



SAL Engineering & Technical Institute

IT Department

DATA SCIENCE PRACTICALS

Year:2021-2022

Name : Patel Ayushkumar

Enrollment no.: 191260116029

In [1]:

```
import pandas as pd
```

In [2]:

```
columns = ['Sr. No.', 'Enrollment Number', 'Name', 'Gender', 'Mobile Number', 'City', 'COA',
sr_no = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23,
Enrollment_Number = [191260116001, 191260116002, 191260116003, 191260116004, 191260116005,
Name = ['ANSARI ASHFAK HUSSAIN ABRAR AHMED', 'BARIA LEVIN SANTOSHBHAI', 'CHAUHAN ABHISHEK R
Gender = ['Male', 'Male', 'Male', 'Male', 'Male', 'Male', 'Male', 'Male', 'Male', 'Male', '
Mobile_no = [9825672577.0, 9884375747.0, 9876428544.0, 9828909545.0, 9831362615.0, 98533622
COA = [88.0, 91.0, 64.0, 85.0, 89.0, 74.0, 79.0, 62.0, 94.0, 70.0, 76.0, 95.0, 73.0, 78.0,
OS = [72.0, 69.0, 93.0, 62.0, 77.0, 68.0, 67.0, 91.0, 62.0, 86.0, 63.0, 94.0, 80.0, 77.0, 8
PEM = [81.0, 85.0, 86.0, 71.0, 88.0, 75.0, 85.0, 75.0, 75.0, 94.0, 70.0, 75.0, 65.0, 60.0,
OOP_1 = [92.0, 66.0, 63.0, 74.0, 90.0, 87.0, 98.0, 85.0, 67.0, 91.0, 82.0, 88.0, 69.0, 51.0
DM = [56.0, 90.0, 86.0, 77.0, 63.0, 85.0, 61.0, 68.0, 74.0, 76.0, 74.0, 82.0, 63.0, 58.0, 6
DE_1B = [73.0, 84.0, 88.0, 77.0, 88.0, 94.0, 89.0, 85.0, 86.0, 95.0, 78.0, 81.0, 87.0, 72.0
City = ['Panchmahal', 'Anand', 'Amreli', 'Ahmedabad', 'Ahmedabad', 'Porbandar', 'Gandhinaga

col_list = ['sr_no', 'Enrollment_Number', 'Name', 'Gender', '']
```

In [3]:

```
df = pd.DataFrame(columns = columns)

df['Sr. No.'] = sr_no
df['Enrollment Number'] = Enrollment_Number
df['Name'] = Name
df['Gender'] = Gender
df['Mobile Number'] = Mobile_no
df['City'] = City
df['COA'] = COA
df['OS'] = OS
df['PEM'] = PEM
df['OOP-1'] = OOP_1
df['DM'] = DM
df['DE-1B'] = DE_1B
df
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | City | COA | OS | PEM |
|-----|------------|----------------------|---|--------|------------------|------------|------|------|------|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | Panchmahal | 88.0 | 72.0 | 81.0 |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | Anand | 91.0 | 69.0 | 85.0 |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | Amreli | 64.0 | 93.0 | 86.0 |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | Ahmedabad | 85.0 | 62.0 | 71.0 |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | Ahmedabad | 89.0 | 77.0 | 88.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 71 | 73 | 201260116515 | PATEL HEET JAYESHBHAI | Male | 9.834835e+09 | Narmada | 69.0 | 64.0 | 65.0 |
| 72 | 74 | 201260116516 | PATEL ANVI DINESHBHAI | Female | 9.883401e+09 | jamnagar | 94.0 | 71.0 | 87.0 |
| 73 | 75 | 201260116517 | HIRPARA ANIKET | Male | 9.898055e+09 | Bhavnagar | 81.0 | 61.0 | 88.0 |
| 74 | 76 | 201260116518 | MEWADA HITEN | Male | 9.871237e+09 | jamnagar | 79.0 | 87.0 | 73.0 |
| 75 | 77 | 201260116519 | BAROT JAY KAMALBHAI | Male | 9.873119e+09 | kheda | 94.0 | 66.0 | 61.0 |

76 rows × 12 columns



In [4]:

```
df.to_csv('Dataset.csv', index=False)
```

In []:

In [1]:

```
# Practical 2
# Perform descriptive analysis and identify the data type.
```

In [2]:

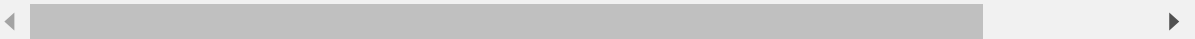
```
import pandas as pd
import numpy as np
import statistics as stats
```

In [3]:

