```
import pandas as pd
```

series

```
mydata1 = ["ananya", "aishwitha", "monisha", "nayana", "gouthmi"]
Ser1 = pd.Series(mydata1)
print(Ser1)
0
        ananya
1
     aishwitha
2
       monisha
3
        nayana
       gouthmi
dtype: object
Ser1[1]
'aishwitha'
mydata1 = ["anany", "aishwitha", "monisha", "nayana", "monisha"]
roll = [1,2,3,4,5]
Ser2 = pd.Series(mydata1,index=roll)
print(Ser2)
1
         anany
2
     aishwitha
3
       monisha
4
        nayana
       monisha
dtype: object
Ser2[2]
'aishwitha'
Ser2.to csv("C:\my files\mydata1.csv")
```

dataframes

```
dict df = pd.DataFrame(mydict)
print(dict df)
  Names
         Age
               City
0
          19
    ram
                ckm
1
          20
    rai
              mnglr
   raju
          19
              bnglr
dict_df.to_csv("C:\my files\mydict.csv")
```

load data

```
df1 = pd.read_csv("C:\my files\sampledata1.CSV")
df1.head()
        name dept
                   sem1
                         sem2
                               sem3
  aishwitha
             ISE
                    8.3
                         8.34
                                9.0
1
     monisha
              ISE
                    NaN 8.20
                                8.9
2
                                8.0
    gouthami
              ISE
                    8.0 8.00
3
                                8.2
             ISE
                    8.0
                          NaN
        pavi
4
             ISE
                    9.0 9.00
      nayana
                                NaN
```

loading large file

```
import pandas as pd
diab df = pd.read csv("C:\my files\diabetcsvsmall.csv")
diab df.head()
   preg
        plas
              pres
                    skin
                          insu
                                mass
                                       pedi
                                             age
                                                            class
   6.0
         148 72.0
                                33.6
                                                 tested positive
0
                    35.0
                                      0.627
                                              50
                             0
                                                 tested negative
1
   1.0
          85 66.0 29.0
                             0 26.6 0.351
                                              31
          183 64.0 0.0 0 23.3 0.672
89 66.0 23.0 94 28.1 0.167
2
                                              32 tested_positive
   8.0
         183 64.0
                             0 23.3 0.672
3
   1.0
                                              21
                                                  tested negative
         137 40.0 35.0
                           168 43.1 2.288
   0.0
                                              33 tested positive
diab df.tail()
          plas
                      skin
                            insu mass
                                         pedi
     preg
                pres
                                               age
                                                              class
97
                48.0
                                        0.323
      1.0
            71
                       NaN
                              76 20.4
                                                22
                                                    tested negative
98
      6.0
            93 50.0
                     30.0
                              64 28.7
                                        0.356
                                                23
                                                    tested negative
           122 90.0
99
                             220 49.7
                                        0.325
                                                    tested positive
      NaN
                     51.0
                                                31
100
     1.0
                72.0
                               0 39.0
                                        1.222
                                                33
           163
                       0.0
                                                    tested positive
101
      1.0
           151 60.0
                       0.0
                               0 26.1
                                        0.179
                                                22
                                                    tested negative
```

access

```
diab_df.loc[12:19,"age"]
```

```
12
      57
13
      59
14
      51
15
      32
16
      31
17
      31
18
      33
19
      32
Name: age, dtype: int64
diab_df.loc[12:19]
    preq
          plas
                 pres
                        skin
                              insu
                                     mass
                                            pedi
                                                   age
                                                                   class
12
                 80.0
                                     27.1
                                           1.441
                                                    57
    10.0
            139
                         0.0
                                 0
                                                        tested negative
13
     1.0
                 60.0
                        23.0
                                     30.1
                                           0.398
            189
                               846
                                                    59
                                                        tested positive
                                     25.8
14
     5.0
            166
                 72.0
                        19.0
                               175
                                           0.587
                                                    51
                                                        tested positive
15
     7.0
            100
                  0.0
                        0.0
                                 0
                                     30.0
                                           0.484
                                                    32
                                                        tested positive
                       47.0
                                     45.8
                                                        tested_positive
16
     0.0
            118
                 84.0
                               230
                                           0.551
                                                    31
17
     7.0
                                     29.6
                                           0.254
                                                        tested positive
            107
                 74.0
                         0.0
                                                    31
                                 0
18
     1.0
                        38.0
                                     43.3
                                                        tested negative
            103
                 30.0
                                83
                                           0.183
                                                    33
19
     1.0
            115
                       30.0
                                96
                                     34.6
                                           0.529
                                                    32 tested_positive
                 70.0
diab_df.iloc[12:19 , 3:8] #datafram.ioc[row_range,column_range]
    skin
          insu
                        pedi
                 mass
                               age
                                57
12
     0.0
                 27.1
                        1.441
              0
13
    23.0
            846
                 30.1
                        0.398
                                59
14
    19.0
            175
                 25.8
                        0.587
                                51
15
     0.0
              0
                 30.0
                       0.484
                                32
16
    47.0
            230
                 45.8
                        0.551
                                31
     0.0
                        0.254
17
                 29.6
                                31
              0
18
    38.0
             83
                 43.3
                        0.183
                                33
```

