

AIR QUALITY ANALYSIS AND PREDICTION IN TAMILNADU

Dataset contains in features

- ▶ Stn_code : Station code. A code is given to each station that recorded the data.
- ▶ Sampling_date: The date when the data was recorded.
- ▶ State: It represents the states whose air quality data is measured.
- ▶ Location: It represents the city whose air quality data is measured.
- ▶ Agency: Name of the agency that measured the data.

DATA EXPLORATION

- ▶ Let us get some insights about the data – the number of entries in each column, the type of entry in each column
- ▶ It represents the type of area where the data was recorded like industrial, residential, etc.
- ▶ Let us see how many types of area
- ▶ NO, because the agency's name has nothing to do with how much polluted the state is. Similarly, stn_code is also unnecessary.
- ▶ It is given in the data description that date

CODE

- ▶ # Rows with missing “types”
- ▶ `null_data = tn[tn.isnull().any(axis=1)]`
- ▶ `null_data.head(20)`

OUTPUT

state	location	type	so2	no2	spm	date	
343962	Tamil Nadu	Tuticorin	NaN	10.2	16.7	194.0	1991-09-01
343984	Tamil Nadu	Tuticorin	NaN	7.2	16.4	82.0	1991-12-01
343985	Tamil Nadu	Tuticorin	NaN	8.2	5.4	47.0	1991-12-01
343986	Tamil Nadu	Tuticorin	NaN	8.0	12.4	44.0	1991-12-01
344142	Tamil Nadu	Tuticorin	NaN	3.7	17.8	80.0	1994-09-01
344143	Tamil Nadu	Tuticorin	NaN	6.4	20.9	68.0	1994-09-01
344147	Tamil Nadu	Tuticorin	NaN	3.7	17.8	80.0	1994-09-01
344148	Tamil Nadu	Tuticorin	NaN	6.4	20.9	68.0	1994-09-01
344155	Tamil Nadu	Coimbatore	NaN	6.0	16.0	55.0	1995-01-01
344156	Tamil Nadu	Coimbatore	NaN	0.9	9.2	44.0	1995-03-01
344157	Tamil Nadu	Coimbatore	NaN	5.1	15.8	78.0	1995-02-01
344158	Tamil Nadu	Coimbatore	NaN	3.9	13.5	71.0	1995-03-01
344159	Tamil Nadu	Coimbatore	NaN	2.3	9.1	56.0	1995-04-01
344160	Tamil Nadu	Coimbatore	NaN	2.2	5.5	42.0	1995-05-01
344161	Tamil Nadu	Coimbatore	NaN	1.4	4.9	31.0	1995-06-01
344162	Tamil Nadu	Coimbatore	NaN	2.3	6.9	29.0	1995-07-01
344163	Tamil Nadu	Coimbatore	NaN	3.7	8.4	51.0	1995-08-01
344164	Tamil Nadu	Coimbatore	NaN	2.3	5.3	40.0	1995-09-01
344165	Tamil Nadu	Coimbatore	NaN	3.1	5.9	33.0	1995-10-01
344166	Tamil Nadu	Coimbatore	NaN	5.1	6.5	41.0	1995-11-01

