In [27]:
```
import numpy as np
import matplotlib.pyplot as plt

import pandas as pd
train=pd.read_csv('/home/student/Downloads/Titanic-Dataset.csv')
train.head()
```

Out[27]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

In [28]:
```
train.shape
```

Out[28]: (891, 12)

In [29]:
```
train.sample(891)
```

Out[29]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
794	795	0	3	Dantcheff, Mr. Ristiu	male	25.0	0	0	349203	7.8
346	347	1	2	Smith, Miss. Marion Elsie	female	40.0	0	0	31418	13.0
372	373	0	3	Beavan, Mr. William Thomas	male	19.0	0	0	323951	8.0
617	618	0	3	Lobb, Mrs. William Arthur (Cordelia K Stanlick)	female	26.0	1	0	A/5. 3336	16.1
424	425	0	3	Rosblom, Mr. Viktor Richard	male	18.0	1	1	370129	20.2
...	...	...	...	...	...	...	...	...	...	
31	32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	female	NaN	1	0	PC 17569	146.5
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631	7.2
356	357	1	1	Bowerman, Miss. Elsie Edith	female	22.0	0	1	113505	55.0
610	611	0	3	Andersson, Mrs. Anders Johan (Alfrida Konstant...	female	39.0	1	5	347082	31.2
234	235	0	2	Leyson, Mr. Robert William Norman	male	24.0	0	0	C.A. 29566	10.5

891 rows × 12 columns

In [30]:
```
train.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 [31]:
```
train.isnull().sum()
```

Out[31]:
```
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 [32]:
```
train.drop(["Cabin"],axis=1,inplace=True)
```

In [40]:
```
train['Age'].fillna(train['Age'].mean(),inplace=True)
```

```
/tmp/ipykernel_6051/900713126.py:1: FutureWarning: A value is tryi
ng to be set on a copy of a DataFrame or Series through chained as
signment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will n
ever 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[co
l].method(value) instead, to perform the operation inplace on the
original object.


  train['Age'].fillna(train['Age'].mean(),inplace=True)
```

In [41]:
```
train['Embarked'].value_counts()
```

Out[41]:
```
Embarked
S    646
C    168
Q     77
Name: count, dtype: int64
```

In [38]:
```
train["Embarked"].fillna("S",inplace=True)
```

In [37]:
```
train.isnull().sum()
```

Out[37]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

In [43]:
```
#Data Analysis
train.describe()
```

Out[43]:

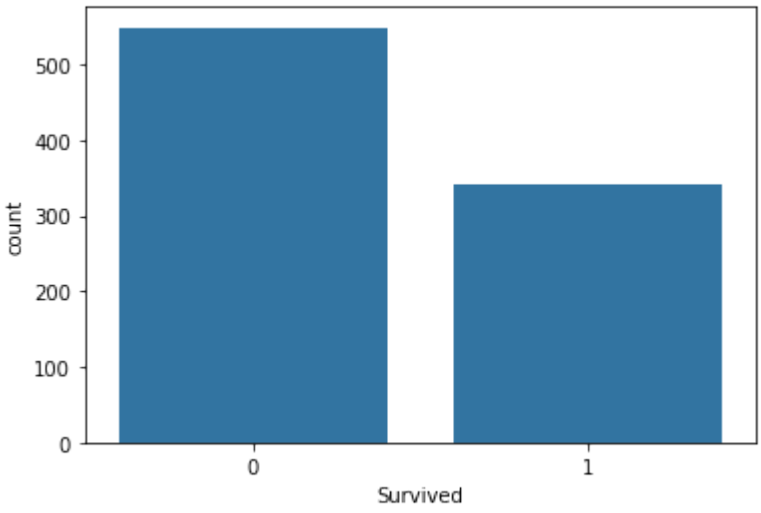
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fa
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.00000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.20420
std	257.353842	0.486592	0.836071	13.002015	1.102743	0.806057	49.69342
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.00000
25%	223.500000	0.000000	2.000000	22.000000	0.000000	0.000000	7.91040
50%	446.000000	0.000000	3.000000	29.699118	0.000000	0.000000	14.45420
75%	668.500000	1.000000	3.000000	35.000000	1.000000	0.000000	31.00000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.32920

In [46]:
```
import seaborn as sns
train['Survived'].value_counts()
```

Out[46]:
```
Survived
0    549
1    342
Name: count, dtype: int64
```

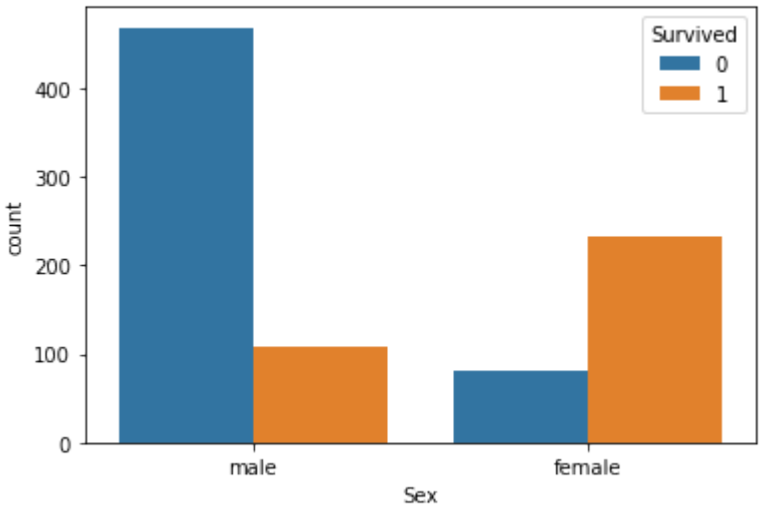
In [47]:
```
sns.countplot(x="Survived",data=train)
```

Out[47]: <Axes: xlabel='Survived', ylabel='count'>



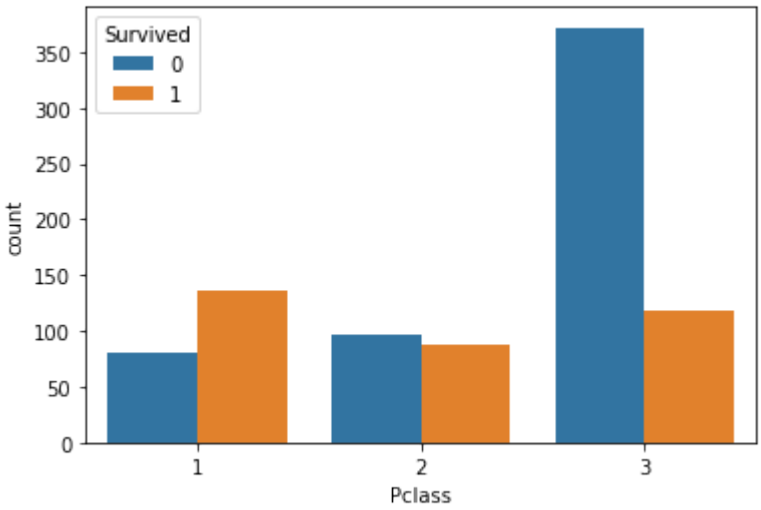
In [48]:
```
sns.countplot(x="Sex",hue="Survived",data=train)
```

Out[48]: <Axes: xlabel='Sex', ylabel='count'>



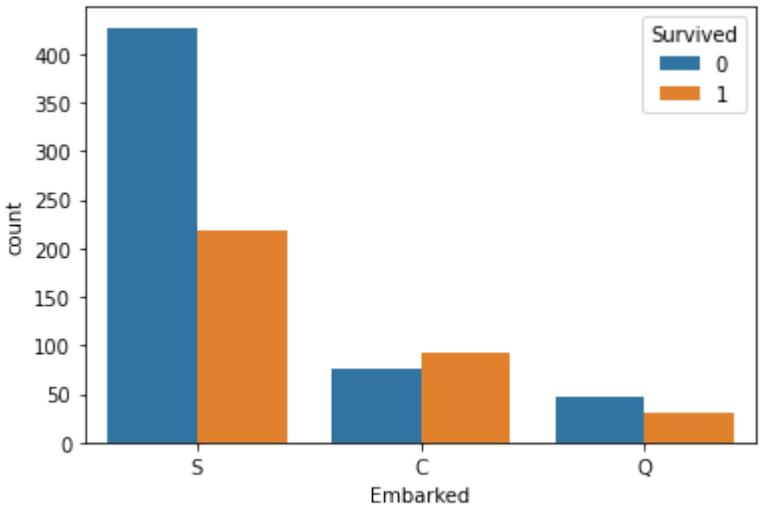
In [49]:
```
sns.countplot(x="Pclass",hue="Survived",data=train)
```

