

PHASE 3

Loading and preprocessing the dataset

Printing the Intial structure of the Dataset

Stn Code	Sampling	State	City/Town	Location c	Agency	Type of Loc	SO2	NO2	RSPM/PM	SPM
38	5/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.6	17.16667	73.33333	149.6667
38	7/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.15	20.28333	61.33333	150.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.45	20.51667	75	114.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.75	18.18333	120	197.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.78	17.32	96.5	216
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	17.11667	20.86667	81.33333	159.3333
38	2/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.48333	20.66667	82.33333	185.6667
38	4/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.93333	18.65	79	179.3333
38	9/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.36667	19.5	96.66667	181.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.16667	18.38333	52.66667	125.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	18.63333	50.66667	128.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.56667	21.48333	85	157
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.55	19.78333	88.66667	272.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14.71667	22.33333	83	197.3333
38	2/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.633333	16.61667	94.66667	193.3333
38	4/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.38333	16.58333	118.3333	335.3333
38	9/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.7	22.13333	56	173.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	8.65	14.36667	53.33333	107.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.71667	19.36667	69.66667	167.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	16.6	111	298.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.41667	20.15	58	222.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15.35	20.5	67.66667	279.3333
38	1/4/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.316667	17.31667	88	253

air quality

A	B	C	D	E	F	G	H	I	J	K	L	M
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	16	19	117	NA		
773	5/6/2015	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	20	23	120	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	10	14	106	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	17	20	120	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	22	24	123	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	19	22	118	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	124	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	22	141	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	137	NA		
773	8/5/2015	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	23	141	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	23	144	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	134	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	20	136	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	NA	NA	102	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	22	142	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	11	18	80	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	8	17	116	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	24	121	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	22	118	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	NA	NA	104	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	24	132	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	25	147	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	24	138	NA		

air quality



8332	237	1/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	35	35	NA	
8333	237	5/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	30	48	NA	
8334	237	8/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	32	41	NA	
8335	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	26	41	NA	
8336	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	27	47	NA	
8337	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	45	NA	
8338	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	7	23	59	NA	
8339	237	2/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	29	55	NA	
8340	237	5/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	30	51	NA	
8341	237	9/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	34	35	NA	
8342	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	28	40	NA	
8343	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	40	NA	
8344	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	43	NA	
8345	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	27	78	NA	
8346	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	47	NA	
8347	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	33	37	NA	
8348	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	23	40	NA	
8349	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	26	47	NA	
8350	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	22	50	NA	
8351	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	24	43	NA	
8352	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	29	59	NA	
8353	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	25	46	NA	
8354	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	6	28	46	NA	
8355	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	28	45	NA	

air quality



Introduction to the attribute given in the datasets

TYPE OF PARAMETER GIVEN IN AIR QUALITY DATASETS OF TAMIL NADU:

Stn Code	Sampling	State	City/Town	Location c	Agency	Type of Loc	SO2	NO2	RSPM/PM	SPM
159	8/2/2010	Tamil Nad	Chennai	Madras M	National E	Residenti	3.166667	9.166667	80	
159	#####	Tamil Nad	Chennai	Madras M	National E	Residenti	4.666667	6.333333	51.66667	

1 .STN CODE: It refers to station code of the Tamil Nadu

District.

1. 0-150 : Unhealthy and sensitive
2. 150-200:unhealthy
3. 201-300: very unhealthy

1. 2. SAMPLING Date: It refers to the sampling date at which No2 and so2 measured.

3. STATE: It refers to the Tamil Nadu.

4. City/Town/village: List of village givens in the datasets.

1. Chennai
2. Coimbatore
3. Cuddalore
4. Madurai
5. Thoothukudi
6. Salem
7. Trichy
8. Mettur

LOCATION: It refers to the Location the so₂ and no₂ measured.

Agency: Refers to the Industry emitting No₂ and so₂

Type of Location: It refers to the type of location such resident and street .

SO₂: It refers to the sulphur dioxide.

NO2: It refers to the Nitrogen dioxide.

RSPM: Respirable suspended Particulate Matter.

SPM: Suspended Particulate Matter.

Loading the Initial structure of the Datasets without preprocessing

Stn Code	Sampling	State	City/Town	Location	Agency	Type of Loc	SO2	NO2	RSPM/PM	SPM
38	5/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.6	17.16667	73.33333	149.6667
38	7/1/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.15	20.28333	61.33333	150.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.45	20.51667	75	114.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.75	18.18333	120	197.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.78	17.32	96.5	216
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	17.11667	20.86667	81.33333	159.3333
38	2/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.48333	20.66667	82.33333	185.6667
38	4/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.93333	18.65	79	179.3333
38	9/2/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.36667	19.5	96.66667	181.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.16667	18.38333	52.66667	125.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	18.63333	50.66667	128.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.56667	21.48333	85	157
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.55	19.78333	88.66667	272.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14.71667	22.33333	83	197.3333
38	2/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.633333	16.61667	94.66667	193.3333
38	4/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.38333	16.58333	118.3333	335.3333
38	9/3/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12.7	22.13333	56	173.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	8.65	14.36667	53.33333	107.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.71667	19.36667	69.66667	167.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10.18333	16.6	111	298.6667
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11.41667	20.15	58	222.3333
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15.35	20.5	67.66667	279.3333
38	1/4/2010	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	9.316667	17.31667	88	252