```
df = pd.read_csv('Dataset.csv')
df.head()
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | COA | OS | PEM | OOP-1 | DM |
|---|---------|-------------------|---|--------|---------------|------|------|------|-------|------|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | 88.0 | 72.0 | 81.0 | 92.0 | 56.0 |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | 91.0 | 69.0 | 85.0 | 66.0 | 90.0 |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | 64.0 | 93.0 | 86.0 | 63.0 | 86.0 |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | 85.0 | 62.0 | 71.0 | 74.0 | 77.0 |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | 89.0 | 77.0 | 88.0 | 90.0 | 63.0 |



In [4]:

```
df.dtypes
```

Out[4]:

```
Sr. No.          int64
Enrollment Number  int64
Name             object
Gender           object
Mobile Number     float64
COA              float64
OS               float64
PEM              float64
OOP-1            float64
DM               float64
DE-1B            float64
City             object
dtype: object
```

In [5]:

```
np.mean(df['COA'])
```

Out[5]:

```
78.77631578947368
```

In [6]:

```
stats.mode(df['COA'])
```

Out[6]:

```
89.0
```

In [7]:

```
np.min(df['COA'])
```

Out[7]:

```
60.0
```

In [8]:

```
np.max(df['COA'])
```

Out[8]:

```
98.0
```

In [9]:

```
np.percentile(df['COA'], 25)
```

Out[9]:

```
69.75
```

In [10]:

```
np.percentile(df['COA'], 50)
```

Out[10]:

78.0

In [11]:

```
np.percentile(df['COA'], 75)
```

Out[11]:

89.0

In [12]:

```
df['COA'].std()
```

Out[12]:

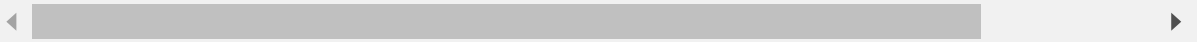
10.768594689138753

In [13]:

```
df.describe()
```

Out[13]:

| | Sr. No. | Enrollment Number | Mobile Number | COA | OS | PEM | OOP-1 | |
|-------|-----------|----------------------|------------------|-----------|-----------|-----------|-----------|------|
| count | 76.000000 | 7.600000e+01 | 7.600000e+01 | 76.000000 | 76.000000 | 76.000000 | 76.000000 | 76.0 |
| mean | 38.986842 | 1.937601e+11 | 9.864197e+09 | 78.776316 | 76.684211 | 78.789474 | 75.065789 | 74.8 |
| std | 22.520209 | 4.358899e+09 | 2.948073e+07 | 10.768595 | 10.570664 | 11.979222 | 14.501803 | 12.4 |
| min | 1.000000 | 1.912601e+11 | 9.812817e+09 | 60.000000 | 60.000000 | 60.000000 | 51.000000 | 55.0 |
| 25% | 19.750000 | 1.912601e+11 | 9.834689e+09 | 69.750000 | 68.000000 | 66.750000 | 63.000000 | 64.0 |
| 50% | 38.500000 | 1.912601e+11 | 9.869449e+09 | 78.000000 | 77.000000 | 81.000000 | 75.500000 | 76.0 |
| 75% | 58.250000 | 1.937601e+11 | 9.891906e+09 | 89.000000 | 86.250000 | 89.250000 | 89.000000 | 86.0 |
| max | 77.000000 | 2.012601e+11 | 9.909774e+09 | 98.000000 | 97.000000 | 98.000000 | 98.000000 | 95.0 |



In []:

In [1]:

```
# Practical 3
# Implement a method to find out variation in data. For example the difference
# between highest and lowest marks in each subject semester wise.
```

In [2]:

```
import pandas as pd
import numpy as np
import statistics as stats
```

In [3]:

```
df = pd.read_csv('Dataset.csv')
df.head()
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | City | COA | OS | PEM |
|---|---------|-------------------|---|--------|---------------|------------|------|------|------|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | Panchmahal | 88.0 | 72.0 | 81.0 |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | Anand | 91.0 | 69.0 | 85.0 |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | Amreli | 64.0 | 93.0 | 86.0 |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | Ahmedabad | 85.0 | 62.0 | 71.0 |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | Ahmedabad | 89.0 | 77.0 | 88.0 |

In [4]:

