In [97]:
```
# here we implimenting the all requared librarys

import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
```

In [98]:
```
# now we are putting the data set in a variable

titanic_1=pd.read_csv("C:\\Users\\prajwal patil\\Downloads\\train.csv")
titanic_2=pd.read_csv("C:\\Users\\prajwal patil\\Downloads\\test.csv")
titanic_1.head()
#titanic_2.head()
```

Out[98]:

	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 [99]:
```
titanic_1.shape
```

Out[99]:
```
(891, 12)
```

In [100]:
```
titanic_1.describe().transpose()
```

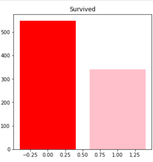
Out[100]:

	count	mean	std	min	25%	50%	75%	max
PassengerId	891.0	446.000000	257.353842	1.00	223.5000	446.0000	668.5	891.0000
Survived	891.0	0.383838	0.486592	0.00	0.0000	0.0000	1.0	1.0000
Pclass	891.0	2.308642	0.836071	1.00	2.0000	3.0000	3.0	3.0000
Age	714.0	29.699118	14.526497	0.42	20.1250	28.0000	38.0	80.0000
SibSp	891.0	0.523008	1.102743	0.00	0.0000	0.0000	1.0	8.0000
Parch	891.0	0.381594	0.806057	0.00	0.0000	0.0000	0.0	6.0000
Fare	891.0	32.204208	49.693429	0.00	7.9104	14.4542	31.0	512.3292

In [101]:
```
titanic_1['Survived'].value_counts()
```

Out[101]:
```
0    549
1    342
Name: Survived, dtype: int64
```

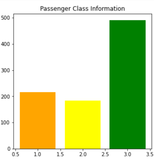
In [102]:
```
# creating a bar plot to see survived peoples.

plt.figure(figsize=(5,5))
plt.bar(list(titanic_1['Survived'].value_counts().keys()),list(titanic_1['Survived'].value_counts()),color=["red","pink"])
plt.title("Survived")
plt.show()
```



In [103]:
```
titanic_1['Pclass'].value_counts()
```

Out[103]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

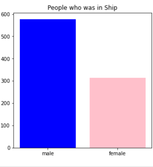
In [104]:
```
# barplot for passenger class

plt.figure(figsize=(5,5))
plt.bar(list(titanic_1['Pclass'].value_counts().keys()),list(titanic_1['Pclass'].value_counts()),color=["green","Orange","yellow"])
plt.title("Passenger Class Information")
plt.show()
```



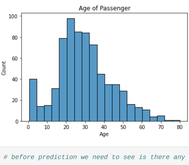
In [105]:
```
titanic_1['Sex'].value_counts()
```

Out[105]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [106]:
```
# bar plot to see the how many Male and Female in the ship
plt.figure(figsize=(5,5))
plt.bar(list(titanic_1['Sex'].value_counts().keys()),list(titanic_1['Sex'].value_counts()),color=["blue","pink"])
plt.title("People who was in Ship")
plt.show()
```



In [107]:
```
# making an a histplot to see the Age of peoples

sns.histplot(titanic_1['Age'])
plt.title("Age of Passenger")
plt.show()
```



In [108]:
```
# before prediction we need to see is there any null values or not

titanic_1['Survived'].isnull()
```

Out[108]:
```
0      False
1      False
2      False
3      False
4      False
       ...
886    False
887    False
888    False
889    False
890    False
Name: Survived, Length: 891, dtype: bool
```

In [109]:
```
sum(titanic_1['Survived'].isnull())
```

Out[109]:
```
0
```

In [110]:
```
titanic_1['Age'].isnull()
```

Out[110]:
```
0      False
1      False
2      False
3      False
4      False
       ...
886    False
887    False
888     True
889    False
890    False
Name: Age, Length: 891, dtype: bool
```

In [111]:
```
# here are so many null values in a Age column

sum(titanic_1['Age'].isnull())
```

Out[111]:
```
177
```

In [112]:
```
# droping the null values

titanic_1=titanic_1.dropna()
```

In [113]:
```
# by using the sum opretor checking the null values
sum(titanic_1['Age'].isnull())
```

Out[113]:
```
0
```

In [114]:
```
# for prediction putting the independent variable 'Age' in a X, and Dependent variable 'Survived' in Y axis

x_train=titanic_1[['Age']]
y_train=titanic_1[['Survived']]
```

In [115]:
```
# making the priction using the Decision Tree Classifier

from sklearn.tree import DecisionTreeClassifier
```

In [116]:
```
dt = DecisionTreeClassifier()
```

In [117]:
```
dt.fit(x_train,y_train)
```

Out[117]:
```
DecisionTreeClassifier()
```

In [118]:
```
# so there are some null values in test datasets Age colomn

sum(titanic_2['Age'].isnull())
```

Out[118]:
```
86
```

In [119]:
```
titanic_2=titanic_2.dropna()
```

In [120]:
```
sum(titanic_2['Age'].isnull())
```

Out[120]:
```
0
```

In [121]:
```
x_test_set=titanic_2[['Age']]
```

In [122]:
```
# storing the survived peoples in y_pred

y_pred=dt.predict(x_test_set)
```

In [123]:
```
y_pred
```

Out[123]:
```
array([1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0,
       1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1,
       1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1],
      dtype=int64)
```

In [124]:
```
# we are pridicting that the 1st person will servive and 2nd person will not servived
```