

## Phase 3: Development Part 1

### Aim:

To start building our project by loading and preprocessing the dataset.

### Loading of dataset:

Load air quality data set using python and data manipulation libraries like pandas.

### Preprocessing of Data:

It is the process of converting raw data into clean data. Preprocessing includes the following steps,

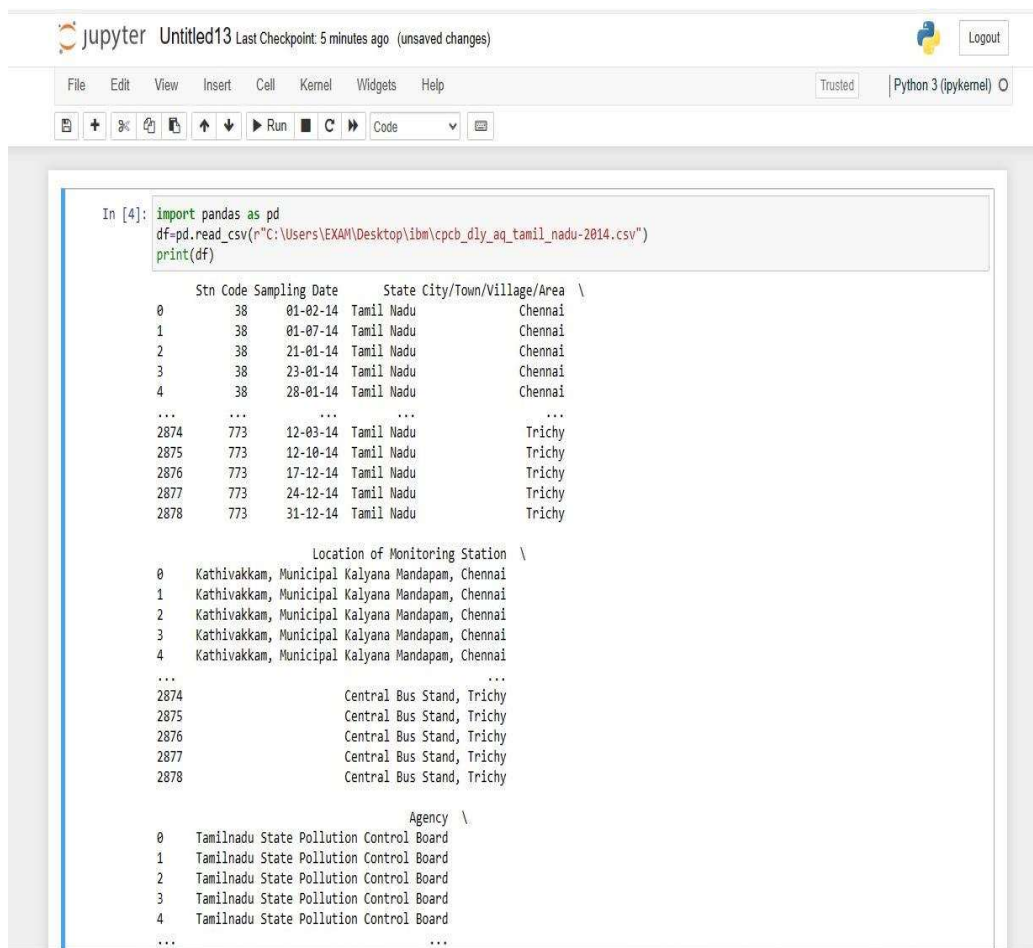
- => Drop columns that are not useful

- => Drop rows with missing values

- => Take care of missing data

Code:

```
#Loading dataset
import pandas as pd
df=pd.read_csv(r"C:\Users\Exam\Desktop\ibm\
cpcb_dly_aq_tamil_nadu-2014.csv")
print(df)
```



The screenshot shows a Jupyter Notebook interface with the following components:

- Header:** Jupyter logo, "Untitled13", "Last Checkpoint: 5 minutes ago (unsaved changes)", Python logo, and "Logout".
- Menu Bar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help.
- Toolbar:** Includes icons for file operations, a "Run" button, and a "Code" dropdown menu.
- Code Cell:** Contains the following code:

```
In [4]: import pandas as pd
df=pd.read_csv(r"C:\Users\EXAM\Desktop\ibm\cpcb_dly_aq_tamil_nadu-2014.csv")
print(df)
```
- Output:** The output of the code cell is a DataFrame with the following structure:

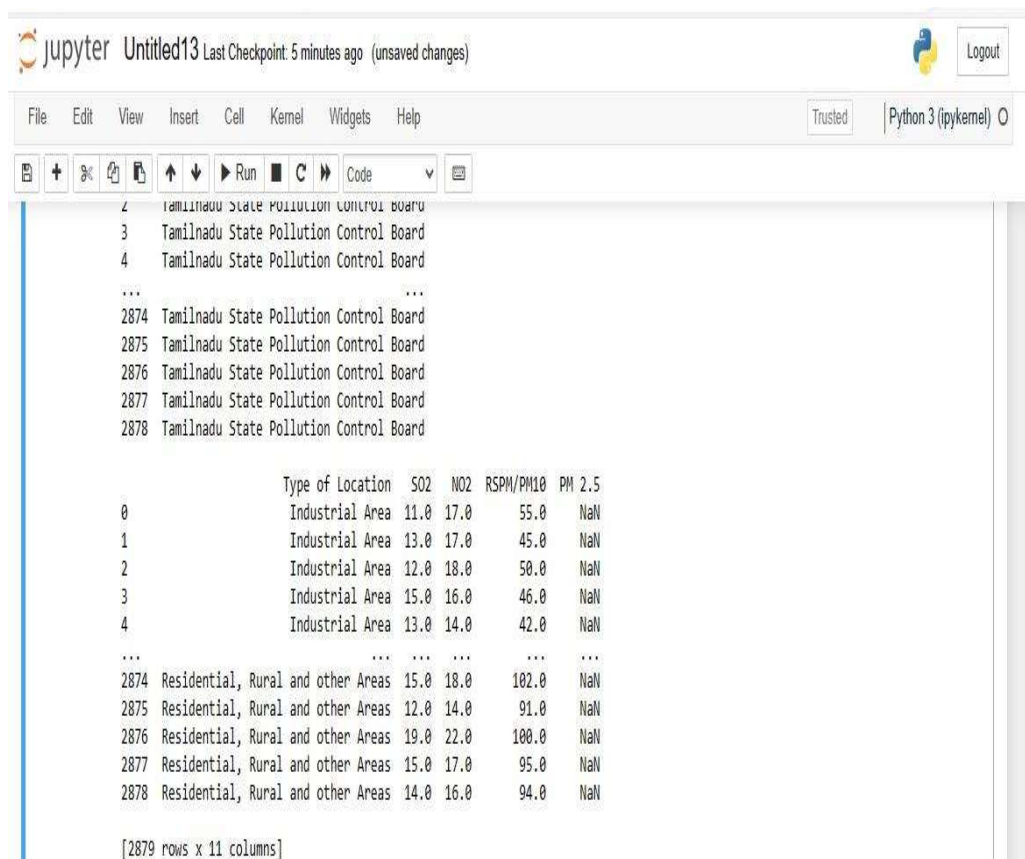
	Stn Code	Sampling Date	State	City/Town/Village/Area \
0	38	01-02-14	Tamil Nadu	Chennai
1	38	01-07-14	Tamil Nadu	Chennai
2	38	21-01-14	Tamil Nadu	Chennai
3	38	23-01-14	Tamil Nadu	Chennai
4	38	28-01-14	Tamil Nadu	Chennai
...	...	...	...	...
2874	773	12-03-14	Tamil Nadu	Trichy
2875	773	12-10-14	Tamil Nadu	Trichy
2876	773	17-12-14	Tamil Nadu	Trichy
2877	773	24-12-14	Tamil Nadu	Trichy
2878	773	31-12-14	Tamil Nadu	Trichy

	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
...	...
2874	Central Bus Stand, Trichy
2875	Central Bus Stand, Trichy
2876	Central Bus Stand, Trichy
2877	Central Bus Stand, Trichy
2878	Central Bus Stand, Trichy

	Agency \
0	Tamilnadu State Pollution Control Board
1	Tamilnadu State Pollution Control Board
2	Tamilnadu State Pollution Control Board
3	Tamilnadu State Pollution Control Board
4	Tamilnadu State Pollution Control Board
...	...



```
2 Tamilnadu State Pollution Control Board
3 Tamilnadu State Pollution Control Board
4 Tamilnadu State Pollution Control Board
...
2874 Tamilnadu State Pollution Control Board
2875 Tamilnadu State Pollution Control Board
2876 Tamilnadu State Pollution Control Board
2877 Tamilnadu State Pollution Control Board
2878 Tamilnadu State Pollution Control Board

      Type of Location  SO2  NO2  RSPM/PM10  PM 2.5
0      Industrial Area  11.0  17.0    55.0    NaN
1      Industrial Area  13.0  17.0    45.0    NaN
2      Industrial Area  12.0  18.0    50.0    NaN
3      Industrial Area  15.0  16.0    46.0    NaN
4      Industrial Area  13.0  14.0    42.0    NaN
...
2874 Residential, Rural and other Areas  15.0  18.0    102.0    NaN
2875 Residential, Rural and other Areas  12.0  14.0     91.0    NaN
2876 Residential, Rural and other Areas  19.0  22.0    100.0    NaN
2877 Residential, Rural and other Areas  15.0  17.0     95.0    NaN
2878 Residential, Rural and other Areas  14.0  16.0     94.0    NaN

[2879 rows x 11 columns]
```