```
for i in list(df.columns)[5:]:
    if(df[i].dtype == 'float64'):
        print(i,"---", np.max(df[i]) - np.min(df[i]))
```

```
COA --- 38.0
OS --- 37.0
PEM --- 38.0
OOP-1 --- 47.0
DM --- 40.0
DE-1B --- 25.0
```


In [5]:

```
# Variance using numpy  
np.var(df['COA'])
```

Out[5]:

114.43680747922438

In [6]:

```
np.var(df['OS'])
```

Out[6]:

110.26869806094184

In [7]:

```
np.var(df['PEM'])
```

Out[7]:

141.6135734072022

In [8]:

```
np.var(df['OOP-1'])
```

Out[8]:

207.5351454293629

In [9]:

```
np.var(df['DM'])
```

Out[9]:

152.5152354570637

In [10]:

```
np.var(df['DE-1B'])
```

Out[10]:

54.78324099722991

In []:

In [1]:

```
# Practical 4  
# 4 Plot the graph showing result of student in each semester.
```

In [2]:

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

In [3]:

```
df = pd.read_csv('Dataset.csv')  
df
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | City | COA | OS | PEM |
|-----|------------|----------------------|---|--------|------------------|------------|------|------|------|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | Panchmahal | 88.0 | 72.0 | 81.0 |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | Anand | 91.0 | 69.0 | 85.0 |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | Amreli | 64.0 | 93.0 | 86.0 |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | Ahmedabad | 85.0 | 62.0 | 71.0 |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | Ahmedabad | 89.0 | 77.0 | 88.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 71 | 73 | 201260116515 | PATEL HEET JAYESHBHAI | Male | 9.834835e+09 | Narmada | 69.0 | 64.0 | 65.0 |
| 72 | 74 | 201260116516 | PATEL ANVI DINESHBHAI | Female | 9.883401e+09 | jamnagar | 94.0 | 71.0 | 87.0 |
| 73 | 75 | 201260116517 | HIRPARA ANIKET | Male | 9.898055e+09 | Bhavnagar | 81.0 | 61.0 | 88.0 |
| 74 | 76 | 201260116518 | MEWADA HITEN | Male | 9.871237e+09 | jamnagar | 79.0 | 87.0 | 73.0 |
| 75 | 77 | 201260116519 | BAROT JAY KAMALBHAI | Male | 9.873119e+09 | kheda | 94.0 | 66.0 | 61.0 |

76 rows × 12 columns



In [4]:

```
for i in list(df.columns)[6:11]:
    df[i].fillna(df[i].mean(), inplace=True)

df.head()
```

Out[4]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | City | COA | OS | PEM |
|---|---------|-------------------|---|--------|---------------|------------|------|------|------|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | Panchmahal | 88.0 | 72.0 | 81.0 |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | Anand | 91.0 | 69.0 | 85.0 |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | Amreli | 64.0 | 93.0 | 86.0 |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | Ahmedabad | 85.0 | 62.0 | 71.0 |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | Ahmedabad | 89.0 | 77.0 | 88.0 |



In [6]:

```
Enroll_no = float(input("Enter Enrollment number: "))
temp_df = df.loc[df['Enrollment Number'] == Enroll_no]

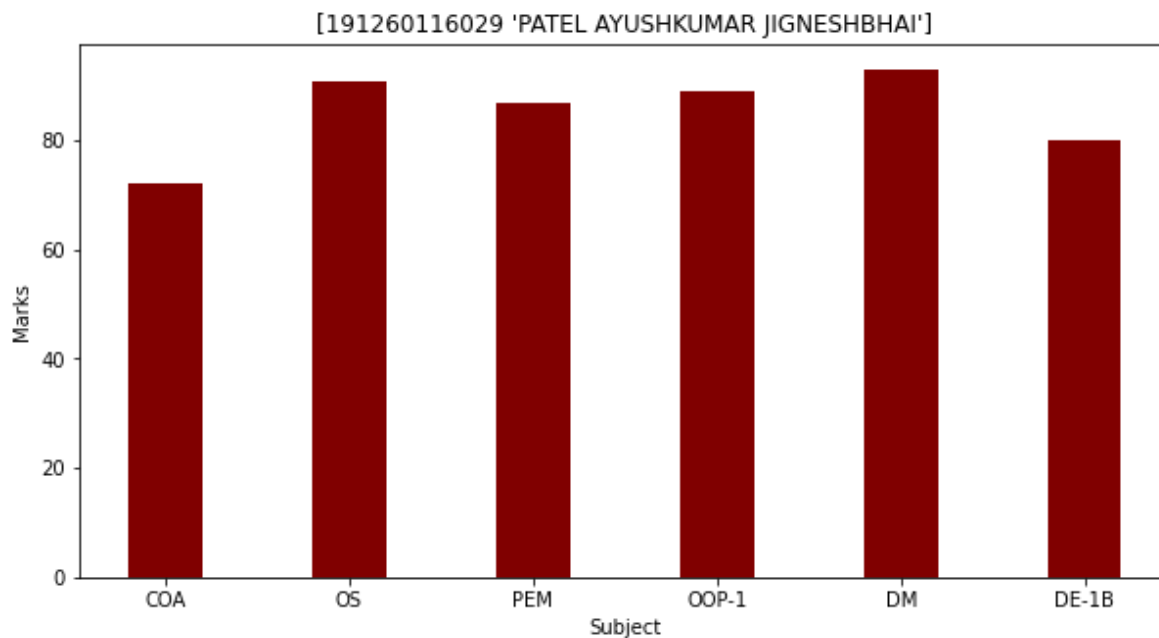
x = list(temp_df.columns)[6:]
df_list = list(temp_df.values)
y = df_list[0][6:]

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

plt.bar(x, y, color = 'maroon', width = 0.4)

plt.xlabel("Subject")
plt.ylabel("Marks")
plt.title(f"{df_list[0][1:3]}")
plt.show()
```

