

GROUP ASSIGNMENT # 02

STATISTICS AND PROBABILITY THEORY

NAMES	REGISTRATION		
SUDAIS KHAN	FA20-BSE-042		
ASADULLAH	FA20-BSE-028		
UBAID UR REHMAN	FA20-BSE-012		
MUHAMMAD ALI MALIK	FA20-BSE-058		
SECTION	BSE-5B		

> Submitted To Sir Atta Ullah

Question 1: Why data visualization is important? Explain different methods of data visualization.

Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs. This makes the data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns, and outliers within large data sets.

Why is Data Visualization Important?

Data visualization can help by delivering data in the most efficient way possible. As one of the essential steps in the business intelligence process, data visualization takes the raw data, models it, and delivers the data so that conclusions can be reached. In advanced analytics, data scientists are creating machine learning algorithms to better compile essential data into visualizations that are easier to understand and interpret.

Data visualization uses visual data to communicate information in a manner that is universal, fast, and effective. This practice can help companies identify which areas need to be improved, which factors affect customer satisfaction and dissatisfaction, and what to do with specific products (where should they go and who should they be sold to). Visualized data gives stakeholders, business owners, and decision-makers a better prediction of sales volumes and future growth.

List of Methods to Visualize Data

Column Chart: It is also called a vertical bar chart where each category is represented by a rectangle. The height of the rectangle is proportional to the values that are plotted.

Bar Graph: It has rectangular bars in which the lengths are proportional to the values which are represented.

Stacked Bar Graph: It is a bar style graph that has various components stacked together so that apart from the bar, the components can also be compared to each other.

Stacked Column Chart: It is like stacked bar; however, the data is stacked horizontally.

Area Chart: It combines the line chart and bar chart to show how the numeric values of one or more groups change over the progress of a viable area.

Dual Axis Chart: It combines a column chart and a line chart and then compares the two variables.

Line Graph: The data points are connected through a straight line; therefore, creating a representation of the changing trend.

Mekko Chart: It can be called a two-dimensional stacked chart with varying column widths.

Pie Chart: It is a chart where various components of a data set are presented in the form of a pie which represents their proportion in the entire data set.

Waterfall Chart: With the help of this chart, the increasing effect of sequentially introduced positive or negative values can be understood.

Bubble Chart: It is a multi-variable graph that is a hybrid of Scatter Plot and a Proportional Area Chart.

Scatter Plot Chart: It is also called a scatter chart or scatter graph. Dots are used to denote values for two different numeric variables.

Bullet Graph: It is a variation of a bar graph. A bullet graph is used to swap dashboard gauges and meters.

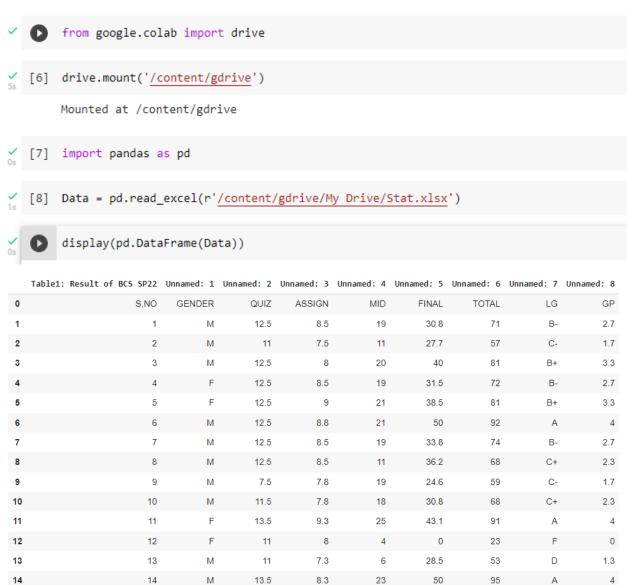
Funnel Chart: The chart determines the flow of users with the help of a business or sales process.

Heat Map: It is a technique of data visualization that shows the level of instances as color in two dimensions.

Question 2: (Using Python) The results of two classes are given. Present these data sets of by frequency Distribution, histogram, frequency curve, simple bar chart, multiple bar chart, component bar chart and pie chart. Describe the performance of each class and make comparison between (a) two classes (b) over all two classes (c) Male and Female Students by class vise and overall.

Table1: Result of BCS SP22

Description: Here we are fetching the data of Excel into python.



15	15	М	11	8	2	13.1	34	F	0
16	16	М	12.5	8.5	13	36.2	70	B-	2.7
17	17	F	13	8.5	15	32.3	69	C+	2.3
18	18	М	12	7.8	13	33.1	66	C+	2.3
19	19	F	12.5	8.5	19	50	90	Α	4
20	20	F	14	9.5	23	50	97	Α	4
21	21	F	12	7.8	18	50	88	A-	3.7
22	22	М	11.5	7.8	20	40	79	В	3
23	23	F	12.5	8.8	21	44.6	87	A-	3.7
24	24	F	13.5	9	24	50	97	Α	4
25	25	F	11.5	8.5	11	19.2	50	D	1.3
26	26	F	13	8.5	12	17.7	51	D	1.3
27	27	F	11	8	9	22.3	50	D	1.3
28	28	М	12	8.8	20	26.9	68	C+	2.3

FREQUENCY DISTRIBUTION:

Description: In this step we have find the frequency distribution of BCS SP22

```
_{\text{De}} [12] import numpy as np
[13] import pandas as pd
def make_frequency_distribution(data, user_input=None, extra=True):
## total number of observations
length = len(data)
## lowest and highest number in the data
lowest = min(data)
highest = max(data)
## total number of class
if user_input == None:
    total_classes = int(np.sqrt(length))
    lowest, highest, total_classes = user_input
## range of the data
range_ = highest - lowest
print(f"Start value: {lowest}")
print(f"End value: {highest}")
print(f"Range: {range_}")
print(f"Total Number of Classes: {total_classes}")
## calculate width
width = range_ / total_classes
```

```
if extra:
        ## class midpoint
        df["Class Midpoint"] = df["Class Intervals"].apply(
           lambda x: (
                ( float(x.split(' ')[0]) + float(x.split(' ')[-1]) ) / 2
        ## relative frequency
       df["Relative Frequency"] = df["Frequency"] / df["Frequency"].sum()
        ## cumulative frequency
        df["Cumulative Frequency"] = df["Frequency"].cumsum()
    return df
## data
test_scores = np.array([
    71,57,81,72,81,92,74,68,59,68,91,23,53,95,34,70,69,66,90,97,88,79,87,97,50,51,50,68
])
## without specifying user input
make_frequency_distribution(test_scores)
```

Start value: 23 End value: 97 Range: 74 Total Number of Classes: 5 Class Width = 14.8

Class Intervals Frequency Class Midpoint Relative Frequency Cumulative Frequency

0	23.0 - under 37.8	2	30.4	0.071429	2
1	37.8 - under 52.6	3	45.2	0.107143	5
2	52.6 - under 67.4	4	60.0	0.142857	9
3	67.4 - under 82.2	11	74.8	0.392857	20
4	82.2 - under 97.0	8	89.6	0.285714	28

Pie Chart:

Description: In this step we have find the males and female student population in BCS SP22.

```
# Table 1: Result of BCS SP22

import matplotlib.pyplot as plt

Gender="Males","Females"

popularity=[15,13]

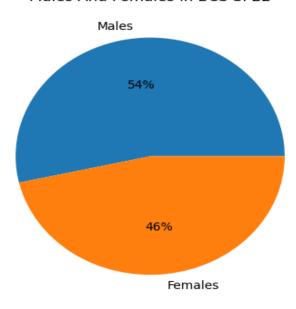
colors = ["#1f77b4", "#ff7f0e"]

plt.pie(popularity_labels=Gender_colors=colors)

plt.title("Males And Females in BSC SP22")

plt.show()
```

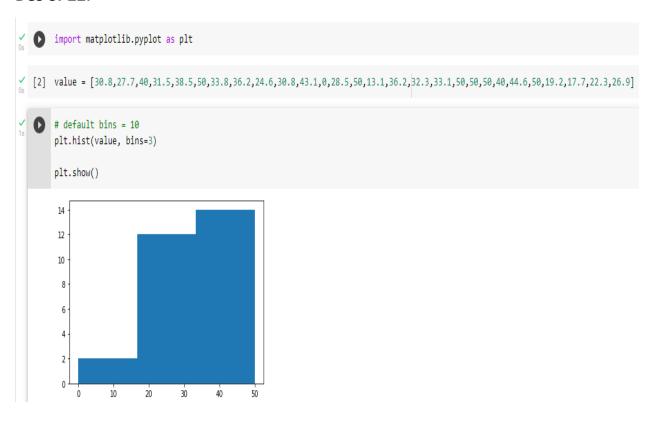
Males And Females in BCS SP22



HISTOGRAM:

FINAL MARKS

Description: In this Step we have make the histogram of Final Marks Attribute of BCS SP22.



Simple Bar Chart:

Description: In this chart we have find the Grades of Males and Females Students of BCS SP22.

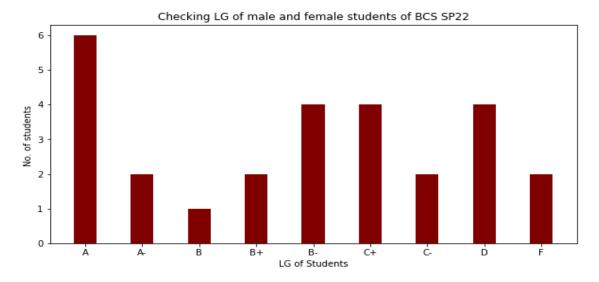
```
import numpy as np
import matplotlib.pyplot as plt

# creating the dataset
data = {'A':6, 'A-':2, 'B':1, 'B+':2, 'B-':4, 'C+':4, 'C-':2, 'D':4, 'F':2}
courses = list(data.keys())
values = list(data.values())

fig = plt.figure(figsize = (10, 5))

# creating the bar plot
plt.bar(courses, values, color ='maroon', width = 0.4)

plt.xlabel("LG of Students")
plt.ylabel("No. of students")
plt.ylabel("No. of students ")
plt.title("Checking LG of male and female students of BCS SP22")
plt.show()
```

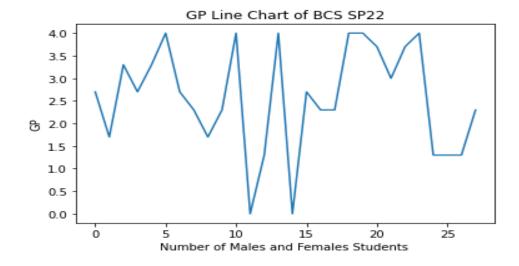


Frequency Curve:

Description: In this chart we have make the frequency curve of GP of Male and Female Students BCS SP22

```
def fnplot(list1):
    plt.plot(list1)
    plt.title("GP Line Chart of BCS SP22")
    plt.xlabel('Number of Males and Females Students ')
    plt.ylabel('GP')
    plt.show()

list1=[2.7,1.7,3.3,2.7,3.3,4,2.7,2.3,1.7,2.3,4,0,1.3,4,0,2.7,2.3,2.3,4,4,3.7,3,3.7,4,1.3,1.3,1.3,2.3]
fnplot(list1)
```



Multiple Bar Chart:

Description: In this chart we are Finding the Quiz and Mid Marks of First 10 Student of BCS SP22.

```
import numpy as np
import matplotlib.pyplot as plt

N = 10
ind = np.arange(N)  # the x locations for the groups
width = 0.27  # the width of the bars

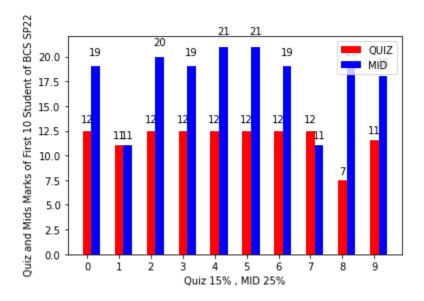
fig = plt.figure()
ax = fig.add_subplot(111)

yvals = [12.5,11,12.5,12.5,12.5,12.5,12.5,12.5,7.5,11.5]
rects1 = ax.bar(ind, yvals, width, color='r')
kvals = [19,11,20,19,21,21,19,11,19,18]
rects2 = ax.bar(ind+width, kvals, width, color='b')

ax.set_ylabel(' Quiz and Mids Marks of First 10 Student of BCS SP22')
ax.set_xticks(ind)
ax.legend( (rects1[0], rects2[0]), ('QUIZ', 'MID') )
ax.set_xlabel("Quiz 15% , MID 25%")
```

```
autolabel(rects1)
autolabel(rects2)

plt.show()
```



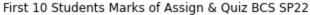
Component Bar Chart:

Description: In this chart we are finding the Assign & Quiz marks of first 10 Students of BCS SP22.

```
import matplotlib.pyplot as plt

# create data
x = [' S1',' S2','S3','S4','S5','S6','S7','S8','S9','S10']
y1 = [8.5,7.5,8,8.5,9,8.8,8.5,8.5,7.8,7.8]
y2 = [12.5,11,12.5,12.5,12.5,12.5,12.5,12.5,7.5,11.5]

# plot bars in stack manner
plt.title("First 10 Students Marks of Assign & Quiz BCS SP22")
plt.xlabel("Blue=Quiz & Red=Assign ")
plt.bar(x, y1, color='r')
plt.bar(x, y2, bottom=y1, color='b')
plt.show()
```



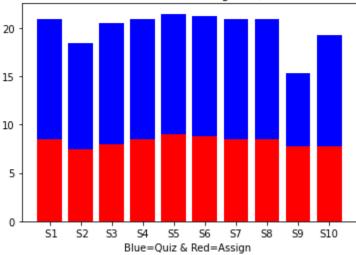
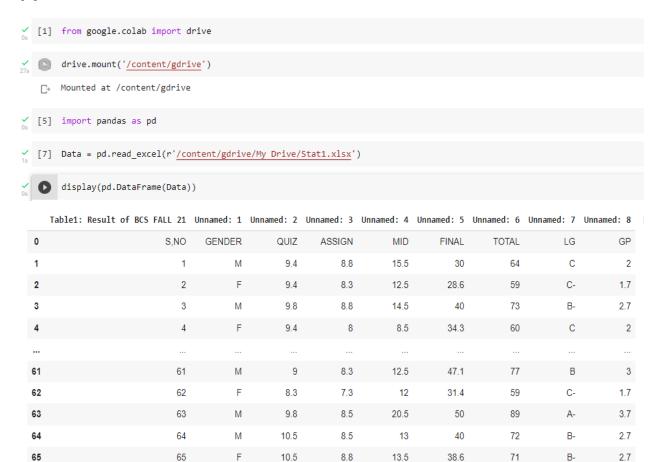


Table2: Result of BCS FALL 21

Description: In this step we are fetching the data from Excel file into python of BCS FALL 21



66 rows × 9 columns

FREQUENCY DISTRIBUTION:

Description: Here We have Find the Frequency Distribution of BCS FALL 21.

```
[9] import numpy as np
[10] import pandas as pd
✓ ♠ def make_frequency_distribution(data, user_input=None, extra=True):
          ## total number of observations
          length = len(data)
          ## lowest and highest number in the data
          lowest = min(data)
          highest = max(data)
          ## total number of class
          if user_input == None:
              total_classes = int(np.sqrt(length))
              lowest, highest, total_classes = user_input
          ## range of the data
          range_= highest - lowest
          print(f"Start value: {lowest}")
          print(f"End value: {highest}")
          print(f"Range: {range_}")
          print(f"Total Number of Classes: {total_classes}")
          ## calculate width
          width = range_ / total_classes
           ## list of all class intervals
           class_intervals = [
               np.round(start,3) for start in np.linspace(lowest, highest, total_classes+1)
           print(f"Class\ Width = \{np.round(width, 3)\}", end="\n'")
           ## calculate frequency for each class
          hist, _ = np.histogram(data, bins=class_intervals)
           ## frequency table
           df = pd.DataFrame(
               {
                   "Class Intervals": [
                       f"{first} - under {second}" \
                       for first, second in zip(class_intervals, class_intervals[1:])
                   "Frequency": hist
```

Start value: 3 End value: 92 Range: 89 Total Number of Classes: 8 Class Width = 11.125

Class Intervals Frequency Class Midpoint Relative Frequency Cumulative Frequency



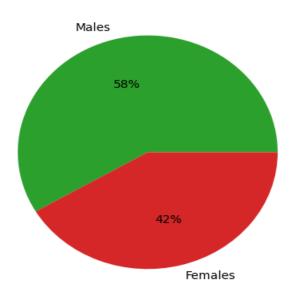
	Class Intervals	rrequency	Class miupoint	Relative Frequency	cumulative Frequency
0	3.0 - under 14.125	2	8.5625	0.030769	2
1	14.125 - under 25.25	0	19.6875	0.000000	2
2	25.25 - under 36.375	0	30.8125	0.000000	2
3	36.375 - under 47.5	2	41.9375	0.030769	4
4	47.5 - under 58.625	14	53.0625	0.215385	18
5	58.625 - under 69.75	25	64.1875	0.384615	43
6	69.75 - under 80.875	12	75.3125	0.184615	55
7	80.875 - under 92.0	10	86.4375	0.153846	65

Pie Chart:

Description: In this pie chart we have find the population of males and female students of BCS FALL 21.

```
# Table 2: Result of BCS FALL 21
import matplotlib.pyplot as plt
Gender="Males", "Females"
popularity=[38,27]
colors = ["#2ca02c", "#d62728"]
plt.pie(popularity,labels=Gender,colors=colors)
plt.title("Males And Females in BCS FALL 21")
plt.show()
```

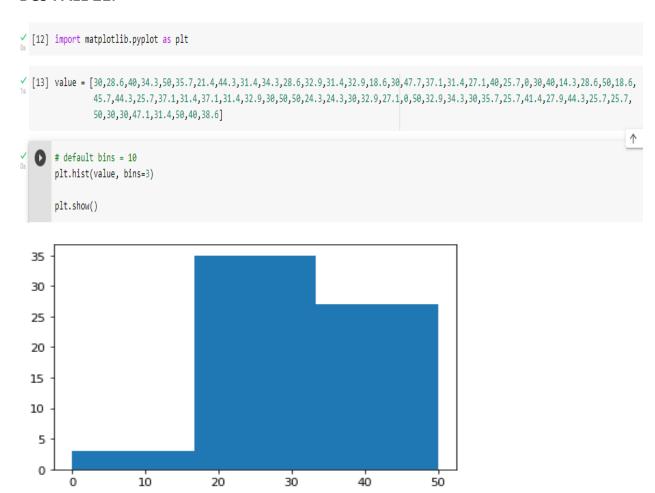
Males And Females in BCS FALL 21



HISTOGRAM:

FINAL MARKS

Description: In this Step we have make the histogram of Final Marks Attribute of BCS FALL 21.



Simple Bar Chart:

In this chart we have find the Grades of BCS FALL 21.

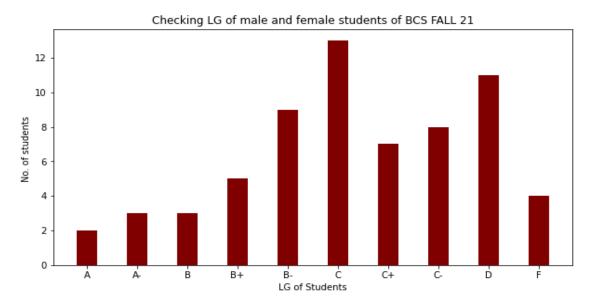
```
import numpy as np
import matplotlib.pyplot as plt

# creating the dataset
data = {'A':2,'A-':3, 'B':3,'B+':5,'B-':9,'C':13,'C+':7,'C-':8,'D':11,'F':4}
courses = list(data.keys())
values = list(data.values())

fig = plt.figure(figsize = (10, 5))

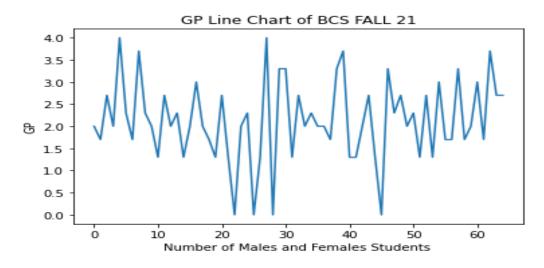
# creating the bar plot
plt.bar(courses, values, color ='maroon', width = 0.4)

plt.xlabel("LG of Students")
plt.ylabel("No. of students ")
plt.title("Checking LG of male and female students of BCS FALL 21")
plt.show()
```



Frequency Curve:

Description: In this chart we have make the frequency curve of GP of Male and Female Students BCS SP22



Multiple Bar Chart:

Description: In this chart we are Finding the Quiz and Mid Marks of First 10 Student of BCS FALL 21.

```
import numpy as np
import matplotlib.pyplot as plt

N = 10
ind = np.arange(N)  # the x locations for the groups
width = 0.27  # the width of the bars

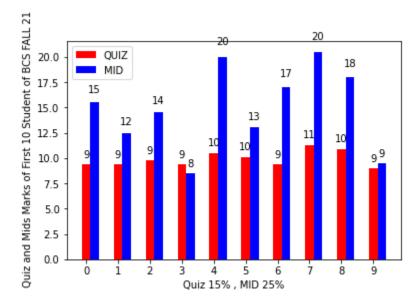
fig = plt.figure()
ax = fig.add_subplot(111)

yvals = [9.4,9.4,9.8,9.4,10.5,10.1,9.4,11.3,10.9,9]
rects1 = ax.bar(ind, yvals, width, color='r')
kvals = [15.5,12.5,14.5,8.5,20,13,17,20.5,18,9.5]
rects2 = ax.bar(ind+width, kvals, width, color='b')

ax.set_ylabel('Quiz and Mids Marks of First 10 Student of BCS FALL 21')
ax.set_xticks(ind)
ax.legend( (rects1[0], rects2[0]), ('QUIZ', 'MID') )
ax.set_xlabel("Quiz 15%, MID 25%")
```

```
autolabel(rects1)
autolabel(rects2)

plt.show()
```



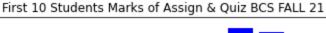
Component Bar Chart:

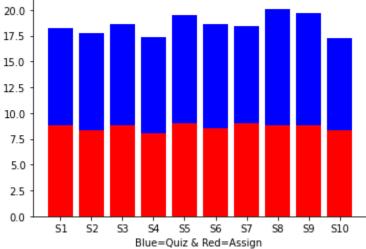
Description: In this chart we are finding the Assign & Quiz marks of first 10 Students of BCS SP22.

```
import matplotlib.pyplot as plt

# create data
x = [' S1',' S2','S3','S4','S5','S6','S7','S8','S9','S10']
y1 = [8.8,8.3,8.8,9,8.5,9,8.8,8.8,8.3]
y2 = [9.4,9.4,9.8,9.4,10.5,10.1,9.4,11.3,10.9,9]

# plot bars in stack manner
plt.title("First 10 Students Marks of Assign & Quiz BCS FALL 21")
plt.xlabel("Blue-Quiz & Red-Assign ")
plt.bar(x, y1, color='r')
plt.bar(x, y2, bottom=y1, color='b')
plt.show()
```





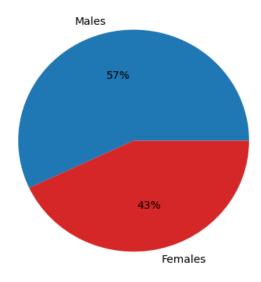
COMPARING OF MALES & FEMALES AND CLASSES OF BOTH CLASSES

COMPARE TWO CLASSES PIE CHART:

Description: In this chart we have Checking the Combine population of Males and Females Student in Both Classes.

```
# Combine Males and Females of BCS SP22 and BCS FALL 21
import matplotlib.pyplot as plt
Gender="Males", "Females"
popularity=[53,40]
colors = ["#1f77b4", "#d62728"]
plt.pie(popularity_labels=Gender_colors=colors_autopct='%1.0f%%')
plt.title("Combine Males and Females of BCS SP22 and BCS FALL 21")
plt.show()
```

Combine Males and Females of BCS SP22 and BCS FALL 21



COMPARE HISTOGRAM TO BOTH CLASSES MALES:

FINAL MARKS

Description: In this step we have make the histogram of Final Marks Attribute of Both Classes Males.

