```
In [3]: import pandas as pd
```

```
In [4]: import numpy as np
```

```
In [5]: titanic = pd.read_csv(r'C:\Users\hp\Documents\Gen  AI and Data Science\11Nov cla
```

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

Out[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
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]: titanic.describe()
```

Out[7]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329

```
In [8]: del titanic["Name"]
        titanic.head()
```

Out[8]:

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

```python
In [9]: del titanic["Ticket"]
        titanic.head()
```

Out[9]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Cabin	Embarked
0	1	0	3	male	22.0	1	0	7.2500	NaN	S
1	2	1	1	female	38.0	1	0	71.2833	C85	C
2	3	1	3	female	26.0	0	0	7.9250	NaN	S
3	4	1	1	female	35.0	1	0	53.1000	C123	S
4	5	0	3	male	35.0	0	0	8.0500	NaN	S

```python
In [10]: del titanic["Fare"]
         titanic.head()
```

Out[10]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Cabin	Embarked
0	1	0	3	male	22.0	1	0	NaN	S
1	2	1	1	female	38.0	1	0	C85	C
2	3	1	3	female	26.0	0	0	NaN	S
3	4	1	1	female	35.0	1	0	C123	S
4	5	0	3	male	35.0	0	0	NaN	S

```python
In [11]: del titanic["Cabin"]
         titanic.head()
```

Out[11]:

	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 [12]:
```python
def getNumber(str):
    if str =="male":
        return 1
    else:
        return 2
titanic["Gender"]=titanic["Sex"].apply(getNumber)

titanic.head()
```

Out[12]:

	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 [14]:
```python
del titanic["Sex"]
titanic.head()
```

Out[14]:

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

In [20]:
```python
titanic.isnull()
```

Out[20]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Embarked	Gender
0	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
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
886	False	False	False	False	False	False	False	False
887	False	False	False	False	False	False	False	False
888	False	False	False	True	False	False	False	False
889	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False

891 rows × 8 columns

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

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

In [24]: `meanS = titanic[titanic.Survived==1].Age.mean()`

In [26]: `meanS`

Out[26]: `28.343689655172415`

In [28]: `titanic["age"] = np.where(pd.isnull(titanic.Age) & titanic["Survived"]==1, meanS`
`titanic.head()`

Out[28]:

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

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

```
Out[30]:  PassengerId      0
          Survived         0
          Pclass           0
          Age            177
          SibSp            0
          Parch            0
          Embarked         2
          Gender           0
          age            125
          dtype: int64
```

```
In [32]:  meaNS = titanic[titanic.Survived==0].Age.mean()
          meaNS
```

```
Out[32]:  30.62617924528302
```

```
In [36]:  titanic.age.fillna(meaNS, inplace = True)
          titanic.head()
```

C:\Users\hp\AppData\Local\Temp\ipykernel_14236\3513729142.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(meaNS, inplace = True)

Out[36]:

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

```
In [38]:  titanic.isnull().sum()
```

```
Out[38]:  PassengerId      0
          Survived         0
          Pclass           0
          Age            177
          SibSp            0
          Parch            0
          Embarked         2
          Gender           0
          age              0
          dtype: int64
```

```
In [40]:  del titanic["Age"]
          titanic.head()
```

Out[40]:

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

In [42]:
```python
survivedQ = titanic[titanic.Embarked == 'Q'][titanic.Survived == 1].shape[0]
survivedC = titanic[titanic.Embarked == 'C'][titanic.Survived == 1].shape[0]
survivedS = titanic[titanic.Embarked == 'S'][titanic.Survived == 1].shape[0]
print(survivedQ)
print(survivedC)
print(survivedS)
```

```
30
93
217
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\1434170922.py:1: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  survivedQ = titanic[titanic.Embarked == 'Q'][titanic.Survived == 1].shape[0]
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\1434170922.py:2: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  survivedC = titanic[titanic.Embarked == 'C'][titanic.Survived == 1].shape[0]
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\1434170922.py:3: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  survivedS = titanic[titanic.Embarked == 'S'][titanic.Survived == 1].shape[0]
```

In [44]:
```python
survivedQ = titanic[titanic.Embarked == 'Q'][titanic.Survived == 0].shape[0]
survivedC = titanic[titanic.Embarked == 'C'][titanic.Survived == 0].shape[0]
survivedS = titanic[titanic.Embarked == 'S'][titanic.Survived == 0].shape[0]
print(survivedQ)
print(survivedC)
print(survivedS)
```

