In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
titanic=pd.read_csv(r"C:\Users\Jan Saida\OneDrive\Documents\Desktop\Excel sheets\titanic dataset.csv")
```

In [3]:
```
titanic
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [4]:
```
titanic.head()
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [5]:
```
titanic.tail()
```

Out[5]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

In [6]:
```
titanic.isnull().sum()
```

Out[6]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [7]:
```
titanic.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 [8]:
```
titanic.isnull().any()
```

Out[8]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age             True
SibSp          False
Parch          False
Ticket         False
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

In [9]:
```
titanic.describe()
```

Out[9]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
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	32.204208
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	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [10]:
```
titanic.transpose()
```

Out[10]:

	0	1	2	3	4	5	6	7	8	9	...	881	882	883	884	885	886
PassengerId	1	2	3	4	5	6	7	8	9	10	...	882	883	884	885	886	887
Survived	0	1	1	1	0	0	0	0	1	1	...	0	0	0	0	0	0
Pclass	3	1	3	1	3	3	1	3	3	2	...	3	3	2	3	3	2
Name	Braund, Mr. Owen Harris	Cumings, Mrs. John Bradley (Florence Briggs Th...	Heikkinen, Miss. Laina	Futrelle, Mrs. Jacques Heath (Lily May Peel)	Allen, Mr. William Henry	Moran, Mr. James	McCarthy, Mr. Timothy J	Palsson, Master. Gosta Leonard	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	Nasser, Mrs. Nicholas (Adele Achem)	...	Markun, Mr. Johann	Dahlberg, Miss. Gerda Ulrika	Banfield, Mr. Frederick James	Sutehall, Mr. Henry Jr	Rice, Mrs. William (Margaret Norton)	Montvila, Rev. Juozas
Sex	male	female	female	female	male	male	male	male	female	female	...	male	female	male	male	female	male
Age	22.0	38.0	26.0	35.0	35.0	NaN	54.0	2.0	27.0	14.0	...	33.0	22.0	28.0	25.0	39.0	27.0
SibSp	1	1	0	1	0	0	0	3	0	1	...	0	0	0	0	0	0
Parch	0	0	0	0	0	0	0	1	2	0	...	0	0	0	0	5	0
Ticket	A/5 21171	PC 17599	STON/O2. 3101282	113803	373450	330877	17463	349909	347742	237736	...	349257	7552	C.A./SOTON 34068	SOTON/OQ 392076	382652	211536
Fare	7.25	71.2833	7.925	53.1	8.05	8.4583	51.8625	21.075	11.1333	30.0708	...	7.8958	10.5167	10.5	7.05	29.125	13.0
Cabin	NaN	C85	NaN	C123	NaN	NaN	E46	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
Embarked	S	C	S	S	S	Q	S	S	S	C	...	S	S	S	S	Q	S

12 rows × 891 columns

In [11]:
```
del titanic['Name']
titanic.head()
```

Out[11]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	male	35.0	0	0	373450	8.0500	NaN	S

In [12]:
```
del titanic['Ticket']
titanic.tail()
```

Out[12]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Cabin	Embarked
886	887	0	2	male	27.0	0	0	13.00	NaN	S
887	888	1	1	female	19.0	0	0	30.00	B42	S
888	889	0	3	female	NaN	1	2	23.45	NaN	S
889	890	1	1	male	26.0	0	0	30.00	C148	C
890	891	0	3	male	32.0	0	0	7.75	NaN	Q

In [13]:
```
del titanic['Fare']
```

In [14]:
```
del titanic['Cabin']
titanic.head()
```

Out[14]:

	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
2	3	1	3	female	26.0	0	0	S
3	4	1	1	female	35.0	1	0	S
4	5	0	3	male	35.0	0	0	S

In [15]:
```
titanic.isnull().sum()
```

Out[15]:
```
PassengerId      0
Survived         0
Pclass           0
Sex              0
Age            177
SibSp            0
Parch            0
Embarked         2
dtype: int64
```

In [16]:
```
def age_filter(str):
    if str == 'male':
        return 1
    else:
        return 2
```

```
titanic['Gender']=titanic['Sex'].apply(age_filter)
titanic.head()
```

Out[16]:

	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	2
2	3	1	3	female	26.0	0	0	S	2
3	4	1	1	female	35.0	1	0	S	2
4	5	0	3	male	35.0	0	0	S	1

In [17]:
```
mean_1=titanic[titanic.Survived == 1].Age.mean()
mean_1
```

Out[17]:
```
28.343689655172415
```

In [18]:
```
titanic['age']=np.where(pd.isnull(titanic.Age)&titanic.Survived==1,mean_1,titanic.Age)
titanic.head()
```

Out[18]:

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

In [19]:
```
titanic.isnull().sum()
```

Out[19]:
```
PassengerId      0
Survived         0
Pclass           0
Sex              0
Age            177
SibSp            0
Parch            0
Embarked         2
Gender           0
age            125
dtype: int64
```