skin,insu,mass,pedi,age,pre,pres ==> independent(feature)

class ==> depedent(target)

```
diab df.rename(columns ={"plas":"glucose"})
     preq
           glucose pres skin insu
                                      mass
                                              pedi
class
                    72.0
               148
                          35.0
                                       33.6
                                            0.627
                                                     50
      6.0
tested positive
      1.0
                85
                    66.0
                          29.0
                                   0
                                      26.6
                                            0.351
                                                     31
tested negative
                           0.0
                                      23.3
                                                     32
               183
                    64.0
                                   0
                                            0.672
      8.0
tested positive
                89
                    66.0
                          23.0
                                   94
                                       28.1
                                            0.167
                                                     21
      1.0
tested negative
      0.0
               137
                    40.0
                          35.0
                                  168
                                       43.1 2.288
                                                     33
```

```
tested positive
97
      1.0
                71
                   48.0
                           NaN
                                  76
                                      20.4 0.323
                                                    22
tested negative
      6.0
                93
                    50.0
                          30.0
                                  64
                                      28.7 0.356
                                                    23
tested negative
99
      NaN
               122
                    90.0
                          51.0
                                 220
                                      49.7
                                          0.325
                                                    31
tested positive
100
      1.0
               163
                   72.0
                           0.0
                                   0
                                      39.0
                                           1.222
                                                    33
tested positive
101
      1.0
               151 60.0
                           0.0
                                   0 26.1 0.179
                                                    22
tested negative
[102 rows x 9 columns]
diab df.head()
   preg
         plas pres
                     skin
                           insu
                                 mass
                                        pedi
                                              age
                                                             class
0
   6.0
          148
              72.0
                    35.0
                                 33.6
                                      0.627
                                               50
                                                   tested positive
                              0
   1.0
                     29.0
                                                   tested negative
1
           85
              66.0
                              0
                                 26.6
                                      0.351
                                               31
2
   8.0
          183
              64.0
                      0.0
                                 23.3 0.672
                                               32
                                                   tested positive
                              0
                                                   tested negative
3
   1.0
          89
                     23.0
                             94
                                 28.1
              66.0
                                       0.167
                                               21
4
   0.0
          137
              40.0
                    35.0
                            168
                                 43.1 2.288
                                               33
                                                  tested positive
diab df.rename(columns ={"plas":"glucose"},inplace = True)
#datafram.rename(colums={"old":"new"},inplace = True) if inplace is
not given it will just show but not save
diab df.head()
        glucose
                             insu
                                                                class
   preg
                  pres
                        skin
                                    mass
                                           pedi
                                                 age
   6.0
             148
                 72.0
                        35.0
                                 0
                                    33.6
                                          0.627
                                                  50
                                                      tested positive
                  66.0
                                          0.351
                                                      tested negative
   1.0
                        29.0
                                    26.6
                                                  31
1
              85
                                 0
2
   8.0
             183
                  64.0
                        0.0
                                    23.3
                                                  32
                                                      tested positive
                                 0
                                          0.672
3
   1.0
             89
                  66.0
                        23.0
                                94
                                    28.1
                                          0.167
                                                  21
                                                      tested negative
4
   0.0
             137
                  40.0
                        35.0
                               168 43.1
                                         2.288
                                                  33 tested positive
diab df["qlucose in mmol"]=diab df["qlucose"]/18.018
#dataframe["new col name"] = content
#converting glucose from mg to mmol and creating new col
diab df.head(12)
   preg glucose
                   pres skin insu mass
                                            pedi age
0
     6.0
              148
                   72.0 35.0
                                  0
                                     33.6 0.627
                                                   50
                                                       tested positive
     1.0
               85
                   66.0 29.0
                                  0
                                     26.6 0.351
                                                   31
                                                       tested negative
2
     8.0
              183
                   64.0
                          0.0
                                  0
                                     23.3 0.672
                                                   32 tested positive
```

3	1.0	89	66.0	23.0	94	28.1	0.167	21	tested_negative
4	0.0	137	40.0	35.0	168	43.1	2.288	33	tested_positive
5	5.0	116	74.0	0.0	0	25.6	0.201	30	tested_negative
6	3.0	78	50.0	32.0	88	31.0	0.248	26	tested_positive
7	10.0	115	0.0	0.0	0	35.3	0.134	29	tested_negative
8	2.0	197	70.0	45.0	543	30.5	0.158	53	tested_positive
9	8.0	125	96.0	0.0	0	0.0	0.232	54	tested_positive
10	4.0	110	92.0	0.0	0	37.6	0.191	30	tested_negative
11	10.0	168	74.0	0.0	0	38.0	0.537	34	tested_positive
0 1 2 3 4 5 6 7 8 9 10									

filter and groups

```
fil_age_30less = diab_df[diab_df["age"]<30]</pre>
#new df = your df[condition]
fil_age_30less.head()
    preg glucose
                    pres skin insu
                                              pedi
                                                                    class
                                      mass
                                                    age
     1.0
3
               89
                    66.0
                          23.0
                                  94
                                      28.1
                                             0.167
                                                     21
                                                         tested_negative
                                                         tested_positive
6
     3.0
               78
                    50.0
                          32.0
                                  88
                                      31.0
                                             0.248
                                                     26
                                                         tested_negative
    10.0
              115
                     0.0
                           0.0
                                   0
                                      35.3
                                             0.134
                                                     29
20
     3.0
              126
                    88.0
                          41.0
                                 235
                                      39.3
                                             0.704
                                                     27
                                                         tested_negative
23
     9.0
              119
                    80.0
                          35.0
                                   0
                                      29.0 0.263
                                                     29
                                                         tested_positive
```

```
glucose in mmol
3
           4.939505
6
           4.329004
7
           6.382506
20
           6.993007
23
           6.604507
glucose above100 = diab df[diab df["glucose"]>100]
glucose above100.head(7)
                                                                   class
   preg glucose pres skin insu
                                     mass
                                             pedi
                                                   age
0
    6.0
             148 72.0
                         35.0
                                  0
                                     33.6
                                           0.627
                                                    50
                                                        tested positive
2
    8.0
             183
                  64.0
                          0.0
                                     23.3
                                                    32
                                  0
                                            0.672
                                                        tested positive
    0.0
             137
                  40.0
                         35.0
                                168
                                     43.1
                                            2,288
                                                    33
                                                        tested positive
5
    5.0
             116 74.0
                          0.0
                                  0
                                     25.6
                                           0.201
                                                    30
                                                        tested negative
7
   10.0
             115
                   0.0
                          0.0
                                  0
                                     35.3
                                            0.134
                                                    29
                                                        tested negative
    2.0
             197
                         45.0
                                543
                                     30.5
                                                    53
                                                        tested positive
8
                  70.0
                                            0.158
    8.0
             125
                  96.0
                          0.0
                                  0
                                      0.0
                                           0.232
                                                    54
                                                        tested positive
   glucose in mmol
0
          8.214008
2
         10.156510
4
          7,603508
5
          6.438006
7
          6.382506
8
         10.933511
9
          6.937507
glucose below100 = diab df[diab df["glucose"]<100]</pre>
glucose below100.head(7)
                                                                    class
    preg
          glucose
                   pres skin insu
                                      mass
                                              pedi
                                                    age
/
     1.0
                   66.0
                          29.0
1
               85
                                   0
                                      26.6
                                            0.351
                                                     31
                                                         tested negative
3
     1.0
               89
                   66.0 23.0
                                  94
                                      28.1
                                            0.167
                                                     21
                                                         tested negative
     3.0
               78
                    50.0
                          32.0
                                  88
                                      31.0
                                             0.248
                                                     26
                                                         tested positive
21
     8.0
               99
                   84.0
                           0.0
                                   0
                                      35.4
                                            0.388
                                                     50
                                                         tested negative
```

```
27
     1.0
                97
                    66.0 15.0
                                  140 23.2 0.487
                                                      22 tested negative
32
     3.0
                88
                    58.0
                         11.0
                                   54
                                       24.8
                                             0.267
                                                      22
                                                          tested negative
     6.0
33
                92 92.0
                           0.0
                                    0
                                      19.9
                                             0.188
                                                      28
                                                         tested negative
    glucose in mmol
           \overline{4.717505}
1
3
           4.939505
6
           4.329004
21
           5.494505
27
           5.383505
32
           4.884005
33
           5.106005
```

create a filter data set which has only the rows with age btw 20 and 30

```
age 20 and 30 = diab df[(diab df['age']>20)&(diab df['age']<30)]
age 20and30.head()
   preg glucose
                   pres skin insu mass
                                            pedi age
                                                                 class
3
                   66.0
     1.0
               89
                        23.0
                                 94
                                     28.1 0.167
                                                   21
                                                       tested negative
6
     3.0
               78
                   50.0 32.0
                                 88
                                    31.0
                                          0.248
                                                   26
                                                       tested positive
   10.0
              115
                   0.0 0.0
7
                                  0
                                    35.3 0.134
                                                   29
                                                       tested negative
20
     3.0
              126
                       41.0
                                235 39.3
                                           0.704
                   88.0
                                                   27
                                                       tested negative
23
     9.0
              119
                   80.0 35.0
                                  0 29.0 0.263
                                                   29
                                                       tested positive
   glucose in mmol
3
           4.939505
6
           4.329004
7
           6.382506
20
           6.993007
23
           6.604507
#group by class and calucate the average
grouped by class age = diab df.groupby("class")["age"].mean()
grouped_by_class_age
class
tested negative
                   31.238095
tested positive
                   40.589744
Name: age, dtype: float64
```

```
grouped by class insu = diab df.groupby("class")["insu"].mean()
grouped by class insu
#Result:
# the average insul in level of diabetic people is 114.6
# the average insul in level of non-diabetic people is 114.6
class
tested negative
                    52.571429
tested positive
                   114.692308
Name: insu, dtype: float64
grouped by class glucose= diab df.groupby("class")["glucose"].mean()
grouped by class glucose
class
tested negative
                   107.793651
tested_positive
                   140.666667
Name: glucose, dtype: float64
grouped min age= diab df.groupby("class")["age"].min()
grouped min age
class
tested negative
                   21
tested positive
                   25
Name: age, dtype: int64
grouped max age= diab df.groupby("class")["age"].max()
grouped max age
class
tested negative
                   60
tested positive
                   60
Name: age, dtype: int64
```

cleaning data

handel null

```
diab df.isnull()
    preg glucose pres skin insu
                                 mass
                                      pedi
                                             age
class
          False False False False False
   False
                                           False
                                                False
1
   False
          False
                False False False False
                                           False
                                                 False
2 False
          False
                False False False False
                                           False
                                                 False
          False False False False
3
   False
                                           False False
```

```
4 False
           False False False False False False
97
    False False False False False False
98
    False
           False False False False
                                               False False
99 True
           False False False False False False
100 False
           False
                 False False False False
                                               False
                                                     False
101 False False False False False False False
    glucose in mmol
0
            False
1
            False
2
            False
3
            False
4
            False
. .
              . . .
97
            False
98
            False
99
            False
100
            False
101
            False
[102 rows x 10 columns]
diab df.isnull().sum()
preg
                1
                0
glucose
                1
pres
skin
                1
insu
                0
                1
mass
pedi
                1
                0
age
class
                0
glucose in mmol
                0
dtype: int64
diab df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 102 entries, 0 to 101
Data columns (total 10 columns):
                  Non-Null Count
    Column
                                Dtype
    -----
```

```
0
                       101 non-null
                                        float64
     preq
                                        int64
 1
     glucose
                       102 non-null
 2
     pres
                       101 non-null
                                        float64
 3
     skin
                       101 non-null
                                        float64
 4
     insu
                       102 non-null
                                        int64
 5
                       101 non-null
                                        float64
     mass
 6
     pedi
                       101 non-null
                                        float64
 7
                       102 non-null
                                        int64
     age
 8
     class
                       102 non-null
                                        object
9
     glucose in mmol 102 non-null
                                        float64
dtypes: float64(6), int64(3), object(1)
memory usage: 8.1+ KB
diab_df.dropna(inplace=True)
diab df.info()
<class 'pandas.core.frame.DataFrame'>
Index: 98 entries, 0 to 101
Data columns (total 10 columns):
                       Non-Null Count
#
     Column
                                        Dtype
0
                       98 non-null
                                        float64
     preg
     glucose
                       98 non-null
                                        int64
 1
 2
                       98 non-null
                                        float64
     pres
 3
     skin
                       98 non-null
                                        float64
 4
                       98 non-null
                                        int64
     insu
 5
                       98 non-null
                                        float64
     mass
 6
                                        float64
     pedi
                       98 non-null
 7
     age
                       98 non-null
                                        int64
 8
                       98 non-null
                                        object
     class
 9
     glucose in mmol 98 non-null
                                        float64
dtypes: float64(6), int64(3), object(1)
memory usage: 8.4+ KB
diab df.isnull().sum()
preg
                    0
                    0
glucose
                    0
pres
                    0
skin
                    0
insu
mass
                    0
                    0
pedi
                    0
age
                    0
class
glucose in mmol
                    0
dtype: int64
```

```
diab df.info()
<class 'pandas.core.frame.DataFrame'>
Index: 98 entries, 0 to 101
Data columns (total 10 columns):
#
     Column
                       Non-Null Count
                                       Dtvpe
- - -
     _ _ _ _ _ _
                       _____
 0
                       98 non-null
                                       float64
     preg
 1
     alucose
                       98 non-null
                                        int64
 2
                       98 non-null
                                       float64
     pres
 3
     skin
                       98 non-null
                                       float64
 4
                       98 non-null
                                       int64
     insu
 5
                       98 non-null
                                       float64
     mass
 6
                       98 non-null
                                       float64
     pedi
7
                       98 non-null
                                       int64
     age
 8
     class
                       98 non-null
                                       object
9
     glucose in mmol 98 non-null
                                       float64
dtypes: float64(6), int64(3), object(1)
memory usage: 8.4+ KB
diab df.drop duplicates(inplace = True)
diab df.info()
<class 'pandas.core.frame.DataFrame'>
Index: 96 entries, 0 to 101
Data columns (total 10 columns):
                       Non-Null Count
#
     Column
                                       Dtype
0
                       96 non-null
                                        float64
     preg
                       96 non-null
                                        int64
 1
     glucose
 2
                       96 non-null
                                       float64
     pres
 3
                       96 non-null
                                       float64
     skin
 4
     insu
                       96 non-null
                                       int64
 5
                       96 non-null
                                       float64
     mass
 6
     pedi
                       96 non-null
                                       float64
 7
                       96 non-null
                                       int64
     age
 8
                       96 non-null
                                       object
     class
 9
     glucose in mmol 96 non-null
                                       float64
dtypes: float64(6), int64(3), object(1)
memory usage: 8.2+ KB
```

reading other formats(excle)

```
dia ex = pd.read excel("C:\my files\diabetes.xlsx")
dia ex.head()
         plas
                pres
                      skin
                            insu
                                  mass
                                          pedi
                                                age
                                                                class
   preg
0
          148
                                   33.6
                                                  50
      6
                  72
                        35
                               0
                                         0.627
                                                      tested positive
1
      1
           85
                  66
                        29
                               0
                                  26.6
                                         0.351
                                                  31
                                                      tested negative
```

```
2
      8
          183
                  64
                         0
                               0
                                  23.3
                                        0.672
                                                 32
                                                     tested positive
3
           89
                        23
                                         0.167
      1
                  66
                              94
                                  28.1
                                                 21
                                                     tested negative
4
      0
          137
                 40
                        35
                             168 43.1 2.288
                                                 33
                                                     tested positive
dia ex sheet2 = pd.read excel("C:\my files\")
diabetes.xlsx", sheet name="dora")
dia ex sheet2.head()
  Dead Alive
  yes
          no
1
  yes
          no
  yes
          no
  yes
          no
4 yes
          no
```

