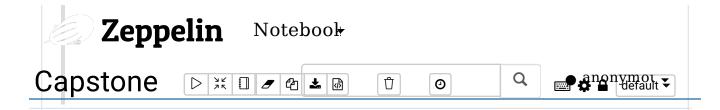
```
FINISHED ▷  🏻 🕸
%pyspark
import os
import glob
import pandas as pd
p = '/home/gokul/Documents/pollution/'
files = glob.glob(os.path.join(p, "*.csv"))
df = (pd.read_csv(f) for f in files)
   = pd.concat(df, ignore_index=True)
print df.head(5)
                                                                           FINISHED ▷  🏻 🕸
%pyspark
df['longitude'] = df['longitude'].astype(str)
df['latitude'] = df['latitude'].astype(str)
df["location"] = df[["longitude" ,"latitude"]].apply(lambda x: ','.join(x), axis=1)
                                                                           FINISHED ▷  🏻 🕸
%pyspark
arr = df.location.unique()
print "Number of sensors is" + len(arr)
                                                                             READY ▷  🏻 🕸
%pyspark
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import StringIO
def show(p):
import plotly.plotly as py
import plotly.graph_objs as go
py.sign_in('gkrishnan', 'raCcZvH0wUkyyr2YlpIL')
trace = go.Scatter( x = data['timestamp'], y = data['ozone'] )
```

1 of 3 03/25/2017 11:55 PM



```
READY ▷  🏻 🕸
%pyspark
for col in df.columns: df[col] = df[col].astype(str)
scl = [[0.0, 'rgb(242,240,247)'],[0.2, 'rgb(218,218,235)'],[0.4, 'rgb(188,189,220)'],
\ [0.6, 'rgb(158,154,200)'],[0.8, 'rgb(117,107,177)'],[1.0, 'rgb(84,39,143)']]
df['text'] = df['ozone']
mapplot = [ dict( type='choropleth', colorscale = scl, autocolorscale = False,
locations = df['latitude'], z = df['ozone'].astype(float),
locationmode = 'USA-states', text = df['text'],
marker = dict( line = dict ( color = 'rqb(255,255,255)', width = 2 ) ),
colorbar = dict( title = "Millions USD") ) ]
layout = dict( title = '2011 US Agriculture Exports by State
(Hover for breakdown)', geo = dict( scope='usa', projection=dict( type='albers usa' ),
showlakes = True, lakecolor = 'rgb(255, 255, 255)'), )
fig = dict( data=mapplot, layout=layout )
py.iplot( fig, filename='d3-cloropleth-map' )
his = [ go.Histogram( x=x ) ]
layout = dict(title = 'Distribution Of Ozone', xaxis = dict(title = 'Count'),
yaxis = dict(title = 'Ozone'), )
figh1 = dict(data=his, layout=layout)
py.plot(figh1)
```

```
%pyspark

import plotly.plotly as py
import plotly.graph_objs as go
x = data['carbon_monoxide']
his1 = [ go.Histogram( x=x ) ]
layout = dict(title = 'Distribution Of Carbon Monoxide',
xaxis = dict(title = 'Count'), yaxis = dict(title = 'Carbon Monoxide'), )
figh2 = dict(data=his1, layout=layout)
py.plot(figh2)
```

2 of 3 03/25/2017 11:55 PM



3 of 3 03/25/2017 11:55 PM