In [20]:
```
mean_0=titanic[titanic.Survived==0].Age.mean()
```

```
mean_0
```

Out[20]:
```
30.62617924528302
```

In [21]:
```
titanic['age'].fillna(mean_0,inplace=True)
titanic.head()
```

```
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\1432602589.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.


  titanic['age'].fillna(mean_0,inplace=True)
```

Out[21]:

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

In [22]:
```
titanic.isnull().sum()
```

Out[22]:
```
PassengerId      0
Survived         0
Pclass           0
Sex              0
Age            177
SibSp            0
Parch            0
Embarked         2
Gender           0
age              0
dtype: int64
```

In [23]:
```
del titanic['Age']
titanic.head()
```

Out[23]:

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

In [24]:
```
Survived_C=titanic[titanic.Embarked=='C'][titanic.Survived==1].shape[0]
Survived_Q=titanic[titanic.Embarked=='Q'][titanic.Survived==1].shape[0]
Survived_S=titanic[titanic.Embarked=='S'][titanic.Survived==1].shape[0]
print(Survived_C)
print(Survived_Q)
print(Survived_S)
```

```
93
30
217
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\1746179403.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  Survived_C=titanic[titanic.Embarked=='C'][titanic.Survived==1].shape[0]
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\1746179403.py:2: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  Survived_Q=titanic[titanic.Embarked=='Q'][titanic.Survived==1].shape[0]
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\1746179403.py:3: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  Survived_S=titanic[titanic.Embarked=='S'][titanic.Survived==1].shape[0]
```

In [25]:
```
Survived_C=titanic[titanic.Embarked=='C'][titanic.Survived==0].shape[0]
Survived_Q=titanic[titanic.Embarked=='Q'][titanic.Survived==0].shape[0]
Survived_S=titanic[titanic.Embarked=='S'][titanic.Survived==0].shape[0]
print(Survived_C)
print(Survived_Q)
print(Survived_S)
```

```
75
47
427
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\3810841169.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  Survived_C=titanic[titanic.Embarked=='C'][titanic.Survived==0].shape[0]
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\3810841169.py:2: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  Survived_Q=titanic[titanic.Embarked=='Q'][titanic.Survived==0].shape[0]
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\3810841169.py:3: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  Survived_S=titanic[titanic.Embarked=='S'][titanic.Survived==0].shape[0]
```

In [26]:
```
def emb_filter(str):
    if str=='C':
        return 1
    elif str=='Q':
        return 2
```

```
    else:
        return 3

titanic['Embark']=titanic['Embarked'].apply(emb_filter)
titanic.head()
```

Out[26]:

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

In [27]:
```
del titanic['Embarked']
titanic.head()
```

Out[27]:

	PassengerId	Survived	Pclass	Sex	SibSp	Parch	Gender	age	Embark
0	1	0	3	male	1	0	1	22.0	3
1	2	1	1	female	1	0	2	38.0	1
2	3	1	3	female	0	0	2	26.0	3
3	4	1	1	female	1	0	2	35.0	3
4	5	0	3	male	0	0	1	35.0	3

In [28]:
```
titanic.rename(columns={'Embark':'Embarked'},inplace=True)
titanic.head()
```

Out[28]:

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

In [29]:
```
titanic.dropna(inplace=True)
titanic.head()
```

Out[29]:

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

In [30]:
```
titanic.isnull().sum()
```

Out[30]:
```
PassengerId    0
Survived       0
Pclass         0
Sex            0
SibSp          0
Parch          0
Gender         0
age            0
Embarked       0
dtype: int64
```

In [31]:
```
titanic
```

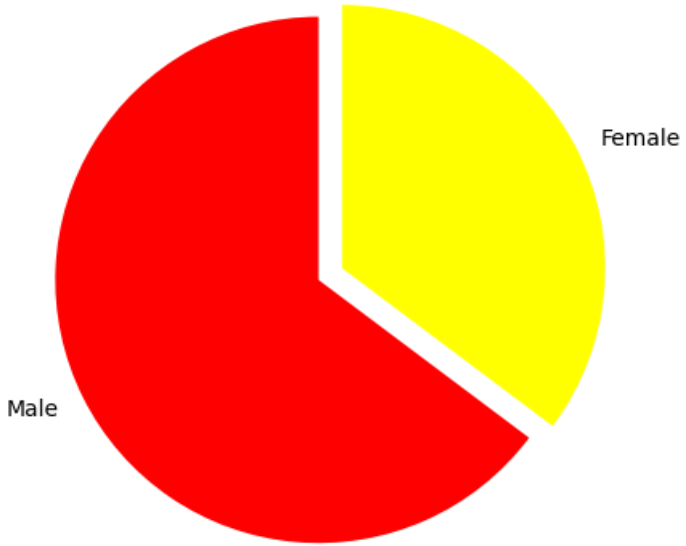
Out[31]:

