In [9]:
```
#Importing Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
#Reading the Given Csv
#Load Dataset
df=pd.read_csv("E:/Student/Titanic-Dataset.csv")
df.head()
```

Out[9]:

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

In [11]:
```
#EXPLORE AND CLEAN DATA:
#Check missing values(.isnull())
print("CHECKING FOR MISSING VALUES")
print(df.isnull().sum())

#Remove Duplicates
print("CHECKING FOR DUPLICATES")
print(df.duplicated().sum())
print("If there exists any duplicates, Remove them")
df.drop_duplicates(inplace=True)

#Fill Missing Age with median
df['Age'].fillna(df['Age'].median(),inplace=True)
```

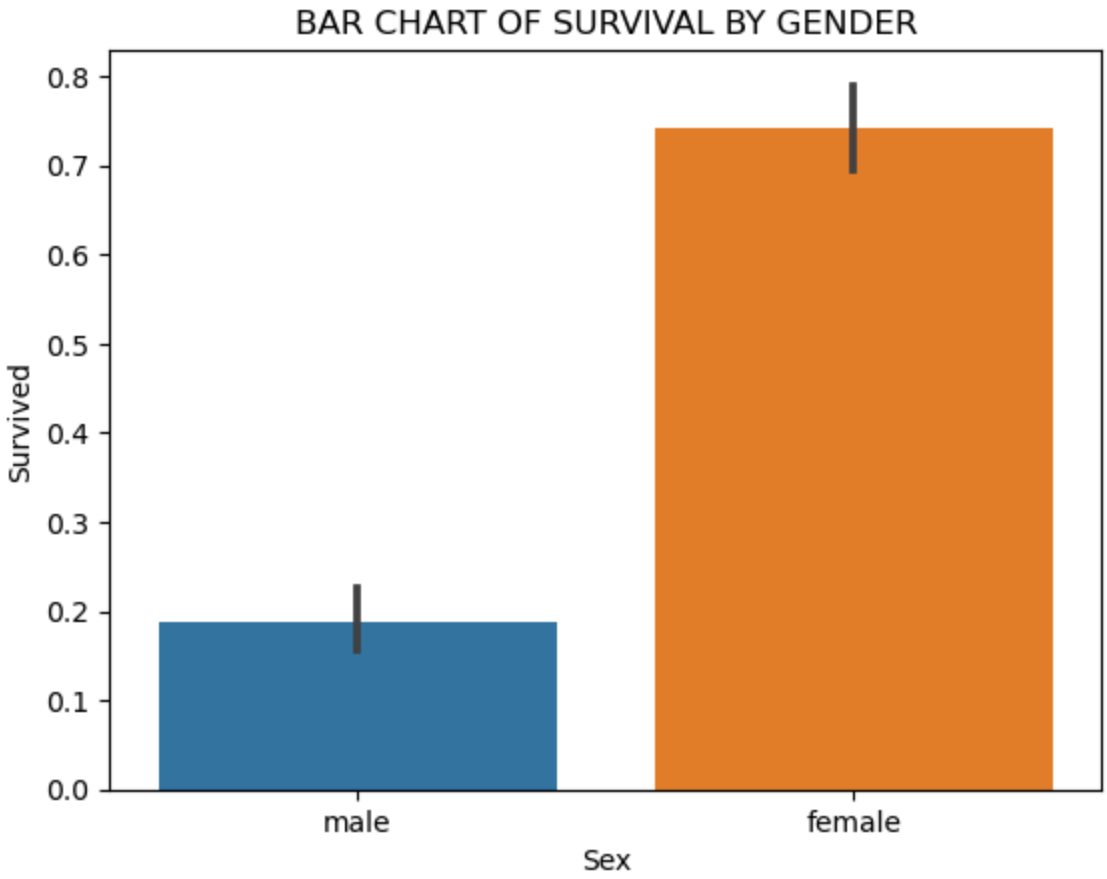
```
CHECKING FOR MISSING VALUES
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
CHECKING FOR DUPLICATES
0
If there exists any duplicates, Remove them
```