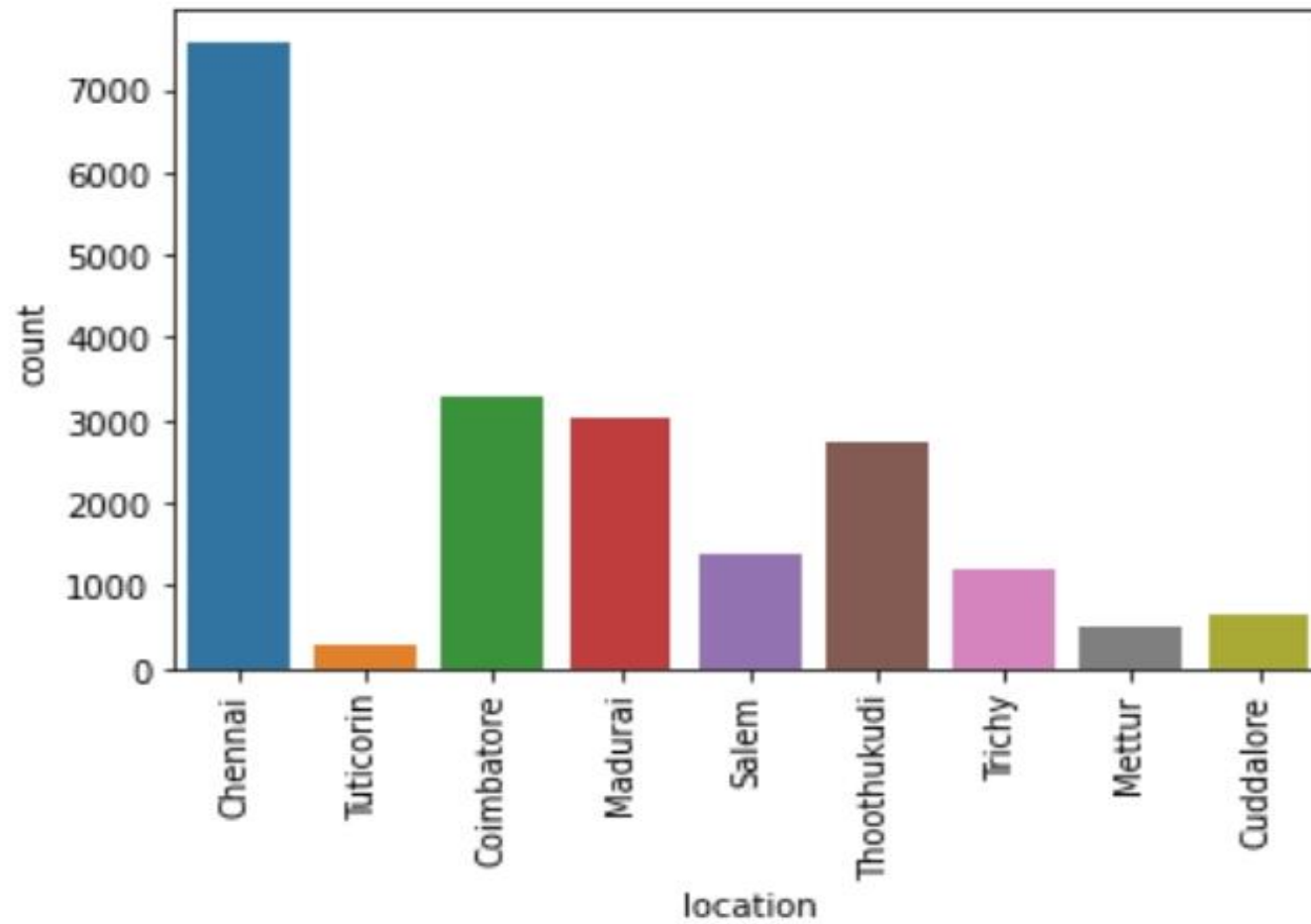
DATA VISUALIZATION

Code: data visualization “

Input:

```
datacount =sns.countplot(x ="location",data = tn);  
datacount.set_xticklabels(datacount.get_xticklabels(), rotation=90);
```

OUTPUT



CALCULATE IN AIR QUALITY

- ▶ The AQI is calculated by converting measured pollutant concentrations to a uniform index which is based on the health effects associated with a pollutant.
- ▶ The health benchmarks used for calculating the AQI are pollutant specific and are established by the EPA through the National Ambient Air Quality Standards.
- ▶ There are several ways to test indoor air quality, including the following:
Purchase an indoor air quality monitor. High-quality monitors will test for particulate matter, chemical pollutants, humidity, carbon monoxide and even formaldehyde.

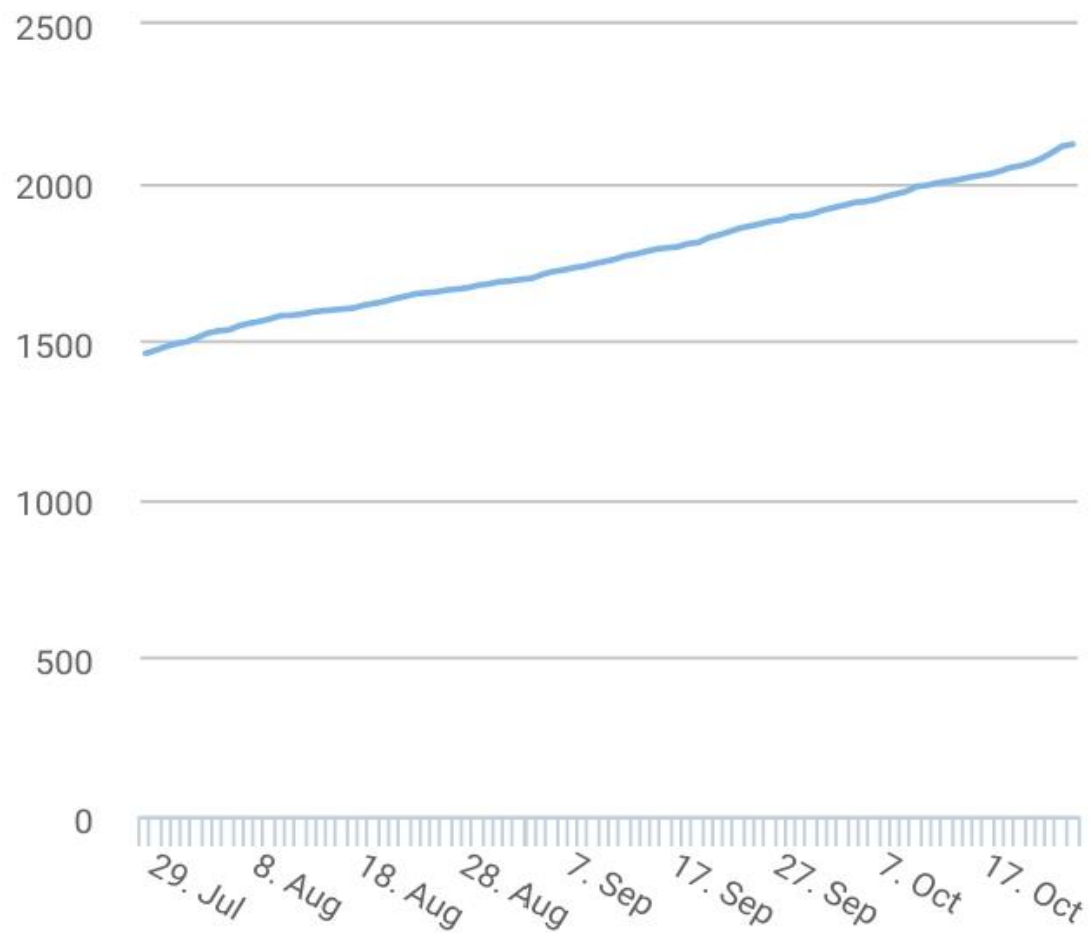
INPUT

```
▶ #Function to calculate so2 individual pollutant index(si)
▶ def calculate_si(so2):
▶     si=0
▶     if (so2<=40):
▶         si= so2*(50/40)
▶     if (so2>40 and so2<=80):
▶         si= 50+(so2-40)*(50/40)
▶     if (so2>80 and so2<=380):
▶         si= 100+(so2-80)*(100/300)
▶     if (so2>380 and so2<=800):
▶         si= 200+(so2-380)*(100/800)
▶     if (so2>800 and so2<=1600):
▶         si= 300+(so2-800)*(100/800)
▶     if (so2>1600):
▶         si= 400+(so2-1600)*(100/800)
▶     return si
▶ data['si']=data['so2'].apply(calculate_si)
▶ df= data[['so2','si']]
▶ df.head()
```

OUTPUT

►	So2	si
►	0	4.8 6.000
►	1	3.1 3.875
►	2	6.2 7.750
►	3	6.3 7.875
►	4	4.7 5.875

Article Metrics (level)



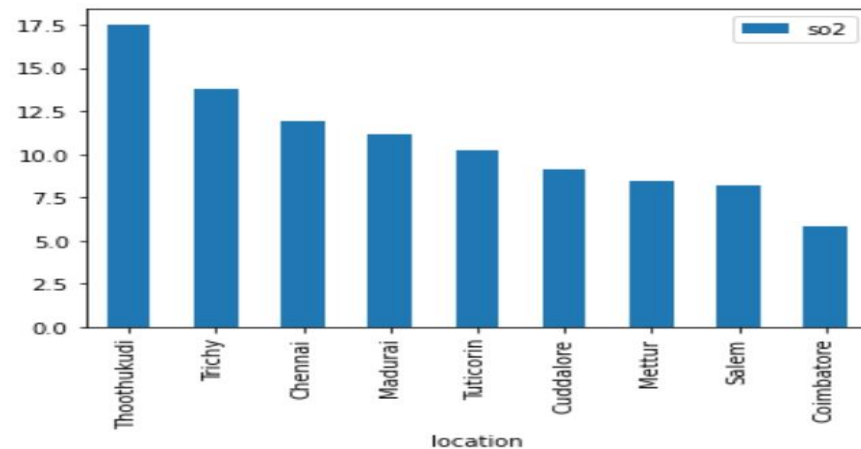
DISTRICT LEVEL PREDICTION

- ▶ INPUT : CODE

- ▶

```
maxso2 = loc.sort_values(by='so2',ascending=False)  
maxso2.loc[:,['so2']].head(10).plot(kind='bar'); # Based on average values
```