text file

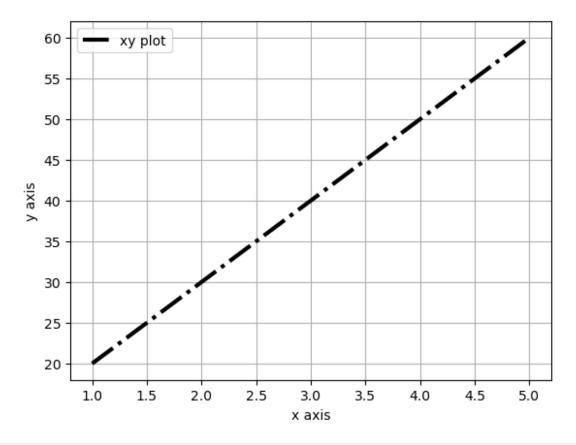
```
df_txt = pd.read_csv("C:\my files\grades.txt")
df txt.head()
  Names Initials SEM1 SEM2 SEM3 Grade
                   Joe K 9.8 10 9.9 A+
0
1
               Rajesh M 8.9 9.1 9.3 A
2
               Kissan V 9.9 9.3 9.2 A
3
                    Mary N 7.7 8 7.1 B
4
                Jeen K 9.8 9.1 9.9 A+
df_txt = pd.read_csv("C:\my files\grades.txt",sep = ' ')
df txt.head()
    Names Initials
                    SEM1
                           SEM2
                                 SEM3 Grade
0
      Joe
                      9.8
                           10.0
                                  9.9
                                          A+
                 K
1
  Rajesh
                 М
                     8.9
                            9.1
                                  9.3
                                           Α
2
                            9.3
  Kissan
                 ٧
                      9.9
                                  9.2
                                           Α
3
                      7.7
                                  7.1
                                           В
     Mary
                            8.0
4
     Jeen
                 K
                      9.8
                            9.1
                                  9.9
                                          A+
```

modifying datatype

```
df txt['SEM1 INT'] = df_txt['SEM1'].astype(int)
df txt.head()
    Names Initials SEM1
                            SEM2
                                  SEM3 Grade
                                               SEM1 INT
                      9.8
                                   9.9
      Joe
                  K
                            10.0
                                           Α+
                                                       9
   Rajesh
                      8.9
                             9.1
                                   9.3
                                                       8
1
                                            Α
                                                       9
2
                  ٧
                      9.9
                             9.3
  Kissan
                                   9.2
                                            Α
3
     Mary
                      7.7
                             8.0
                                   7.1
                                            В
                                                       7
                  N
4
                  Κ
                      9.8
                                                       9
     Jeen
                             9.1
                                   9.9
                                           Α+
```

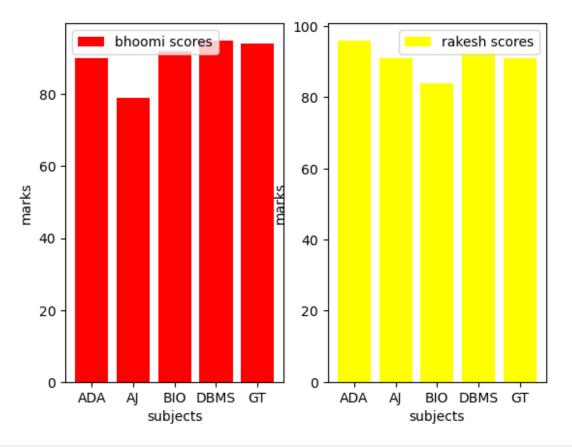
matplotlib

```
x=[1,2,3,4,5]
y=[20,30,40,50,60]
import matplotlib.pyplot as plt
plt.plot(x,y,color='k',label="xy plot",linestyle = '-.',linewidth=3)
plt.xlabel("x axis")
plt.ylabel("y axis")
plt.grid()
plt.legend()
<matplotlib.legend.Legend at 0x14b0d5e9f50>
```



```
import matplotlib.pyplot as plt
sub = ["ADA", "AJ", "BIO", "DBMS", "GT"]
bhoomi = [90,79,92,95,94]
rakesh = [96,91,84,93,91]
#firstplot
plt.subplot(1,2,1)
plt.bar(sub, bhoomi , color = 'red', label = "bhoomi scores")
plt.xlabel("subjects")
plt.ylabel("marks")
plt.legend()
#second plot
plt.subplot(1,2,2)
```

```
plt.bar(sub, rakesh , color = 'yellow', label = "rakesh scores" )
plt.xlabel("subjects")
plt.ylabel("marks")
plt.legend()
plt.show()
```



```
import numpy as np
a =np.array([25,60,5,10])
labe = ["AIML","PYTHON","PANDAS","NUMPY"]
color = ["pink","black","coral","yellow"]
plt.pie(a,labels = labe,colors=color)
plt.legend()
plt.show()
```