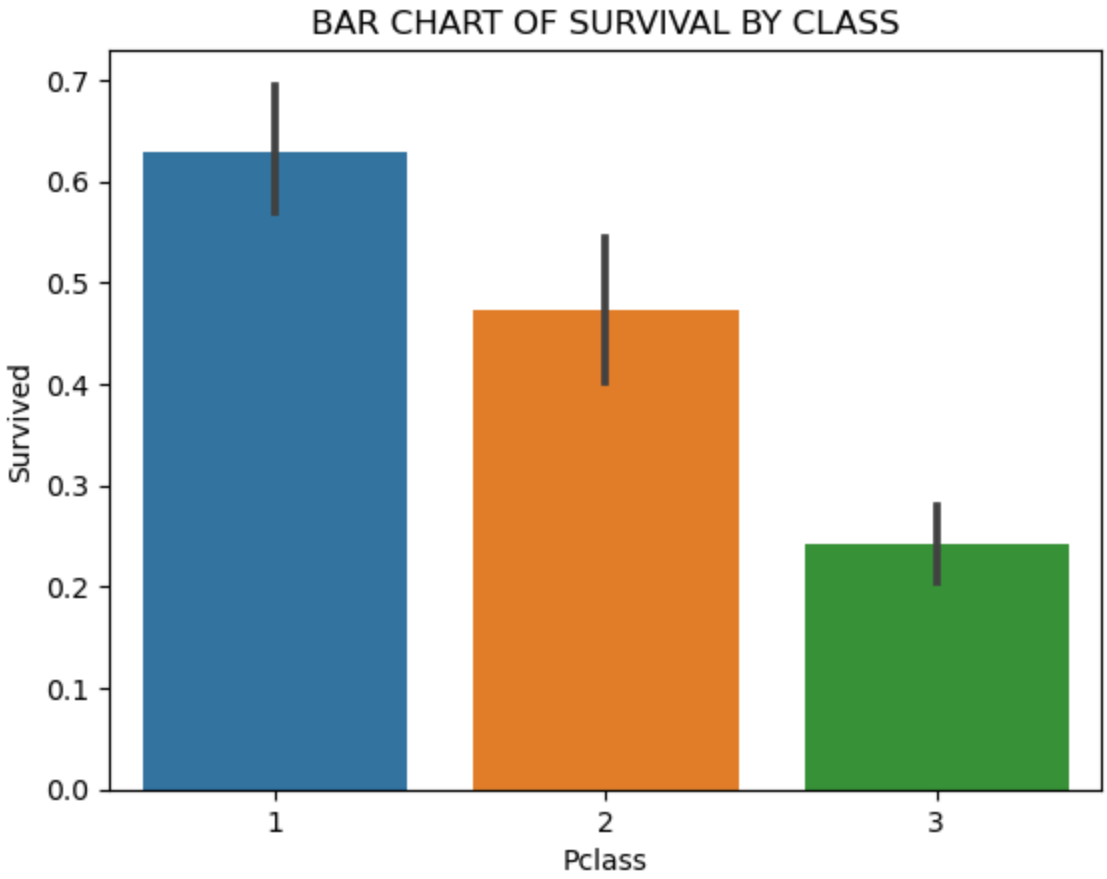
In [21]:
```
#ANALYSIS QUESTIONS

#Who survived more: Males or Females?
print("SURVIVAL RATES BY SEX")
print(df.groupby('Sex')['Survived'].mean())

#Did passenger class affect survival chances?
print("SURVIVAL RATES BY PASSENGER CLASS")
print(df.groupby('Pclass')['Survived'].mean())

#What was the Survival rate by Age Group?
print("SURVIVAL RATES BY AGE GROUP")
bins=[0,12,20,40,60,100]
labels=['Child','Teen','Adult','Middle Age','Senior']
df['AgeGroup']=pd.cut(df['Age'],bins=bins,labels=labels)
print(df.groupby('AgeGroup')['Survived'].mean())
```