Enter Enrollment number: 191260116029



In []:

In []:

In [1]:

```
# Practical 5
# Plot the graph showing the geographical location of students.
```

In [2]:

```
import pandas as pd
import numpy as np
import statistics as stats
import matplotlib.pyplot as plt
```

In [3]:

```
df = pd.read_csv('Dataset.csv')
df.head()
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | COA | OS | PEM | OOP- 1 | DM | DE 1 |
|---|------------|----------------------|---|--------|------------------|------|------|------|-----------|------|---------|
| 0 | 1 | 1.912600e+11 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9825672577 | NaN | 72.0 | 81.0 | 92.0 | 56.0 | 73. |
| 1 | 2 | 1.912600e+11 | BARIA LEVIN SANTOSHBHAI | Male | 9884375747 | 91.0 | 69.0 | 85.0 | 66.0 | 90.0 | 84. |
| 2 | 3 | 1.912600e+11 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9876428544 | 64.0 | 93.0 | 86.0 | 63.0 | 86.0 | 88. |
| 3 | 4 | 1.912600e+11 | DESAI JIGAR ANILBHAI | Male | 9828909545 | 85.0 | 62.0 | 71.0 | 74.0 | 77.0 | Na |
| 4 | 5 | 1.912600e+11 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9831362615 | 89.0 | 77.0 | NaN | 90.0 | 63.0 | 88. |

In [4]:

```
City_list = list(df['City'].unique())
print(City_list)
```

```
['Panchmahal', 'Anand', 'Amreli', 'Ahmedabad', 'Porbandar', 'Gandhinagar',  
'Bhavnagar', 'Somnath', 'Lucknow', 'kheda', 'Ratlam', 'Dahod', 'Gondal', 'Su  
rat', 'junagadh', 'Narmada', 'Palanpur', 'Botad', 'Morbi', 'Patan', 'Navsar  
i', 'Bharuch', 'Vadodra', 'Rangpur', 'jamnagar', 'kutch']
```

In [5]:

```
dd = {}

for i in City_list:
    temp = 0
    for j in list(df['City']):
        if i==j:
            temp = temp + 1
    dd[i] = temp
```

In [6]:

```
city = list(dd.keys())
city_count = list(dd.values())
```

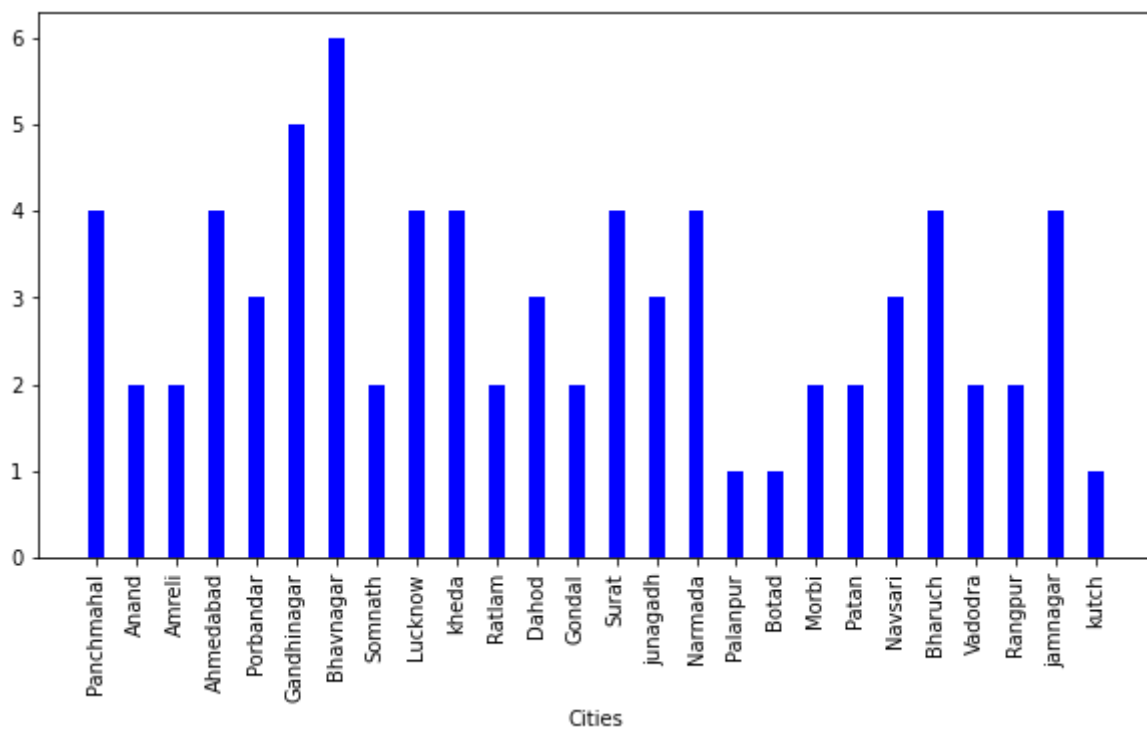
In [7]:

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

# creating the bar plot
plt.bar(city, city_count, color = 'blue',
        width = 0.4)

plt.xlabel("Cities")
plt.xticks(rotation=90)

# plt.ylabel("Male/Female")
plt.show()
```



In [1]:

```
# Practical 6
# 6 Plot the graph showing number of male and female students.
```

In [2]:

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

In [3]:

```
df = pd.read_csv('Dataset.csv')
df.head()
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | COA | OS | PEM | OOP-1 | DM | DE-1 |
|---|---------|-------------------|---|--------|---------------|------|------|------|-------|------|------|
| 0 | 1 | 1.912600e+11 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9825672577 | NaN | 72.0 | 81.0 | 92.0 | 56.0 | 73. |
| 1 | 2 | 1.912600e+11 | BARIA LEVIN SANTOSHBHAI | Male | 9884375747 | 91.0 | 69.0 | 85.0 | 66.0 | 90.0 | 84. |
| 2 | 3 | 1.912600e+11 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9876428544 | 64.0 | 93.0 | 86.0 | 63.0 | 86.0 | 88. |
| 3 | 4 | 1.912600e+11 | DESAI JIGAR ANILBHAI | Male | 9828909545 | 85.0 | 62.0 | 71.0 | 74.0 | 77.0 | Na |
| 4 | 5 | 1.912600e+11 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9831362615 | 89.0 | 77.0 | NaN | 90.0 | 63.0 | 88. |

In [4]:

```
male = 0
female = 0

for i in df['Gender']:
    if i == 'Male':
        male = male + 1
    if i == 'Female':
        female = female + 1

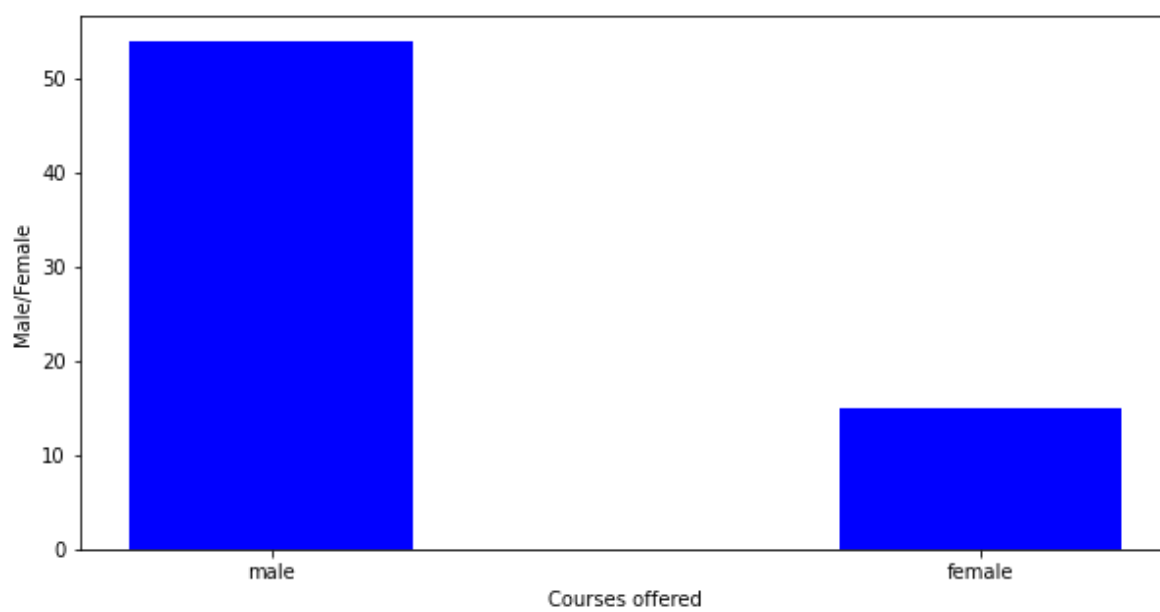
dd = {}
dd['male'] = male
dd['female'] = female
dd
```

Out[4]:

```
{'male': 54, 'female': 15}
```


In [5]:

```
fig = plt.figure(figsize = (10, 5))  
  
# creating the bar plot  
plt.bar(list(dd.keys()), list(dd.values()), color = 'blue',  
        width = 0.4)  
  
plt.xlabel("Courses offered")  
plt.ylabel("Male/Female")  
plt.show()
```



In []:

In [1]:

```
# Practical 7
# Implement a method to treat missing value for gender and missing value for marks.
```

In [2]:

```
import pandas as pd
import numpy as np
```

In [3]:

```
df = pd.read_csv('Dataset.csv')
df
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | COA | OS | PEM | OOP-1 | DM | Final |
|-----|---------|-------------------|---|--------|---------------|------|------|------|-------|------|-------|
| 0 | 1 | 1.912600e+11 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9825672577 | NaN | 72.0 | 81.0 | 92.0 | 56.0 | 72.0 |
| 1 | 2 | 1.912600e+11 | BARIA LEVIN SANTOSHBHAI | Male | 9884375747 | 91.0 | 69.0 | 85.0 | 66.0 | 90.0 | 85.0 |
| 2 | 3 | 1.912600e+11 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9876428544 | 64.0 | 93.0 | 86.0 | 63.0 | 86.0 | 86.0 |
| 3 | 4 | 1.912600e+11 | DESAI JIGAR ANILBHAI | Male | 9828909545 | 85.0 | 62.0 | 71.0 | 74.0 | 77.0 | NaN |
| 4 | 5 | 1.912600e+11 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9831362615 | 89.0 | 77.0 | NaN | 90.0 | 63.0 | 86.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 71 | 73 | 2.012600e+11 | PATEL HEET JAYESHBHAI | Male | 9834834778 | 69.0 | 64.0 | 65.0 | 57.0 | 88.0 | 86.0 |
| 72 | 74 | 2.012600e+11 | PATEL ANVI DINESHBHAI | Female | 9883401326 | NaN | 71.0 | 87.0 | 93.0 | 68.0 | 86.0 |
| 73 | 75 | 2.012600e+11 | HIRPARA ANIKET . | Male | 9898054882 | 81.0 | 61.0 | 88.0 | 82.0 | 76.0 | 90.0 |
| 74 | 76 | 2.012600e+11 | MEWADA HITEN . | Male | 9871237130 | 79.0 | 87.0 | 73.0 | NaN | 95.0 | 72.0 |
| 75 | 77 | 2.012600e+11 | BAROT JAY KAMALBHAI | Male | 9873119124 | 94.0 | 66.0 | 61.0 | 60.0 | 86.0 | 86.0 |

76 rows × 12 columns



In [4]:

```
print(df.isnull().sum())
```

| | |
|-------------------|---|
| Sr. No. | 0 |
| Enrollment Number | 0 |
| Name | 0 |
| Gender | 7 |
| Mobile Number | 0 |
| COA | 6 |
| OS | 5 |
| PEM | 7 |
| OOP-1 | 6 |
| DM | 6 |
| DE-1B | 2 |
| City | 0 |

dtype: int64

In [5]:

```
for i in list(df.columns)[5:11]:
    print(i)
    df[i] = df[i].replace(0, np.nan)
df
```

COA
OS
PEM
OOP-1
DM
DE-1B

Out[5]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | COA | OS | PEM | OOP-1 | DM | DE-1B |
|-----|---------|-------------------|---|--------|---------------|------|------|------|-------|------|-------|
| 0 | 1 | 1.912600e+11 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9825672577 | NaN | 72.0 | 81.0 | 92.0 | 56.0 | 72.0 |
| 1 | 2 | 1.912600e+11 | BARIA LEVIN SANTOSHBHAI | Male | 9884375747 | 91.0 | 69.0 | 85.0 | 66.0 | 90.0 | 81.0 |
| 2 | 3 | 1.912600e+11 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9876428544 | 64.0 | 93.0 | 86.0 | 63.0 | 86.0 | 81.0 |
| 3 | 4 | 1.912600e+11 | DESAI JIGAR ANILBHAI | Male | 9828909545 | 85.0 | 62.0 | 71.0 | 74.0 | 77.0 | NaN |
| 4 | 5 | 1.912600e+11 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9831362615 | 89.0 | 77.0 | NaN | 90.0 | 63.0 | 81.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 71 | 73 | 2.012600e+11 | PATEL HEET JAYESHBHAI | Male | 9834834778 | 69.0 | 64.0 | 65.0 | 57.0 | 88.0 | 81.0 |
| 72 | 74 | 2.012600e+11 | PATEL ANVI DINESHBHAI | Female | 9883401326 | NaN | 71.0 | 87.0 | 93.0 | 68.0 | 81.0 |
| 73 | 75 | 2.012600e+11 | HIRPARA ANIKET | Male | 9898054882 | 81.0 | 61.0 | 88.0 | 82.0 | 76.0 | 91.0 |
| 74 | 76 | 2.012600e+11 | MEWADA HITEN | Male | 9871237130 | 79.0 | 87.0 | 73.0 | NaN | 95.0 | 72.0 |
| 75 | 77 | 2.012600e+11 | BAROT JAY KAMALBHAI | Male | 9873119124 | 94.0 | 66.0 | 61.0 | 60.0 | 86.0 | 81.0 |

76 rows × 12 columns

In [6]:

```
for i in list(df.columns)[5:11]:
    df[i].fillna(df[i].mean(), inplace=True)
# df['COA'].isnull.sum
```

In [7]:

```
df.isnull().sum()
```

Out[7]:

| | |
|-------------------|---|
| Sr. No. | 0 |
| Enrollment Number | 0 |
| Name | 0 |
| Gender | 7 |
| Mobile Number | 0 |
| COA | 0 |
| OS | 0 |
| PEM | 0 |
| OOP-1 | 0 |
| DM | 0 |
| DE-1B | 0 |
| City | 0 |

dtype: int64

In [1]:

```
# Practical 8
# Implement simple regression. Implement linear regression to predict the 5th
# semester result of student.
```

In [2]:

```
import pandas as pd
import numpy as np
from sklearn import linear_model
```

In [3]:

```
df = pd.read_csv('Semester_result_dataset.csv')
df.head()
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | City | SEM_1 | SEM_2 | SEM_3 |
|---|------------|----------------------|---|--------|------------------|------------|-------|-------|-------|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | Panchmahal | 8.8 | 7.2 | |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | Anand | 9.1 | 6.9 | |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | Amreli | 6.4 | 9.3 | |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | Ahmedabad | 8.5 | 6.2 | |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | Ahmedabad | 8.9 | 7.7 | |

In [4]:

```
reg = linear_model.LinearRegression()
```

In [5]:

```
reg.fit(df[['SEM_1', 'SEM_2', 'SEM_3', 'SEM_4']], df.SEM_5)
```

Out[5]:

```
LinearRegression()
```

In [6]:

```
reg.coef_
```

Out[6]:

```
array([-4.22178281e-17,  1.11022302e-16, -9.71445147e-17,  1.00000000e+00])
```

In [7]:

```
reg.predict([[8.17, 9.57, 8.78, 7.98]])
```

Out[7]:

```
array([8.98])
```

In []:

In [1]:

```
# Practical 9
# Implement Logistic regression to classify the student as average or clever.
```

In [2]:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
import numpy as np
```

In [3]:

```
df = pd.read_csv("Dataset1.csv")
df.head()
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | City | COA | OS | PEM |
|---|------------|----------------------|---|--------|------------------|------------|------|------|------|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | Panchmahal | 88.0 | 72.0 | 81.0 |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | Anand | 91.0 | 69.0 | 85.0 |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | Amreli | 64.0 | 93.0 | 86.0 |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | Ahmedabad | 85.0 | 62.0 | 71.0 |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | Ahmedabad | 89.0 | 77.0 | 88.0 |

In [4]:

```
df['Performance'] = df['PERFORMANCE'].apply(lambda x: True if (x == 'CLEVER') else False)
df.head()
```

Out[4]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | City | COA | OS | PEM |
|---|---------|-------------------|---|--------|---------------|------------|------|------|------|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | Panchmahal | 88.0 | 72.0 | 81.0 |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | Anand | 91.0 | 69.0 | 85.0 |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | Amreli | 64.0 | 93.0 | 86.0 |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | Ahmedabad | 85.0 | 62.0 | 71.0 |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | Ahmedabad | 89.0 | 77.0 | 88.0 |

In [5]:

```
X = df[['COA', 'OS', 'PEM', 'OOP-1', 'DM', 'DE-1B']]
y = df.Performance
```

In [6]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42)
```

In [7]:

```
model = LogisticRegression()
```

In [8]:

```
model.fit(X_train, y_train)
```

Out[8]:

```
LogisticRegression()
```

In [9]:

```
model.score(X_test, y_test)
```

Out[9]:

```
0.8846153846153846
```

In [10]:

```
dd = np.array([80, 89, 75, 45, 63, 41])
result = model.predict(dd.reshape(1,-1))

if result[0] == True:
    print('CLEVER')
else:
    print('AVERAGE')
# print(dd, type(dd))
```

AVERAGE

In []:

In [1]:

```
# Practical 10  
# 10 Implement decision tree to classify the student as average or clever.
```

In [2]:

```
import pandas as pd  
import numpy as np  
from sklearn.preprocessing import LabelEncoder  
from sklearn import tree  
from sklearn.model_selection import train_test_split
```

In [3]:

```
df = pd.read_csv('Semester_result_dataset1.csv')  
df.head()
```

Out[3]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | City | SEM_1 | SEM_2 | SE |
|---|------------|----------------------|---|--------|------------------|------------|-------|-------|----|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | Panchmahal | 8.8 | 7.2 | |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | Anand | 9.1 | 6.9 | |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | Amreli | 6.4 | 9.3 | |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | Ahmedabad | 8.5 | 6.2 | |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | Ahmedabad | 8.9 | 7.7 | |

In [4]:

```
df['Performance'] = df['PERFORMANCE'].apply(lambda x: True if x=='CLEVER' else False)
df.head()
```

Out[4]:

| | Sr. No. | Enrollment Number | Name | Gender | Mobile Number | City | SEM_1 | SEM_2 | SE |
|---|------------|----------------------|---|--------|------------------|------------|-------|-------|----|
| 0 | 1 | 191260116001 | ANSARI ASHFAK HUSSAIN ABRAR AHMED | Male | 9.825673e+09 | Panchmahal | 8.8 | 7.2 | |
| 1 | 2 | 191260116002 | BARIA LEVIN SANTOSHBHAI | Male | 9.884376e+09 | Anand | 9.1 | 6.9 | |
| 2 | 3 | 191260116003 | CHAUHAN ABHISHEK RAMKUNVAR | Male | 9.876429e+09 | Amreli | 6.4 | 9.3 | |
| 3 | 4 | 191260116004 | DESAI JIGAR ANILBHAI | Male | 9.828910e+09 | Ahmedabad | 8.5 | 6.2 | |
| 4 | 5 | 191260116005 | DHARAJIYA JIGNESHKUMAR GOPALBHAI | Male | 9.831363e+09 | Ahmedabad | 8.9 | 7.7 | |

In [5]:

```
inputs = df[['SEM_1', 'SEM_2', 'SEM_3', 'SEM_4']]
target = df['Performance']
```

In [6]:

```
X = inputs
y = target
```

In [7]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42)
```

In [8]:

```
model = tree.DecisionTreeClassifier()
```

In [9]:

```
model.fit(X_train, y_train)
```

Out[9]:

```
DecisionTreeClassifier()
```

In [10]:

```
model.score(X_test, y_test)
```

Out[10]:

0.7692307692307693