20

COMPARE SIMPLE BAR CHART TO BOTH CLASSES OF FEMALES:

Checking LG

Description: In this chart we have checking the LG of Both Classes of Females

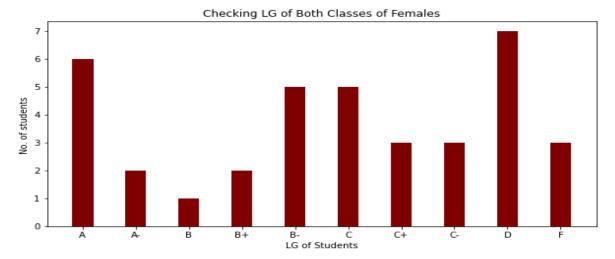
```
import numpy as np
import matplotlib.pyplot as plt

# creating the dataset
data = {'A':6,'A-':2, 'B':1,'B+':2,'B-':5,'C':5,'C+':3,'C-':3,'D':7,'F':3}
courses = list(data.keys())
values = list(data.values())

fig = plt.figure(figsize = (10, 5))

# creating the bar plot
plt.bar(courses, values, color ='maroon', width = 0.4)

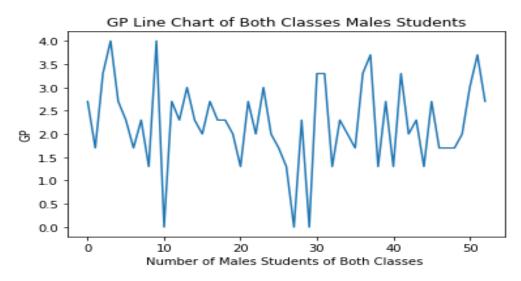
plt.xlabel("LG of Students")
plt.ylabel("No. of students")
plt.title("Checking LG of Both Classes of Females")
plt.show()
```



COMPARE FREQUENCY CURVE OF BOTH CLASSES OF MALES:

Checking GP

Description: In this Frequency Curve we are checking the Both Classes Males Students GP.

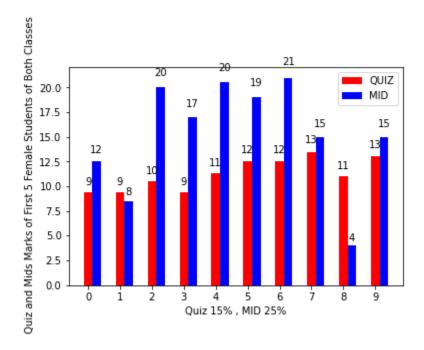


COMBINE MULTIPLE BAR CHART OF BOTH CLASSES OF FIRST 5 FEMALES STUDENTS:

Quiz & Mid

Description: In this Chart we are checking Quiz and Mid marks of Both Classes of First 5 Females Students in the tables.

```
import numpy as np
import matplotlib.pyplot as plt
N = 10
ind = np.arange(N) # the x locations for the groups
width = 0.27 # the width of the bars
fig = plt.figure()
ax = fig.add_subplot(111)
yvals = [9.4,9.4,10.5,9.4,11.3, 12.5,12.5,13.5,11,13]
rects1 = ax.bar(ind, yvals, width, color='r')
kvals = [12.5,8.5,20,17,20.5, 19,21,15,4,15]
rects2 = ax.bar(ind+width, kvals, width, color='b')
ax.set_ylabel(' Quiz and Mids Marks of First 5 Female Students of Both Classes')
ax.set_xticks(ind)
ax.legend( (rects1[0], rects2[0]), ('QUIZ', 'MID') )
ax.set_xlabel("Quiz 15% , MID 25%")
autolabel(rects1)
autolabel(rects2)
plt.show()
```



COMBINE COMPONENT BAR CHART OF BOTH CLASSES OF FIRST 5 MALES STUDENTS:

Assign & Quiz

Description: In this Chart we are checking Assign & Quiz marks of Both Classes of First 5 Males Students In the tables.

```
import matplotlib.pyplot as plt

# create data
x = [' S1',' S2','S3','S4','S5','S6','S7','S8','S9','S10']
y1 = [8.5,7.5,8,8.8,8.5, 8.8,8.8,8.5,8.8,8.3]
y2 = [12.5,11,12.5,12.5, 12.5, 9.4,9.8,10.1,10.9,9]

# plot bars in stack manner
plt.title("First 5 Males Students Marks of Assign & Quiz of Both Classes")
plt.xlabel("Blue=Quiz & Red=Assign ")
plt.bar(x, y1, color='r')
plt.bar(x, y2, bottom=y1, color='b')
plt.show()
```

First 5 Males Students Marks of Assign & Quiz of Both Classes