```
47
75
427
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\3240960939.py:1: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  survivedQ = titanic[titanic.Embarked == 'Q'][titanic.Survived == 0].shape[0]
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\3240960939.py:2: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  survivedC = titanic[titanic.Embarked == 'C'][titanic.Survived == 0].shape[0]
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\3240960939.py:3: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  survivedS = titanic[titanic.Embarked == 'S'][titanic.Survived == 0].shape[0]
```

In [46]:
```python
titanic.dropna(inplace=True)
titanic.head()
```

Out[46]:

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

In [48]:
```python
titanic.isnull().sum()
```

Out[48]:
```
PassengerId    0
Survived       0
Pclass         0
SibSp          0
Parch          0
Embarked       0
Gender         0
age            0
dtype: int64
```

In [50]:
```python
titanic.rename(columns = {'age':'Age'}, inplace=True)
titanic.head()
```

Out[50]:

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

In [52]:
```python
titanic.rename(columns={'Gender':'Sex'}, inplace=True)
titanic.head()
```

Out[52]:

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

In [54]:
```python
def getEmb(str):
    if str=="S":
        return 1
    elif str=='Q':
        return 2
    else:
```

```
        return 3
titanic["Embark"]=titanic["Embarked"].apply(getEmb)
titanic.head()
```

Out[54]:

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

In [56]:
```python
del titanic['Embarked']
titanic.rename(columns={'Embark':'Embarked'}, inplace=True)
titanic.head()
```

Out[56]:

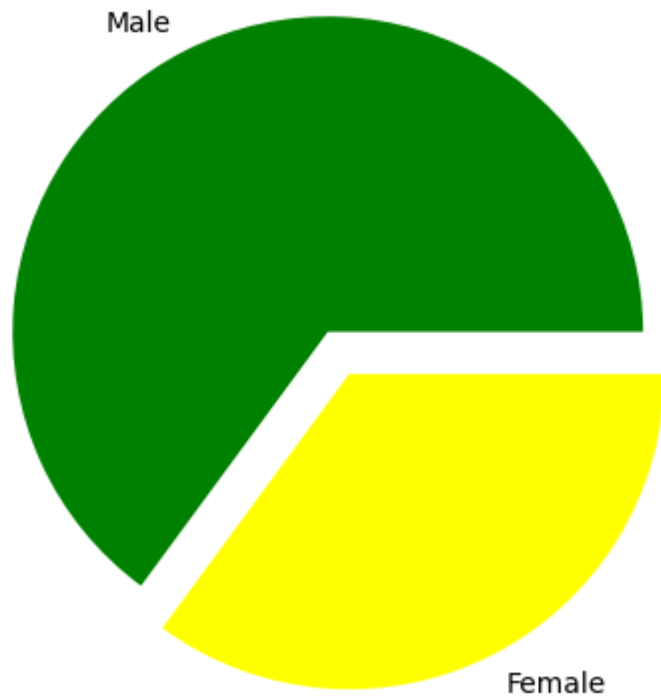|	PassengerId	Survived	Pclass	SibSp	Parch	Sex	Age	Embarked
0	1	0	3	1	0	1	22.0	1
1	2	1	1	1	0	2	38.0	3
2	3	1	3	0	0	2	26.0	1
3	4	1	1	1	0	2	35.0	1
4	5	0	3	0	0	1	35.0	1

In [58]:
```python
import matplotlib.pyplot as plt
from matplotlib import style

males = (titanic['Sex'] == 1).sum()

females = (titanic['Sex'] == 2).sum()
print(males)
print(females)
p = [males, females]
plt.pie(p,
        labels = ['Male', 'Female'],
        colors = ['green', 'yellow'],
        explode = (0.15, 0),
        startangle = 0)
plt.axis('equal')
plt.show()
```

```
577
312
```

```
In [60]: MaleS=titanic[titanic.Sex==1][titanic.Survived==1].shape[0]
         print(MaleS)
         MaleN=titanic[titanic.Sex==1][titanic.Survived==0].shape[0]
         print(MaleN)
         FemaleS=titanic[titanic.Sex==2][titanic.Survived==1].shape[0]
         print(FemaleS)
         FemaleN=titanic[titanic.Sex==2][titanic.Survived==0].shape[0]
         print(FemaleN)
```

```
109
468
231
81
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\1193566603.py:1: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  MaleS=titanic[titanic.Sex==1][titanic.Survived==1].shape[0]
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\1193566603.py:3: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  MaleN=titanic[titanic.Sex==1][titanic.Survived==0].shape[0]
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\1193566603.py:5: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  FemaleS=titanic[titanic.Sex==2][titanic.Survived==1].shape[0]
C:\Users\hp\AppData\Local\Temp\ipykernel_14236\1193566603.py:7: UserWarning: Bool
ean Series key will be reindexed to match DataFrame index.
  FemaleN=titanic[titanic.Sex==2][titanic.Survived==0].shape[0]
```
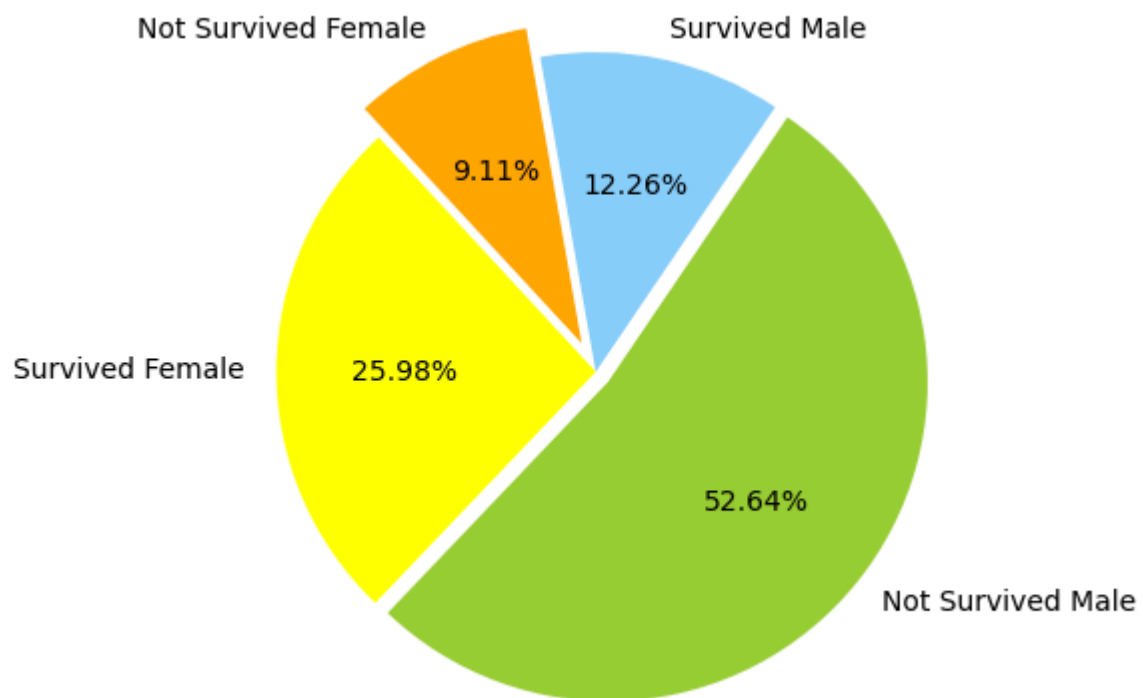
```
In [62]: chart=[MaleS,MaleN,FemaleS,FemaleN]
         colors=['lightskyblue','yellowgreen','Yellow','Orange']
         labels=["Survived Male","Not Survived Male","Survived Female","Not Survived Fema
         explode=[0,0.05,0,0.1]
         plt.pie(chart,labels=labels,colors=colors,explode=explode,startangle=100,counter
         plt.axis("equal")
         plt.show()
```

Not Survived Female

Survived Male

9.11%

12.26%

Survived Female

25.98%

52.64%

Not Survived Male

In [ ]: