OBSERVING WIND QUALITY IN INDIA

PYTHON CODING

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()

|  |
| --- |

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

replacements = {

'state': {

r'Uttaranchal': 'Uttarakhand',

}

}

data.replace(replacements, regex=True, inplace=True)

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})

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)]

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)

#Heatmap Pivot with State as Row, Year as Col, No2 as Value

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

ax.set\_title('{} by state and year'.format('so2'))

sns.heatmap(data.pivot\_table('so2', index='state',

columns=['year'],aggfunc='median',margins=True),

annot=True,cmap="YlGnBu", linewidths=.5, ax=ax,cbar\_kws={'label': 'Annual Average'})

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)

#Heatmap Pivot with State as Row, Year as Col, So2 as Value

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

ax.set\_title('{} by state and year'.format('no2'))

sns.heatmap(data.pivot\_table('no2', index='state',

columns=['year'],aggfunc='median',margins=True),

annot=True,cmap="YlGnBu", linewidths=.5, ax=ax,cbar\_kws={'label': 'Annual Average'})