**Test: Data Analysis using Python**

Approach:

**The Objective:**

Perform analysis of a dataset containing vehicle speed information Python script that works on dataset. It will generate a report with at least 20 visualizations and text descriptions.

Tools and Techniques:

.iPython/Jupyter notebook and pandas.

.matplotlib, plot.ly etc.

**Dataset columns are:**

.latitude, longitude

.timestamp

.speed

.speedLimit,

.functionalClass (a description of the road type)

.controlledAccess (if yes, this is a controlled-access road, like a highway)

.isHighway (a different kind of road classification scheme)

.state (the state in which we put this road)

.vehicle\_id

.vehicle\_type

.driver\_id (if null: missing ID)

[1]import pandas as pd

import matplotlib.pyplot as plt

data = [[34.640141,-77.938721,64,70,2,1208979,5693],

[34.64373,-77.9394,64,70,2,1208979,5693],

[34.551254,-78.78463,1,55,4,1262441,916461],

[35.613663,-78.470596,62,70,2,1280223,1145487],

[35.614525,-78.47467,61,70,2,1280223,1145487],

[34.860115,-77.999969,64,70,2,1208979,5693],

[34.328571,-78.926056,23,55,4,550280,500201],

[35.756538,-78.699715,3,55,5,1280223,1145487],

[34.402813,-78.916901,65,70,2,550280,500201],

[35.039543,-78.149475,64,70,2,1208979,5693],

[35.042751,-78.150772,64,70,2,1208979,5693],

[34.443115,-78.959885,64,70,2,550280,500201],

[34.590424,-79.068939,65,65,2,550280,500201],

[35.191341,-78.28302,64,70,2,1208979,5693],

[34.55217,-78.784126,27,55,4,1262441,916461],

[34.741947,-78.997086,65,65,1,550280,500201],

[34.751602,-78.996979,65,65,1,550280,500201],

[35.362625,-78.365662,65,55,3,1208979,5693],

[35.367558,-78.366119,50,55,3,1208979,5693]

]

df = pd.DataFrame(data, columns = ['latitude', 'longitude',

'speed\_mph', 'speedlimit\_mph',

'functionalClass', 'vehicle\_id','driver\_id'] )

df.hist()

plt.show()

[2]

import pandas as pd

import matplotlib.pyplot as plt

data = [[34.640141,-77.938721],

[34.64373,-77.9394],

[34.551254,-78.78463],

[35.613663,-78.470596],

[35.614525,-78.47467],

[34.860115,-77.999969],

[34.328571,-78.926056],

[35.756538,-78.699715],

[34.402813,-78.916901],

[35.039543,-78.149475],

[35.042751,-78.150772],

[34.443115,-78.959885],

[34.590424,-79.068939],

[35.191341,-78.28302],

[34.55217,-78.784126],

[34.741947,-78.997086],

[34.751602,-78.996979],

[35.362625,-78.365662],

[35.367558,-78.366119]

]

df = pd.DataFrame(data, columns = ['latitude', 'longitude'] )

df.plot.bar()

plt.bar(df['latitude'], df['longitude'])

plt.xlabel("latitude")

plt.ylabel("longitude")

plt.show()

[3]

df.plot.box()

plt.boxplot(df['longitude'])

plt.show()

[4]

df.plot.box()

plt.boxplot(df['latitude'])

plt.show()

[5]

import pandas as pd

import matplotlib.pyplot as plt

data = [[64,5693],

[64,5693],

[1,916461],

[62,1145487],

[61,1145487],

[64,5693],

[23,500201],

[3,1145487],

[65,500201],

[64,5693],

[64,5693],

[64,500201],

[65,500201],

[64,5693],

[27,916461],

[65,500201],

[65,500201],

[65,5693],

[50,5693]

]

df = pd.DataFrame(data, columns = ['speed\_mph', 'driver\_id'] )

plt.scatter(df['speed\_mph'], df['driver\_id'])

plt.show()

Thanks...