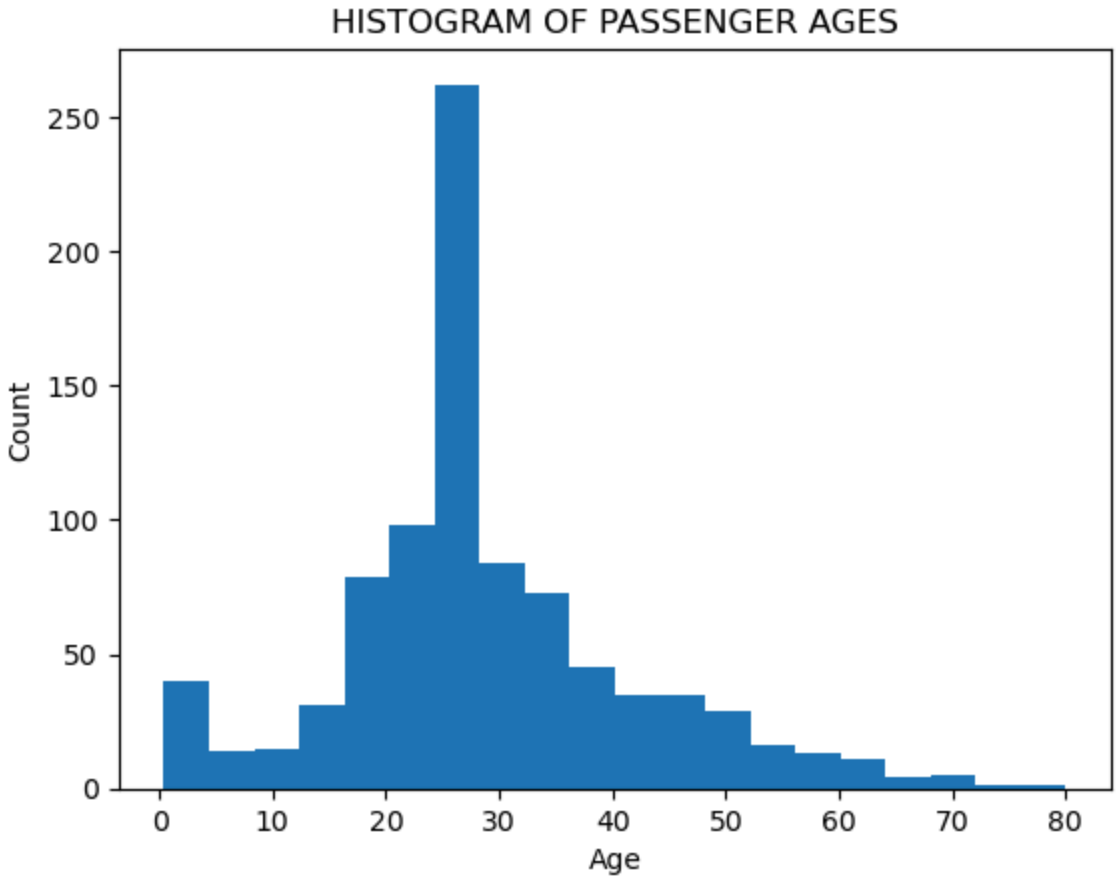
```
SURVIVAL RATES BY SEX
Sex
female    0.742038
male      0.188908
Name: Survived, dtype: float64
SURVIVAL RATES BY PASSENGER CLASS
Pclass
1    0.629630
2    0.472826
3    0.242363
Name: Survived, dtype: float64
SURVIVAL RATES BY AGE GROUP
AgeGroup
Child         0.579710
Teen          0.381818
Adult         0.364769
Middle Age    0.390625
Senior        0.227273
Name: Survived, dtype: float64
```

In [22]:
```
#VISUALISATION
#Bar chart of survival by Gender
sns.barplot(x="Sex",y="Survived",data=df)
plt.title("BAR CHART OF SURVIVAL BY GENDER")
plt.show()
```

```
#Bar Chart of survival by Class
sns.barplot(x="Pclass",y="Survived",data=df)
plt.title("BAR CHART OF SURVIVAL BY CLASS")
plt.show()

#Histogram of Passenger Ages
df["Age"].plot(kind="hist",bins=20,title="Passenger Ages")
plt.title("HISTOGRAM OF PASSENGER AGES")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()
```




BAR CHART OF SURVIVAL BY CLASS
Survived
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1
2
3
Pclass


HISTOGRAM OF PASSENGER AGES
Count
0
50
100
150
200
250
0
10
20
30
40
50
60
70
80
Age