

```
In [1]: import numpy as np
```

```
In [2]: import pandas as pd
```

```
In [4]: import warnings
warnings.filterwarnings('ignore')
```

```
In [5]: data = pd.read_csv(r'C:\Users\lenovo\Desktop\NIT FILES\17th - ML\TITANIC PROJECT\Titanic Dataset.csv')
```

```
In [6]: data.tail()
```

```
Out[6]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
--	-------------	----------	--------	------	-----	-----	-------	-------	--------	------

886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0
-----	-----	---	---	-----------------------	------	------	---	---	--------	------

887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
-----	-----	---	---	------------------------------	--------	------	---	---	--------	------

888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4
-----	-----	---	---	--	--------	-----	---	---	------------	------

889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0
-----	-----	---	---	-----------------------	------	------	---	---	--------	------

890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7
-----	-----	---	---	---------------------	------	------	---	---	--------	-----



```
In [7]: data.describe()
```

```
Out[7]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch
--	-------------	----------	--------	-----	-------	-------

count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204160
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910452
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454269
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.001754
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329160



In [8]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   PassengerId  891 non-null    int64
1   Survived     891 non-null    int64
2   Pclass       891 non-null    int64
3   Name         891 non-null    object
4   Sex          891 non-null    object
5   Age          714 non-null    float64
6   SibSp        891 non-null    int64
7   Parch        891 non-null    int64
8   Ticket       891 non-null    object
9   Fare         891 non-null    float64
10  Cabin        204 non-null    object
11  Embarked     889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

In [9]: `data.head(2)`

Out[9]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
--	-------------	----------	--------	------	-----	-----	-------	-------	--------	------

0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833

In [10]: `del data['Name']`

In [11]: `del data['Ticket']`
`del data['Fare']`
`del data['Cabin']`

In [12]: `data.head(2)`

Out[12]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Embarked
0	1	0	3	male	22.0	1	0	S
1	2	1	1	female	38.0	1	0	C

In [13]: `# Convert male as 1 and female as 0`

```
In [14]: def getvalue(str):
        if str == 'male':
            return 1
        else:
            return 0
        data['Gender'] = data['Sex'].apply(getvalue)
        data.head()
```

```
Out[14]:
```

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Embarked	Gender
0	1	0	3	male	22.0	1	0	S	1
1	2	1	1	female	38.0	1	0	C	0
2	3	1	3	female	26.0	0	0	S	0
3	4	1	1	female	35.0	1	0	S	0
4	5	0	3	male	35.0	0	0	S	1

```
In [15]: # delet the attribute - sex
```

```
In [16]: del data['Sex']
```

```
In [17]: data.head(3)
```

```
Out[17]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Embarked	Gender
0	1	0	3	22.0	1	0	S	1
1	2	1	1	38.0	1	0	C	0
2	3	1	3	26.0	0	0	S	0

```
In [18]: data.head()
```

```
Out[18]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Embarked	Gender
0	1	0	3	22.0	1	0	S	1
1	2	1	1	38.0	1	0	C	0
2	3	1	3	26.0	0	0	S	0
3	4	1	1	35.0	1	0	S	0
4	5	0	3	35.0	0	0	S	1

```
In [19]: data.dtypes
```

```
Out[19]: PassengerId      int64
Survived      int64
Pclass        int64
Age           float64
SibSp         int64
Parch         int64
Embarked      object
Gender        int64
dtype: object
```

```
In [20]: data.isnull().sum()
```

```
Out[20]: PassengerId      0
Survived      0
Pclass        0
Age           177
SibSp         0
Parch         0
Embarked      2
Gender        0
dtype: int64
```

```
In [22]: # calculate the average age of survive passenger
meanS = data[data.Survived==1].Age.mean()
meanS
```

```
Out[22]: 28.343689655172415
```

```
In [23]: # calculate the average age of not survive passenger
meanNS = data[data.Survived==0].Age.mean()
meanNS
```

```
Out[23]: 30.62617924528302
```

```
In [24]: # fill the age where passenger are survive
data.Age.fillna(meanS,inplace=True)
# fill the age where the passenger are not survive
data.Age.fillna(meanNS,inplace=False)
data.head()
```

```
Out[24]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Embarked	Gender
0	1	0	3	22.0	1	0	S	1
1	2	1	1	38.0	1	0	C	0
2	3	1	3	26.0	0	0	S	0
3	4	1	1	35.0	1	0	S	0
4	5	0	3	35.0	0	0	S	1

```
In [25]: data.isnull().sum()
```

```
Out[25]: PassengerId    0
         Survived      0
         Pclass       0
         Age          0
         SibSp        0
         Parch        0
         Embarked     2
         Gender       0
         dtype: int64
```

```
In [26]: # Lets check wheather the Embarked is affect to reach tour target
         data[['Embarked', 'Survived']].value_counts() # valuecount
```

```
Out[26]: Embarked  Survived
         S          0          427
                   1          217
         C          1           93
                   0           75
         Q          0           47
                   1           30
         Name: count, dtype: int64
```

```
In [27]: # Lets drop the missing value at embarked attribute
         data.dropna(inplace=True)
```

```
In [28]: data.isnull().sum() # check missing value
```

```
Out[28]: PassengerId    0
         Survived      0
         Pclass       0
         Age          0
         SibSp        0
         Parch        0
         Embarked     0
         Gender       0
         dtype: int64
```

```
In [29]: # Lets rename some columns
         # rename gender to sex
         data.rename(columns={'Gender': 'Sex'}, inplace=True)
         data.head()
```

```
Out[29]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Embarked	Sex
0	1	0	3	22.0	1	0	S	1
1	2	1	1	38.0	1	0	C	0
2	3	1	3	26.0	0	0	S	0
3	4	1	1	35.0	1	0	S	0
4	5	0	3	35.0	0	0	S	1

```
In [30]: data.columns
```

```
Out[30]: Index(['PassengerId', 'Survived', 'Pclass', 'Age', 'SibSp', 'Parch',
               'Embarked', 'Sex'],
              dtype='object')
```

```
In [31]: # change embarked to numeric value s==1,c==2,q==3
def getemb(str):
    if str == 'S':
        return 1
    elif str == 'C':
        return 2
    else:
        return 3
data['Embarked']=data['Embarked'].apply(getemb)
```

```
In [32]: data.head()
```

```
Out[32]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Embarked	Sex
0	1	0	3	22.0	1	0	1	1
1	2	1	1	38.0	1	0	2	0
2	3	1	3	26.0	0	0	1	0
3	4	1	1	35.0	1	0	1	0
4	5	0	3	35.0	0	0	1	1

```
In [33]: data.dtypes
```

```
Out[33]: PassengerId      int64
Survived      int64
Pclass        int64
Age           float64
SibSp         int64
Parch         int64
Embarked      int64
Sex           int64
dtype: object
```

```
In [34]: data.rename(columns={'Embarked':'Embark'},inplace=True) # rename the embarked co
```

```
In [35]: data.head()
```

```
Out[35]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Embark	Sex
0	1	0	3	22.0	1	0	1	1
1	2	1	1	38.0	1	0	2	0
2	3	1	3	26.0	0	0	1	0
3	4	1	1	35.0	1	0	1	0
4	5	0	3	35.0	0	0	1	1

```
In [36]: data=data.rename(columns={'Survived':'Survive'}) # rename the survived to surviv
```

```
In [37]: data.head()
```

Out[37]:

	PassengerId	Survive	Pclass	Age	SibSp	Parch	Embark	Sex
0	1	0	3	22.0	1	0	1	1
1	2	1	1	38.0	1	0	2	0
2	3	1	3	26.0	0	0	1	0
3	4	1	1	35.0	1	0	1	0
4	5	0	3	35.0	0	0	1	1

In [38]: *#successfully done EDA*

In [39]: *#now - Visualization*

In [40]: `import matplotlib.pyplot as plt`
`from matplotlib import style`

In [41]: `male = (data['Sex']==1).sum()`
`print('Male numbers:-',male)`

Male numbers:- 577

In [42]: `female = (data['Sex']==0).sum()`
`print('Female numbers:-',female)`

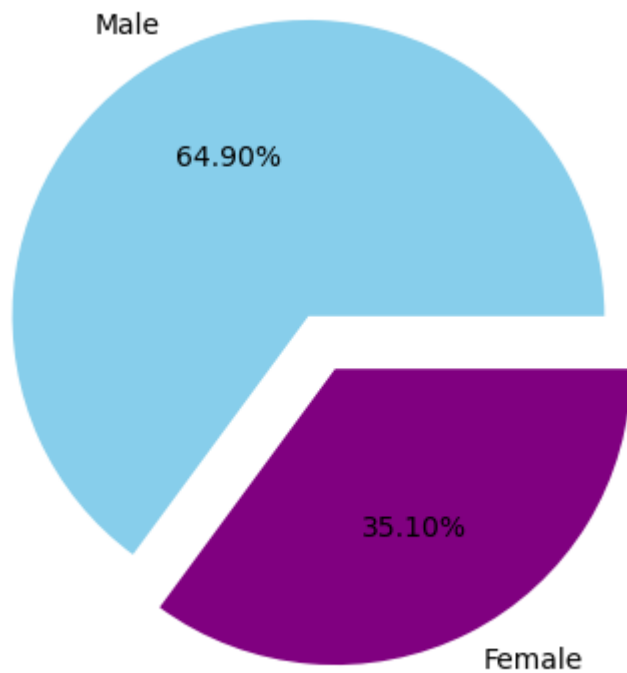
Female numbers:- 312

In [43]: `p=[male,female]`
`p`

Out[43]: [577, 312]

In [47]: `plt.pie(p,labels=["Male",'Female'],colors=['skyblue','purple'],explode=(0.2,0),a`

Out[47]: (`<matplotlib.patches.Wedge at 0x2611bc16840>`,
`<matplotlib.patches.Wedge at 0x2611c35a510>`],
`[Text(-0.586705655416278, 1.1600760638434688, 'Male'),`
`Text(0.4964432468906962, -0.9816028232521663, 'Female')]`,
`[Text(-0.3610496341023248, 0.7138929623652115, '64.90%'),`
`Text(0.27078722557674334, -0.5354197217739088, '35.10%')]`)



```
In [51]: # check totalnumber of male survive
maleS=((data['Sex']==1)&(data['Survive']==1)).sum()
maleS
```

Out[51]: 109

```
In [52]: # check total number of male not survive
maleN=((data['Sex']==1)&(data['Survive']==0)).sum()
maleN
```

Out[52]: 468

```
In [49]: # check total number of female survive
femaleS=((data['Sex']==0) & (data['Survive']==1)).sum()
femaleS
```

Out[49]: 231

```
In [50]: # check total number of female not survive
femaleN=((data['Sex']==0) & (data['Survive']==0)).sum()
femaleN
```

Out[50]: 81

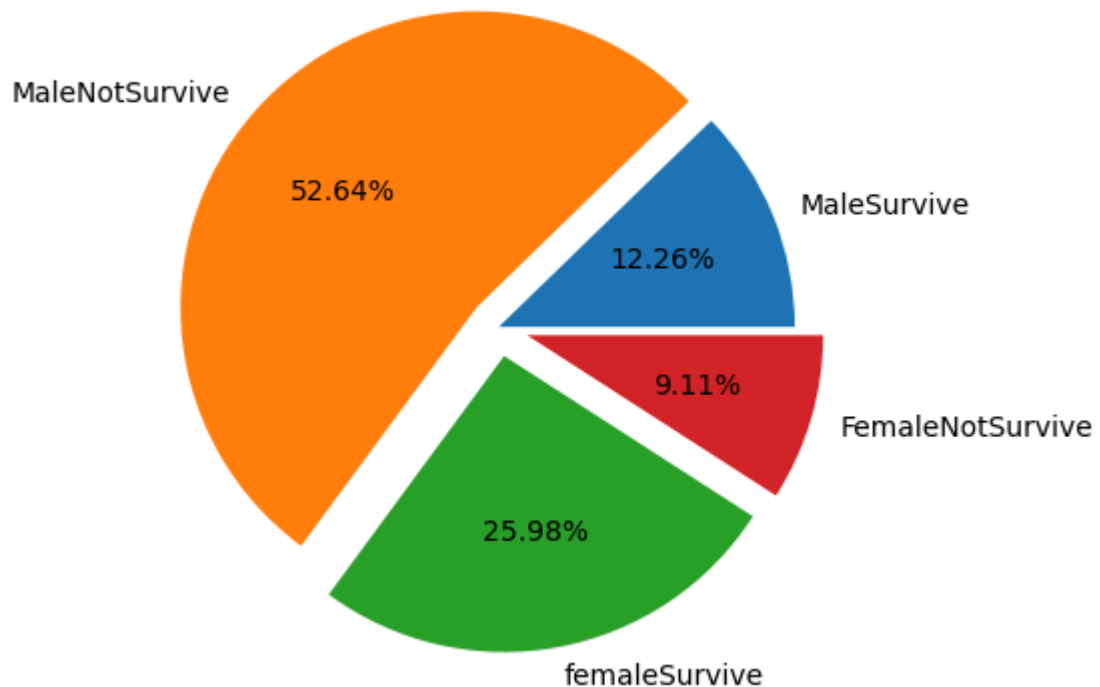
```
In [53]: # bring into one list to do pie chart
pc=[maleS,maleN,femaleS,femaleN]
pc
```

Out[53]: [109, 468, 231, 81]

```
In [55]: #plot piechart
plt.pie(pc,autopct="%.2f%",explode=(0,0.1,0.1,0.1),labels=['MaleSurvive','MaleN',
```



```
Out[55]: ([<matplotlib.patches.Wedge at 0x2611c3bc500>,  
<matplotlib.patches.Wedge at 0x2611c3bc590>,  
<matplotlib.patches.Wedge at 0x2611c3bde50>,  
<matplotlib.patches.Wedge at 0x2611c3be4b0>],  
[Text(1.0193999086759307, 0.4133083911457694, 'MaleSurvive'),  
Text(-0.9042435214866014, 0.7888876053339351, 'MaleNotSurvive'),  
Text(0.21718824283848098, -1.1801818788528882, 'femaleSurvive'),  
Text(1.1511740756789812, -0.33881890071931253, 'FemaleNotSurvive')],  
[Text(0.5560363138232348, 0.22544094062496509, '12.26%'),  
Text(-0.5274753875338507, 0.4601844364447954, '52.64%'),  
Text(0.12669314165578055, -0.6884394293308514, '25.98%'),  
Text(0.6715182108127391, -0.1976443587529323, '9.11%')])
```



At the end we conclude that from all the passengers who travel in the taitanic

- Most of the male are not able to survive around - 52.64%