air quality

A	B	C	D	E	F	G	H	I	J	K	L	M
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	16	19	117	NA		
773	5/6/2015	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	20	23	120	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	10	14	106	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	17	20	120	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	22	24	123	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	19	22	118	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	124	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	22	141	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	137	NA		
773	8/5/2015	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	23	141	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	23	144	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	20	134	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	20	136	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	NA	NA	102	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	22	142	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	11	18	80	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	8	17	116	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	24	121	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	13	22	118	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	NA	NA	104	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	14	24	132	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	25	147	NA		
773	#####	Tamil Nad	Trichy	Central Bu	Tamilnadu	Residenti	15	24	138	NA		

air quality



8332	237	1/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	35	35	NA	
8333	237	5/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	30	48	NA	
8334	237	8/8/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	32	41	NA	
8335	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	26	41	NA	
8336	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	27	47	NA	
8337	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	45	NA	
8338	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	7	23	59	NA	
8339	237	2/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	29	55	NA	
8340	237	5/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	30	51	NA	
8341	237	9/9/2014	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	34	35	NA	
8342	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	28	40	NA	
8343	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	40	NA	
8344	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	43	NA	
8345	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	27	78	NA	
8346	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	25	47	NA	
8347	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	33	37	NA	
8348	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	23	40	NA	
8349	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	26	47	NA	
8350	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	4	22	50	NA	
8351	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	24	43	NA	
8352	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	29	59	NA	
8353	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	25	46	NA	
8354	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	6	28	46	NA	
8355	237	#####	Tamil Nad	Coimbato	SIDCO Off	Tamilnadu	Industrial	5	28	45	NA	

air quality



Preprocessing of the datasets using python

Manipulation Libraries Using pandas and Other libraries:

```
import pandas as pd

# Load the dataset
df = pd.read_csv('air quality.csv')

# Display the initial structure of the dataset
print("Initial structure of the dataset:")
print(df.head())

# Handling missing values
# Drop rows with missing values
df.dropna(inplace=True)

# Removing duplicates
df.drop_duplicates(inplace=True)

# Dealing with outliers (you can use statistical methods like Z-score or IQR)
# Example: Removing outliers in a specific column

# Standardizing data formats (converting to appropriate data types if necessary)
# Example: Converting a column to a datetime format
df['Sampling Date'] = pd.to_datetime(df['Sampling Date'])

# After cleaning, you might want to save the cleaned dataset to a new file.
df.to_csv('cleaned_dataset.csv', index=False)
```


Initial structure of the dataset:

	Stn	Code	Sampling Date	State	City/Town/Village/Area \
0	38	5/1/2010	Tamil Nadu	Chennai	
1	38	7/1/2010	Tamil Nadu	Chennai	
2	38	12/1/2010	Tamil Nadu	Chennai	
3	38	1/19/2010	Tamil Nadu	Chennai	
4	38	1/21/2010	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	9.60	17.166667
1	Tamilnadu State Pollution Control Board	Industrial Area	11.15	20.283333
2	Tamilnadu State Pollution Control Board	Industrial Area	12.45	20.516667
3	Tamilnadu State Pollution Control Board	Industrial Area	10.75	18.183333
4	Tamilnadu State Pollution Control Board	Industrial Area	9.78	17.320000

	RSPM/PM10	SPM
0	73.333333	149.666667
1	61.333333	150.333333
2	75.000000	114.666667
3	120.000000	197.666667
4	96.500000	216.000000

Cleaned dataset:

	Stn	Code	Sampling Date	State	City/Town/Village/Area \
0	38	2010-05-01	Tamil Nadu	Chennai	
1	38	2010-07-01	Tamil Nadu	Chennai	
2	38	2010-12-01	Tamil Nadu	Chennai	
3	38	2010-01-19	Tamil Nadu	Chennai	
4	38	2010-01-21	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	9.60	17.166667
1	Tamilnadu State Pollution Control Board	Industrial Area	11.15	20.283333
2	Tamilnadu State Pollution Control Board	Industrial Area	12.45	20.516667
3	Tamilnadu State Pollution Control Board	Industrial Area	10.75	18.183333
4	Tamilnadu State Pollution Control Board	Industrial Area	9.78	17.320000

	RSPM/PM10	SPM
0	73.333333	149.666667
1	61.333333	150.333333
2	75.000000	114.666667
3	120.000000	197.666667
4	96.500000	216.000000

Preprocess Information after removing Null Values and NAN Values:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1811 entries, 0 to 12230
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Stn Code                             1811 non-null   int64
1   Sampling Date                       1811 non-null   datetime64[ns]
2   State                               1811 non-null   object
3   City/Town/Village/Area              1811 non-null   object
4   Location of Monitoring Station       1811 non-null   object
5   Agency                              1811 non-null   object
6   Type of Location                    1811 non-null   object
7   SO2                                 1811 non-null   float64
8   NO2                                 1811 non-null   float64
9   RSPM/PM10                          1811 non-null   float64
10  SPM                                 1811 non-null   float64
dtypes: datetime64[ns](1), float64(4), int64(1), object(5)
```