75%	764.000000	15.000000	25.000000	78.000000	NaN
max	773.000000	49.000000	71.000000	269.000000	NaN

DATA VISUALIZATION

BAR CAHRT

Program:

```
In [13]: import pandas as pd
from matplotlib import pyplot as plt

# Read CSV into pandas
data = pd.read_excel("Book1.xlsx")
data.head()
df = pd.DataFrame(data)

City= df['City'].head(8)
SO2 = df['SO2'].head(8)

# Figure Size
fig = plt.figure(figsize =(10, 7))

# Horizontal Bar Plot
plt.bar(City[0:10], SO2[0:10])

# Show Plot
plt.show()
```

Output: