In [41]:
```
#SANAKALACHERUVU BHASKAR
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [42]:
```
import os
```

In [43]:
```
df=pd.read_csv("C:/Users/Dell/Downloads/Titanic-Dataset.csv")
```

In [44]:
```
df
```

Out[44]:

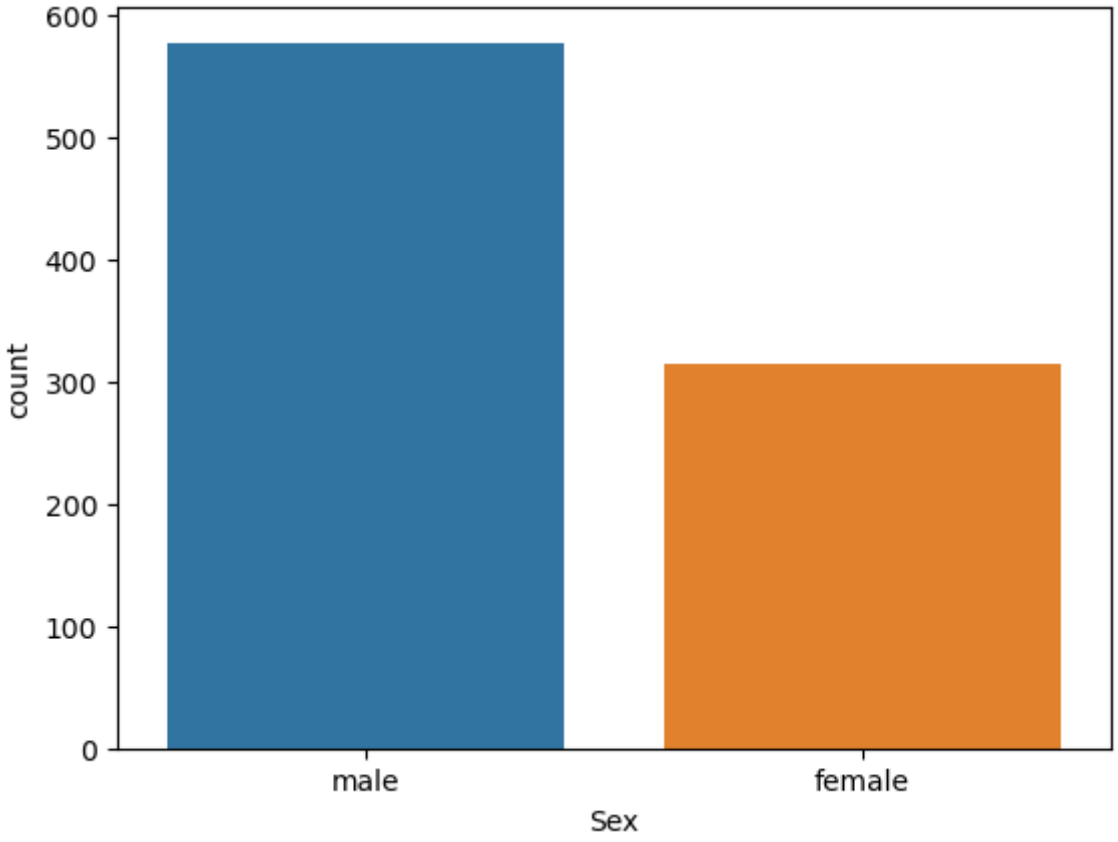
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Ca
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	N
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	N
3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

891 rows × 12 columns

DataVisualzation

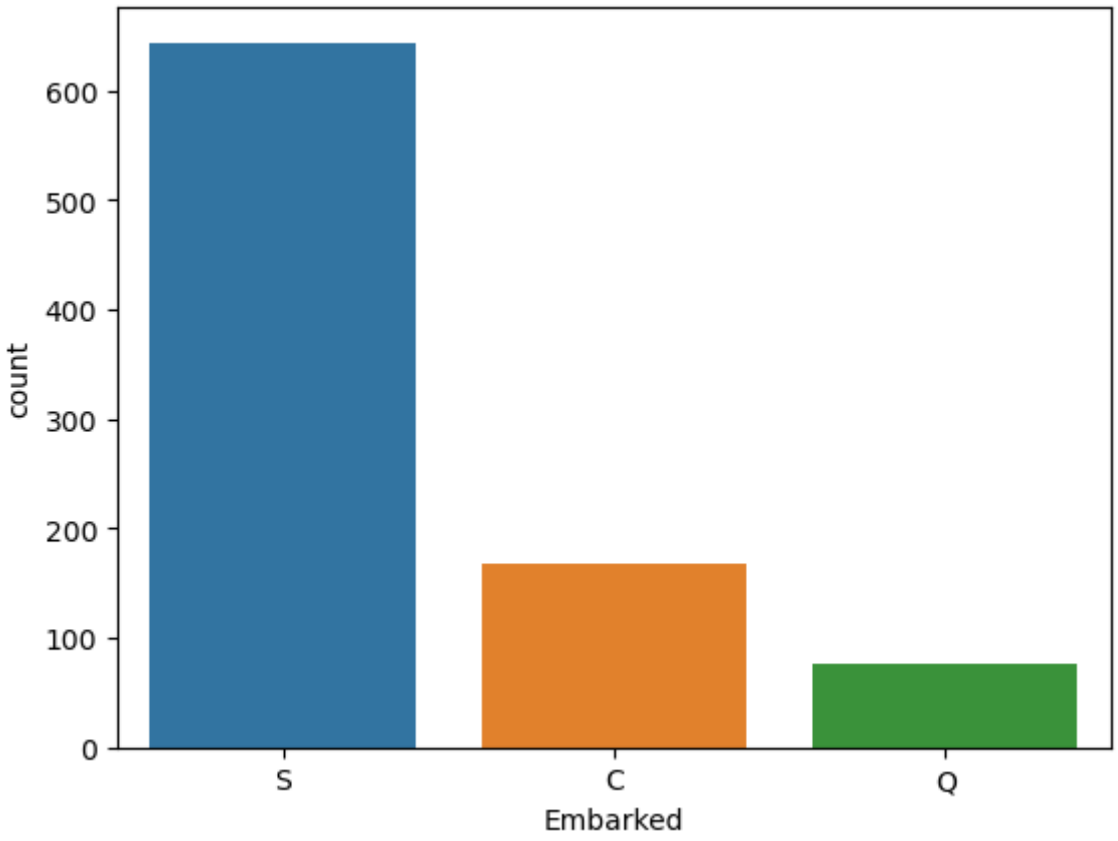
In [45]:
```
sns.countplot(x='Sex',data=df)
```

Out[45]:
```
<Axes: xlabel='Sex', ylabel='count'>
```



In [46]:
```
sns.countplot(x='Embarked',data=df)
```

Out[46]:
```
<Axes: xlabel='Embarked', ylabel='count'>
```



Null Values In DataSet

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
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
```

Dropping unused coloums like PassengerId,Name,Ticket

In [12]:
```
df=df.drop(["PassengerId","Name","Ticket"],axis=1)
```

In [13]:
```
df
```

Out[13]:

	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	0	3	male	22.0	1	0	7.2500	S
1	1	1	female	38.0	1	0	71.2833	C
2	1	3	female	26.0	0	0	7.9250	S
3	1	1	female	35.0	1	0	53.1000	S
4	0	3	male	35.0	0	0	8.0500	S
...	...	...	...	...	...	...	...	...
886	0	2	male	27.0	0	0	13.0000	S
887	1	1	female	19.0	0	0	30.0000	S
888	0	3	female	NaN	1	2	23.4500	S
889	1	1	male	26.0	0	0	30.0000	C
890	0	3	male	32.0	0	0	7.7500	Q

891 rows × 8 columns

Replacing Null Values

In [16]:
```
df["Age"]=df["Age"].fillna(np.mean(df["Age"]))
```

In [17]:
```
df
```

Out[17]:

	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	0	3	male	22.000000	1	0	7.2500	S
1	1	1	female	38.000000	1	0	71.2833	C
2	1	3	female	26.000000	0	0	7.9250	S
3	1	1	female	35.000000	1	0	53.1000	S
4	0	3	male	35.000000	0	0	8.0500	S
...	...	...	...	...	...	...	...	...
886	0	2	male	27.000000	0	0	13.0000	S
887	1	1	female	19.000000	0	0	30.0000	S
888	0	3	female	29.699118	1	2	23.4500	S
889	1	1	male	26.000000	0	0	30.0000	C
890	0	3	male	32.000000	0	0	7.7500	Q

891 rows × 8 columns

Encoding Sex,Embabarked

In [25]:
```
df=df.replace({"sex":{"male":1,"female":0},"Embarked":{"S":0,"C":1,"Q":2}})
```

independent Variable

In [29]:
```
x=df.iloc[:,0:7]
x
```

Out[29]:

	Survived	Pclass	Sex	Age	SibSp	Parch	Fare
0	0	3	male	22.000000	1	0	7.2500
1	1	1	female	38.000000	1	0	71.2833
2	1	3	female	26.000000	0	0	7.9250
3	1	1	female	35.000000	1	0	53.1000
4	0	3	male	35.000000	0	0	8.0500
...	...	...	...	...	...	...	...
886	0	2	male	27.000000	0	0	13.0000
887	1	1	female	19.000000	0	0	30.0000
888	0	3	female	29.699118	1	2	23.4500
889	1	1	male	26.000000	0	0	30.0000
890	0	3	male	32.000000	0	0	7.7500

891 rows × 7 columns

Dependent Variable

In [30]:
```
y=df.iloc[:,7]
y
```

Out[30]:
```
0      0.0
1      1.0
2      0.0
3      0.0
4      0.0
      ...
886    0.0
887    0.0
888    0.0
889    1.0
890    2.0
Name: Embarked, Length: 891, dtype: float64
```

In [ ]:
```

```