	PassengerId	Survived	Pclass	Sex	SibSp	Parch	Gender	age	Embarked
0	1	0	3	male	1	0	1	22.000000	3
1	2	1	1	female	1	0	2	38.000000	1
2	3	1	3	female	0	0	2	26.000000	3
3	4	1	1	female	1	0	2	35.000000	3
4	5	0	3	male	0	0	1	35.000000	3
...	...	...	...	...	...	...	...	...	...
886	887	0	2	male	0	0	1	27.000000	3
887	888	1	1	female	0	0	2	19.000000	3
888	889	0	3	female	1	2	2	30.626179	3
889	890	1	1	male	0	0	1	26.000000	1
890	891	0	3	male	0	0	1	32.000000	2

891 rows × 9 columns

In [32]:
```
male = (titanic.Gender ==1).sum()
female = (titanic.Gender ==2).sum()
print('Male count:',male)
print('Female count:',female)
plt.pie([male,female],labels=["Male","Female"],colors=["red","yellow"],explode=(0.10,0),startangle=90)
plt.axis("equal")
plt.show()
```

```
Male count: 577
Female count: 314
```



In [33]:
```
males_survived=titanic[titanic.Gender==1][titanic.Survived==1].shape[0]
print('males_survived:',males_survived)

males_not_survived=titanic[titanic.Gender==1][titanic.Survived==0].shape[0]
print('males_not_survived:',males_not_survived)

females_survived=titanic[titanic.Gender==2][titanic.Survived==1].shape[0]
print('females_survived:',females_survived)

females_not_survived=titanic[titanic.Gender==2][titanic.Survived==0].shape[0]
print('females_not_survived:',females_not_survived)
```

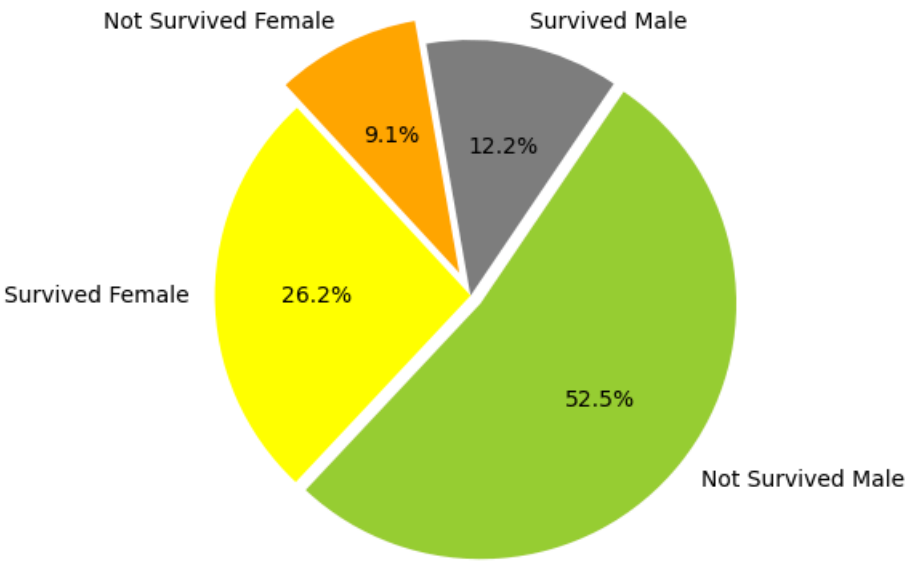
```
males_survived: 109
males_not_survived: 468
females_survived: 233
females_not_survived: 81
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\2127450764.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  males_survived=titanic[titanic.Gender==1][titanic.Survived==1].shape[0]
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\2127450764.py:4: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  males_not_survived=titanic[titanic.Gender==1][titanic.Survived==0].shape[0]
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\2127450764.py:7: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  females_survived=titanic[titanic.Gender==2][titanic.Survived==1].shape[0]
C:\Users\Jan Saida\AppData\Local\Temp\ipykernel_7516\2127450764.py:10: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  females_not_survived=titanic[titanic.Gender==2][titanic.Survived==0].shape[0]
```

In [34]:
```
ch = [males_survived, males_not_survived, females_survived, females_not_survived]
colors = ['grey', 'Yellowgreen', 'Yellow', 'Orange']
```

```
labels = ['Survived Male', 'Not Survived Male', 'Survived Female', 'Not Survived Female']
explode = [0, 0.05, 0, 0.1]
```

In [35]:
```
# Create a pie chart

plt.pie(ch, labels=labels, colors=colors, explode=explode, startangle=100, counterclock=False, autopct='%.1f%%')
plt.axis('equal')
plt.show()
```



In [ ]: