**AIR QUALITY ANALYSIS AND PREDICTION IN TAMILNADU USING MACHINE LEARNING**

**Project Title:** Air Quality Prediction

**Phase 3 :** Development Part 1

**Introduction :**

Air quality can be defined as the measurement of quality of the air we breathe and the concentrations of the pollutants in the air that can cause various health issues. Air quality can be used for various purposes such as the communication of air quality with the public, to plan strategies that can be used to reduce air pollution, and to monitor short term and long term trends.

Air quality can be measured using various machine learning algorithms. Many countries and their environmental agencies in the world use the AQI for the real time spreading of the information on the air quality. Although the basic concepts of air quality are similar, the practical implementations of each can differ. Applying AQIs on a common set of data can show large differences in the index values and concentration of pollutants.

**Given Dataset :**

****

**Necessary step to follow :**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.plotly as py

%matplotlib inline

plt.rcParams['figure.figsize'] = (10, 7)

import warnings

warnings.filterwarnings('ignore')

import os

print(os.listdir("../input"))

['data.csv']

data=pd.read\_csv('../input/data.csv',encoding = "ISO-8859-1")

data.head()

**Output :**

****

**Location Monitoring Station :**

data.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 435742 entries, 0 to 435741

Data columns (total 13 columns):

stn\_code 291665 non-null object

sampling\_date 435739 non-null object

state 435742 non-null object

location 435739 non-null object

agency 286261 non-null object

type 430349 non-null object

so2 401096 non-null float64

no2 419509 non-null float64

rspm 395520 non-null float64

spm 198355 non-null float64

location\_monitoring\_station 408251 non-null object

pm2\_5 9314 non-null float64

date 435735 non-null object

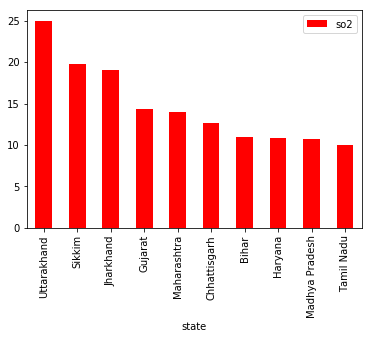
dtypes: float64(5), object(8)

memory usage: 43.2+ MB

**Measuring Places :**

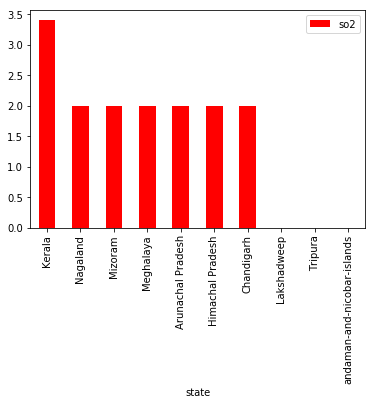
data[['so2','state']].groupby(["state"]).median().sort\_values(by='so2',ascending=False).head(10).plot.bar(color='r')

plt.show()



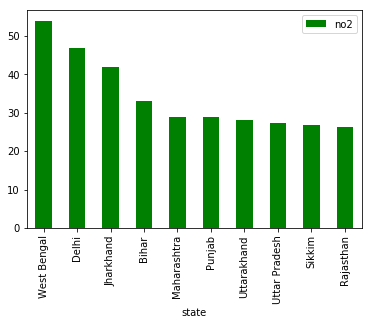
data[['so2','state']].groupby(["state"]).median().sort\_values(by='so2',ascending=False).tail(10).plot.bar(color='r')

plt.show()



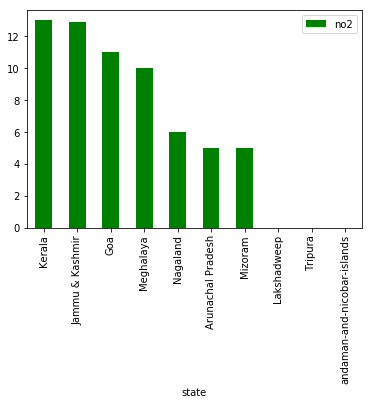
data[['no2','state']].groupby(["state"]).median().sort\_values(by='no2',ascending=False).head(10).plot.bar(color='g')

plt.show()



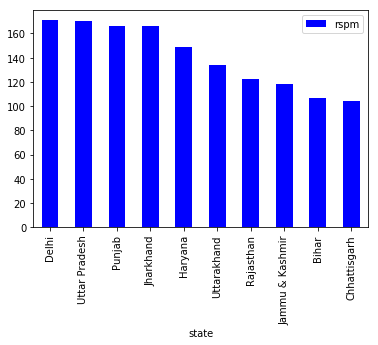
data[['no2','state']].groupby(["state"]).median().sort\_values(by='no2',ascending=False).tail(10).plot.bar(color='g')

plt.show()



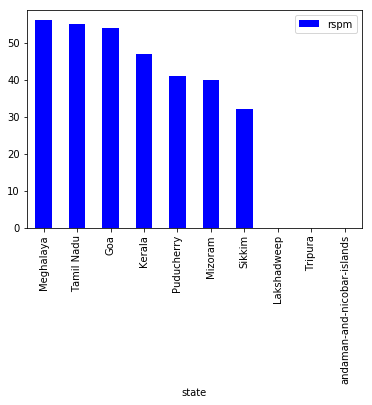
data[['rspm','state']].groupby(["state"]).median().sort\_values(by='rspm',ascending=False).head(10).plot.bar(color='b')

plt.show()



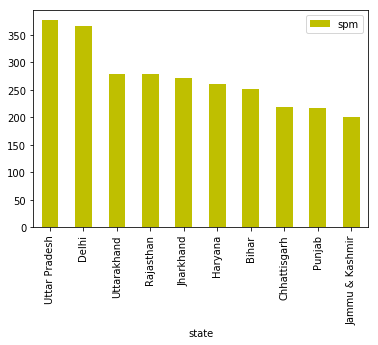
data[['rspm','state']].groupby(["state"]).median().sort\_values(by='rspm',ascending=False).tail(10).plot.bar(color='b')

plt.show()



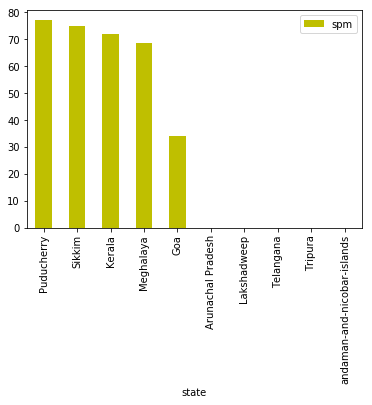
data[['spm','state']].groupby(["state"]).median().sort\_values(by='spm',ascending=False).head(10).plot.bar(color='y')

plt.show()

****

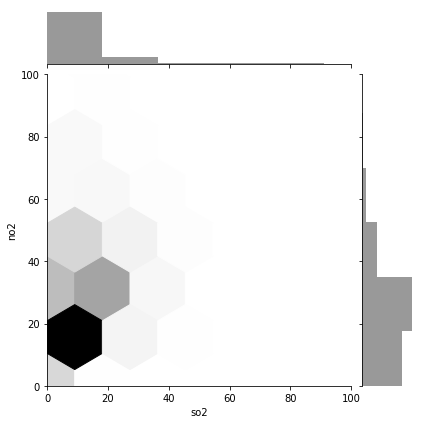
data[['spm','state']].groupby(["state"]).median().sort\_values(by='spm',ascending=False).tail(10).plot.bar(color='y')

plt.show()

****

sns.jointplot(x='so2', y='no2', data=data,kind='hex',color='k',xlim={0,100}, ylim={0,100})

<seaborn.axisgrid.JointGrid at 0x7f7a999be9b0>

****

data['date'] = pd.to\_datetime(data['date'],format='%Y-%m-**%d**') *# date parse*

data['year'] = data['date'].dt.year *# year*

data['year'] = data['year'].fillna(0.0).astype(int)

data = data[(data['year']>0)]

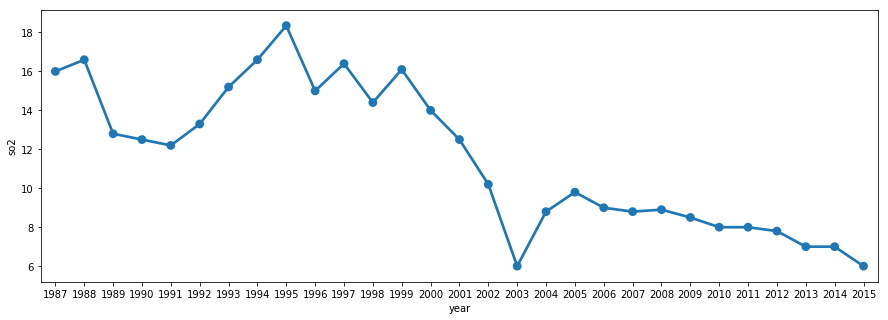
**SO2 ANALYSIS :**

df = data[['so2','year','state']].groupby(["year"]).median().reset\_index().sort\_values(by='year',ascending=False)

f,ax=plt.subplots(figsize=(15,5))

sns.pointplot(x='year', y='so2', data=df)

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f7a975db208>

****

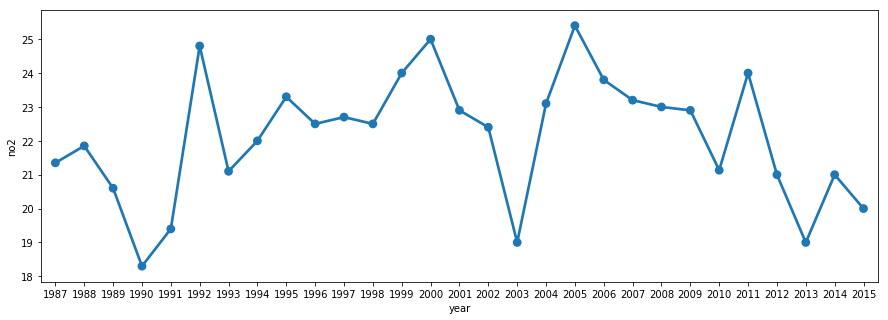
**NO2 ANALYSIS :**

df = data[['no2','year','state']].groupby(["year"]).median().reset\_index().sort\_values(by='year',ascending=False)

f,ax=plt.subplots(figsize=(15,5))

sns.pointplot(x='year', y='no2', data=df)

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f7a90c69d68>

****

**COMMON DATAS FOR AIR QUALITY ANALYSIS :**

When conducting an air quality analysis, it's essential to gather various data to assess the levels of pollutants and their impact on the environment and human health. Here are some common types of data and parameters to consider:

**1. Particulate Matter (PM):**

- PM2.5 (particles with a diameter of 2.5 micrometers or smaller)

- PM10 (particles with a diameter of 10 micrometers or smaller)

**2. Gaseous Pollutants:**

- Nitrogen Dioxide (NO2)

- Sulfur Dioxide (SO2)

- Carbon Monoxide (CO)

- Ozone (O3)

- Volatile Organic Compounds (VOCs)

**3. Meteorological Data:**

- Temperature

- Relative Humidity

- Wind speed and direction

- Atmospheric pressure

- Precipitation

**4. Geographic Data:**

- GPS coordinates or location information

- Topography and land use data

- Elevation

**5. Time Data:**

- Time and date of measurements

- Data collection frequency (e.g., hourly, daily, or continuous monitoring)

**6. Source Emissions Data:**

- Information on local industrial and traffic sources

- Emission inventories

**7. Health Data:**

- Information on health outcomes and population demographics in the area

- Hospital admissions and emergency room visits related to air quality

**8. Air Quality Index (AQI) Data:**

- Calculated AQI values based on pollutant concentrations

**9. Historical Data:**

- Long-term trends and historical records of air quality

- Seasonal variations

**10. Air Quality Monitoring Equipment Data:**

- Calibration and maintenance records

- Sensor types and specifications

**11. Remote Sensing Data:**

- Data from satellites, drones, or other remote sensing technologies for regional analysis

**12. Regulatory Standards:**

- National and local air quality standards and regulations

**13. Data on Chemical Composition:**

- Information on the specific chemical composition of pollutants, including fine particulate matter components (e.g., metals, organic compounds)

**14. Data on Health Impacts:**

- Epidemiological studies and health impact assessments related to air quality

**15. Public Complaints and Feedback:**

- Community reports, complaints, and feedback regarding air quality issues

**16. Weather Forecasts:**

- Short-term and long-term weather forecasts to predict air quality conditions

**17. Land Use and Urban Planning Data:**

- Data on urban development and land use policies that may affect air quality

Collecting and analyzing these data points helps researchers, environmental agencies, and policymakers assess air quality, identify pollution sources, and make informed decisions to mitigate air pollution and protect public health and the environment. The specific data needed can vary depending on the goals and scope of the air quality analysis.

**CONCLUSION:**

The air quality analysis conducted in location has provided valuable insights into the state of air quality in the area. The findings and observations from this study shed light on the environmental and health implications of air pollution, as well as the potential sources contributing to these issues. The assessment of air quality revealed several important points:

* **Pollutant Levels:** The analysis showed that certain pollutants, particularly [highlight specific pollutants, e.g., PM2.5, NO2], frequently exceeded established air quality standards and guidelines. This poses a significant concern for public health and the environment.
* **Sources of Pollution:** Our investigation identified various sources of pollution, including [mention specific sources, such as industrial emissions, traffic, and natural factors]. Understanding these sources is critical for targeted pollution mitigation efforts.
* **Health and Environmental Impacts:** The elevated levels of air pollutants have the potential to have adverse effects on human health, including increased respiratory issues and cardiovascular problems. Additionally, the environment, including local ecosystems, may be experiencing negative consequences due to air pollution.
* **Seasonal and Geographic Variations:** Seasonal and geographic variations in air quality were evident in the data, suggesting the influence of weather patterns and specific geographic factors on pollution levels.
* **Regulatory Compliance:** Regrettably, compliance with air quality regulations is a matter of concern, as [highlight instances of non-compliance]. Addressing these compliance issues is crucial to protecting the well-being of the community.