DATA VISUALISATION IN PYTHON

CHEATSHEET

Why Is Data Visualisation an Important Concept ?

- Because it help us understand distribution, trend, relationship, comparison and composition of data values
- It helps decision makers to quickly examine large piles of data and discover the hidden pattern/insights

" BEAUTY OF AN ART LIES IN THE MESSAGE IT CONVEYS "

WHAT IS REQUIRED TO MAKE VISUALISATION IN PYTHON ?

MATPLOTLIB	SEABORN
Python based plotting library offers matplotlib with a complete 2D support along with limited 3D graphic support. It is useful in producing publication quality figures in interactive environments across platforms.	Being based on matplotlib, seaborn offers various features such as built in themes, color palettes, functions and tools to visualize univariate, bivariate, linear regression, matrices of data, statistical time series etc which lets us to build complex visualizations.

Sample Data Set Used For The VISUALISATION Show Below

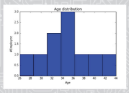
EMPID	Gender	Age	Sales	BMI	Income
1001	M	32	123	Normal	300
1002	F	40	133	Overweight	400
1003	F	37	135	Obesity	500
1004	M	32	137	Underweight	150
1005	F	44	117	Underweight	180
1006	M	36	121	Normal	300
1007	M	32	133	Obesity	500
1008	F	39	140	Normal	520
1009	M	32	133	Normal	450
1010	M	34	133	Underweight	210

Import Data Set:

```
import matplotlib.pyplot as plt
import pandas as pd
df=pd.read_excel("E:/First.xlsx", "Sheet1")
```

Histogram

```
fig=plt.figure()
ax = fig.add_subplot(111)
ax.hist(df['Age'],bins = 7) # Here you can
play with number of bins Labels and Tit
plt.title('Age distribution')
plt.xlabel('Age')
plt.ylabel('#Employee')
plt.show()
```

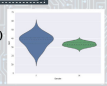


Box Plot

```
importmatplotlib.pyplot as plt
import pandas as pd
fig=plt.figure()
ax = fig.add_subplot(111)
x.boxplot(df['Age'])
plt.show()
```

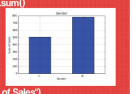
Violin Plot

```
import seaborn as sns
sns.violinplot(df['Age'], df['Gender'])
#Variable Plots
sns.despine()
```



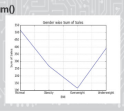
Bar Chart

```
var = df.groupby('Gender').Sales.sum()
#grouped sum of sales at
Gender level
fig = plt.figure()
ax1 = fig.add_subplot(111)
ax1.set_xlabel('Gender')
ax1.set_ylabel('Sum of Sales')
ax1.set_title("Gender wise Sum of Sales")
var.plot(kind='bar')
```



Line Chart

```
var = df.groupby('BMI').Sales.sum()
fig = plt.figure()
ax1 = fig.add_subplot(111)
ax1.set_xlabel('BMI')
ax1.set_ylabel('Sum of Sales')
ax1.set_title("BMI wise Sum
of Sales")
var.plot(kind='line')
```

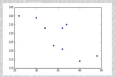


Stacked Column Chart

```
var = df.groupby(['BMI','Gender']).Sales.sum()
var.unstack().plot(kind='bar',stacked=True, color=['red','blue'], grid=False)
```

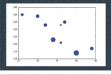
Scatter Plot

```
fig = plt.figure()
ax = fig.add_subplot(111)
ax.scatter(df['Age'],df['Sales'])
plt.show()
```



Bubble Plot

```
fig = plt.figure()
ax = fig.add_subplot(111)
ax.scatter(df['Age'],df['Sales'], s=df['Income'])
plt.show()
```

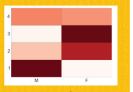


Pie Chart

```
var=df.groupby(['Gender']).sum().stack()
temp=var.unstack()
type(temp)
x_list = temp['Sales']
label_list = temp.index
pyplot.axis('equal') #The pie chart
is oval by default. To make it a
circle use pyplot.axis('equal')
plt.pie(x_list,labels=label_list,autopct="%1if%%")
plt.title("Pastafarianism expenses")
plt.show()
```

Heat Map

```
import numpy as np
data = np.random.rand(4,2)
rows = list('1234') #rows
categories columns =
list('MF') #column categories
fig,ax=plt.subplots()
ax.pcolor(data,cmap=plt.cm.Reds,edgecolors='k')
ax.set_xticks(np.arange(0,2)+0.5)
ax.set_yticks(np.arange(0,4)+0.5)
ax.xaxis.tick_bottom()
ax.yaxis.tick_left()
ax.set_xticklabels(columns,minor=False,fontsize=20)
ax.set_yticklabels(rows,minor=False,fontsize=20)
plt.show())
```



To view the complete guide on data visualisation in python visit here : http://bit.ly/1FjTkRF

Analytics Vidhya
Learn Everything About Analytics