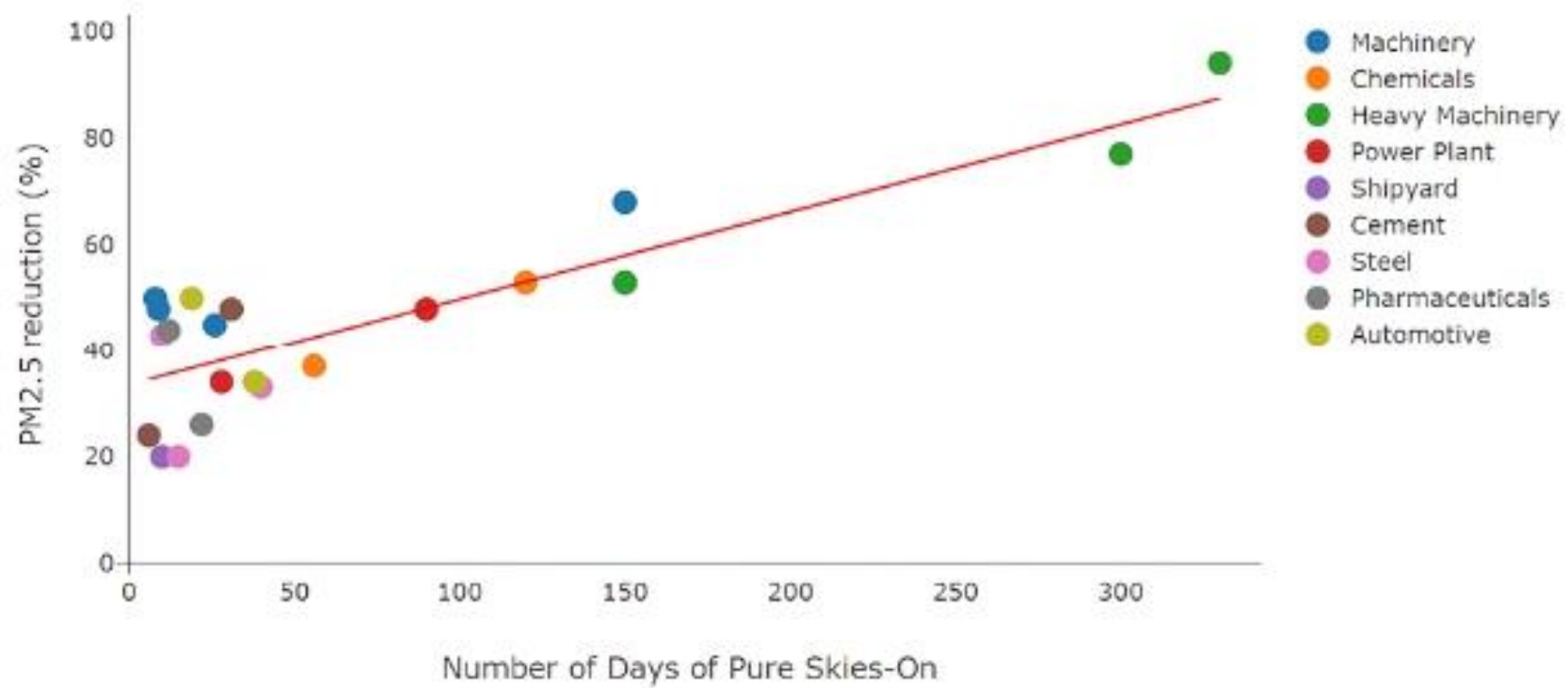
- ▶ OUTPUT :



Air quality visualization

- ▶ The graphical display of data - helps us understand the distribution of air pollutants in the atmosphere. This is hard to do just by looking at a modern air monitor equipment with its digital display.
- ▶
- ▶ By combining real-time monitoring data with python programming, one can easily visualize air monitoring data. Interactive graphs can be created which makes it easier to check air quality, and increasingly diverse colors can visually highlight the air quality level. Visualization of data has a resilient expression (more images and more insightful) than the original data table, which is favorable for further analysis of data.
- ▶
- ▶ In this article, I will share examples of data visualization that have helped Devic Earth to convey more clearly the message on air quality.

OUTPUT



- Thankyou