Python Data Visualisation Matplotlib / Seaborn

BY… SAKEEB SHEIKH

Packages Used

- Matplotlib
- Seaborn

Commands to Install Package

- # To Install Matplotlib Package
- !pip install matplotlib

- # To Install Seaborn Package
- !pip install seaborn

Importing Packages

- import numpy as np
- from numpy.random import randn
- import pandas as pd
- from pandas import Series, DataFrame

****

- import matplotlib.pyplot as plt
- from matplotlib import rcParams
- import seaborn as sb

****

- %matplotlib inline
- rcParams['figure.figsize'] = 5, 4
- sb.set_style('whitegrid')

Types Of Graphs / Charts

- Line Chart
- Bar Chart
- Pie Chart
- Histogram Plot
- Scattered Plot
- Box Plot

Line Chart

- Using Matplotlib
  - x = range(1,10)
  - y = [1,2,3,4,0,4,3,2,1]
  - plt.plot(x, y)

- Using Pandas Object
  - cars = pd.read_csv('mtcars.csv')
  - cars.columns = ['car_names','mpg','cyl','disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am', 'gear', 'carb']
  - mpg = cars['mpg']
  - mpg.plot()

Multiple Lines in Line chart

- df = cars[['cyl', 'wt', 'mpg']]
- df.plot()

Plot **Functions** to Decorate Plot

- ## plt method
- xlim / ylim
- xticks / yticks
- xlabel / ylabel /title
- figure()
- subplots(r,c)　→　fig,(ax1,ax2)=plt.subplot(1,2)
- legend
- ## axes methods
- grid (removes grid)
- set_xticks
- set_xticklabels
- set_title
- set_xlabel
- set_ylabel
- Legend
- annotate

Plot **Argument**

- color(line chart , bar chart) / colors(pie chart)
- width (bar chart)
- align (bar chart)
- ls (line chart) → 'steps', '--'
- lw (line chart)
- marker (line chart)
- Mew (line chart)
- labels (pie chart)
- rotation (axes.xticklabels) (60)
- fontsize (axes.xticklabels) ('medium')
- loc (legend()) ('best')
- kind='barh' (bar chart)

Bar chart

- Using Matplotlib
  - x = range(1,10)
  - y = [1,2,3,4,0,4,3,2,1]
  - **plt.bar(x, y)**

- Using Pandas Object
  - cars = pd.read_csv('mtcars.csv')
  - cars.columns = ['car_names','mpg','cyl','disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am', 'gear', 'carb']
  - mpg = cars['mpg']
  - **mpg.plot(kind='bar')**

Plotting Horizontal Bar Chart

- mpg.plot(kind='barh')

Pie chart

- x = [1,2,3,4,0.5]
- plt.pie(x)
- plt.show()

Saving Plot to Image File

- plt.savefig('pie_chart.jpeg')
- plt.show()

Using Sample function

- x= pd.DataFrame({'a':[1,2,3,4,5,6,7]})
- y = x.sample(5)

- random_state = 2 (any number)

Creating visualizations from time series data

- Importing Data
  - address = 'Superstore-Sales.csv'
  - df = pd.read_csv(address, index_col='Order Date', parse_dates=True)
  - df.head()
- Plot One Column
  - df['Order Quantity'].plot()
- Sampling data randomly
  - df2 = df.sample(n=100, random_state=25, axis=0)

Ploting Time Series Data

- plt.xlabel('Order Date')
- plt.ylabel('Order Quantity')
- plt.title('Superstore Sales')
- df2['Order Quantity'].plot()

histograms, box plots, and scatter plots

- Importing Packages
  - import numpy as np
  - import pandas as pd
  - from pandas import Series, DataFrame
  - from pandas.tools.plotting import scatter_matrix
  - import matplotlib.pyplot as plt
  - from pylab import rcParams
  - import seaborn as sb

  - %matplotlib inline
  - rcParams['figure.figsize'] = 5, 4
  - sb.set_style('whitegrid')

Dataset distributions with histograms

- Using DataFrame Function
  - address = 'datasets/mtcars.csv'
  - cars = pd.read_csv(address)
  - cars.columns = ['car_names','mpg','cyl','disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am', 'gear', 'carb']
  - cars.index = cars.car_names
  - mpg = cars['mpg']
  - mpg.plot(kind='hist')
- Using Matplotlib Funtion
  - plt.hist(mpg)
  - plt.plot()

Seaborn Function for Histogram

- Using Seaborn Function
  - sb.distplot(mpg)

Plotting Scattered Plot

- Using DataFrame Function
  - cars.plot(kind='scatter', x='hp', y='mpg', c=['darkgray'], s=150)

Plotting Regressing Line on Scatter Plot

- Using Seaborn
  - sb.regplot(x='hp', y='mpg', data=cars, scatter=True)

Finding Correlation using Pandas corr() function

- For all the columns in DataFrame
  - cars.corr()

- For all the columns in DataFrame
  - cars[['mpg','cyl','disp', 'hp']].corr()

Generating Scatter Plot matrix(Correlation Plot)

Using Seaborn Function

- For all the Columns
  - sb.pairplot(cars)

- For specific columns
  - sb.pairplot(cars[['mpg','cyl','disp', 'hp']])

Scatter Plot Matrix

- Using DataFrame Method
  - address = 'datasets/mtcars.csv'
  - cars = pd.read_csv(address)
  - cars.columns = ['car_names','mpg','cyl','disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am', 'gear', 'carb']
  - cars.index = cars.car_names
  - cars_df = pd.DataFrame(cars.iloc[:,[1,3,4,6]], columns = ['mpg', 'disp', 'hp', 'wt'])
  - cars_target = cars.iloc[:,9]
  - target_names = [0, 1]
  - cars_df['group'] = pd.Series(cars_target, dtype="category")
  - sb.pairplot(cars_df, hue='group', palette='hls')

Plotting Box Plot

- Using DataFrame Method
  - cars.boxplot(column='mpg', by='am')
  - cars.boxplot(column='wt', by='am')

- Using Seaborn Method
  - sb.boxplot(x='am', y='mpg', data=cars, palette='hls')