## Preprocessing of dataset:

#Drop the useless columns

```
df=df.drop('PM 2.5',axis=1)
```

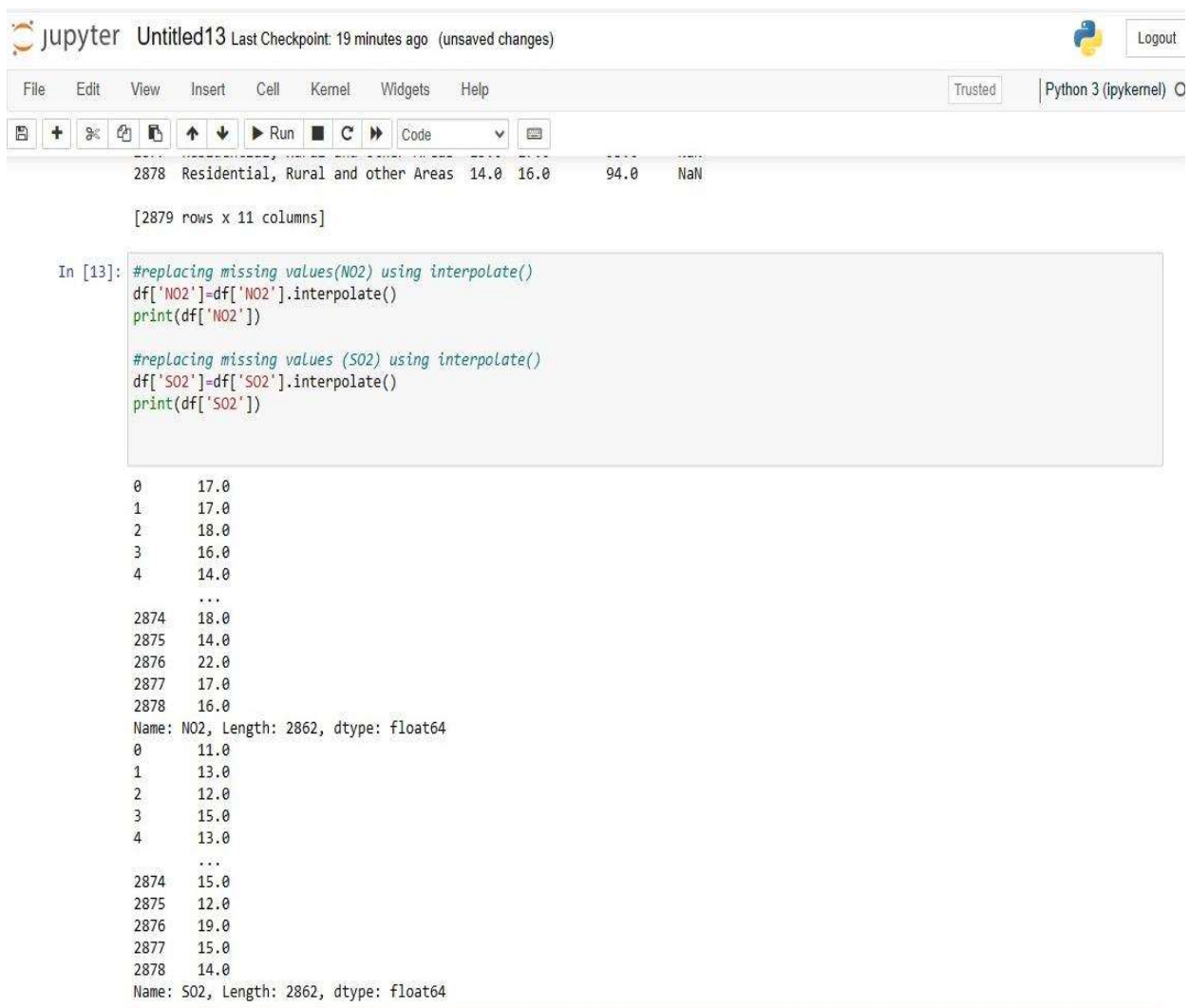
#Take care of missing data

```
df['NO2']=df['NO2'].interpolate()
```

```
print(df['NO2'])
```

```
df['SO2']=df['SO2'].interpolate()
print(df['SO2'])
```

```
#Drop the Null values
df=df.dropna()
```



The image shows a Jupyter Notebook interface with the following components:

- Header:** "jupyter Untitled13 Last Checkpoint: 19 minutes ago (unsaved changes)" and a "Logout" button.
- Menu Bar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help.
- Toolbar:** Includes icons for saving, adding cells, zooming, and running code.
- Code Cell:** Contains the following Python code:

```
In [13]: #replacing missing values(N02) using interpolate()
df['N02']=df['N02'].interpolate()
print(df['N02'])

#replacing missing values (S02) using interpolate()
df['S02']=df['S02'].interpolate()
print(df['S02'])
```
- Output:** Displays the result of the first print statement, showing a series of values for 'N02' (ranging from 14.0 to 18.0) and a second series of values for 'S02' (ranging from 11.0 to 15.0). The output is truncated with ellipses between the two series.