Out[49]: <Axes: xlabel='Pclass', ylabel='count'>



In [50]:
```
sns.countplot(x="Embarked",hue="Survived",data=train)
```

Out[50]: <Axes: xlabel='Embarked', ylabel='count'>

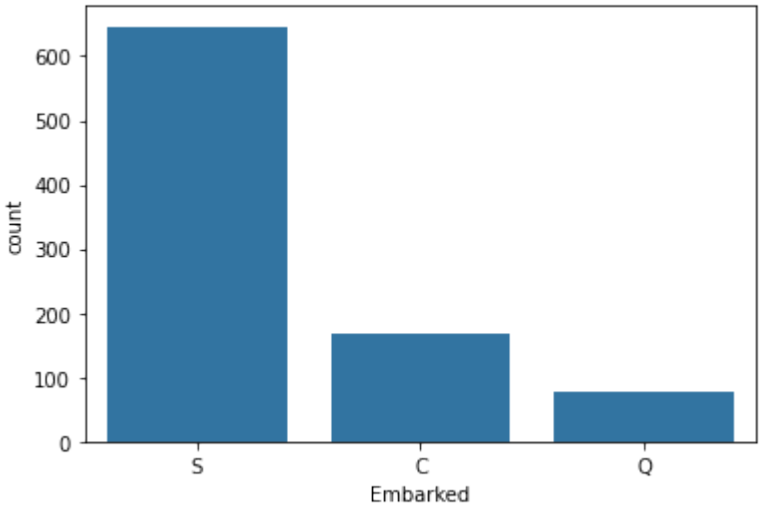


In [54]:
```
train['Embarked'].value_counts()
```

Out[54]:
```
Embarked
S    646
C    168
Q     77
Name: count, dtype: int64
```

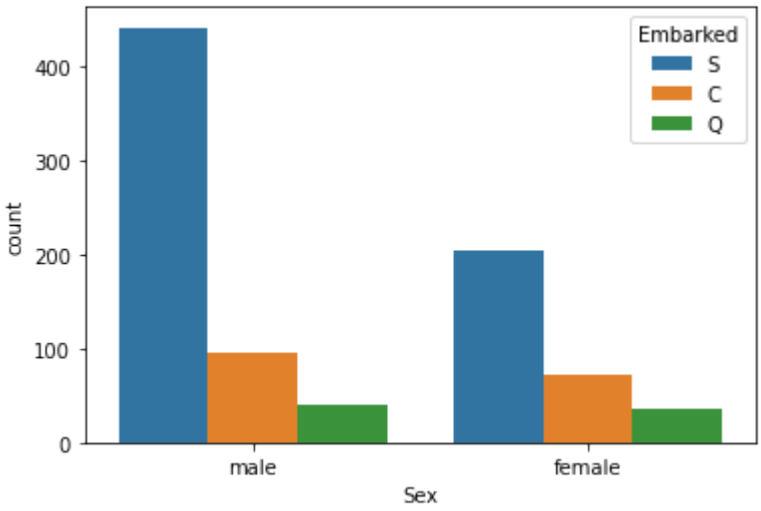
In [55]:
```
sns.countplot(x="Embarked",data=train)
```

Out[55]: <Axes: xlabel='Embarked', ylabel='count'>



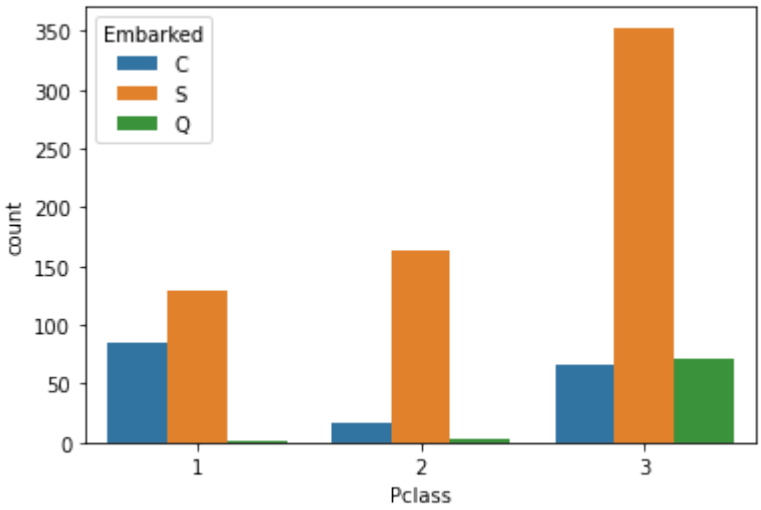
In [56]:
```
sns.countplot(x="Sex",hue="Embarked",data=train)
```

Out[56]: <Axes: xlabel='Sex', ylabel='count'>

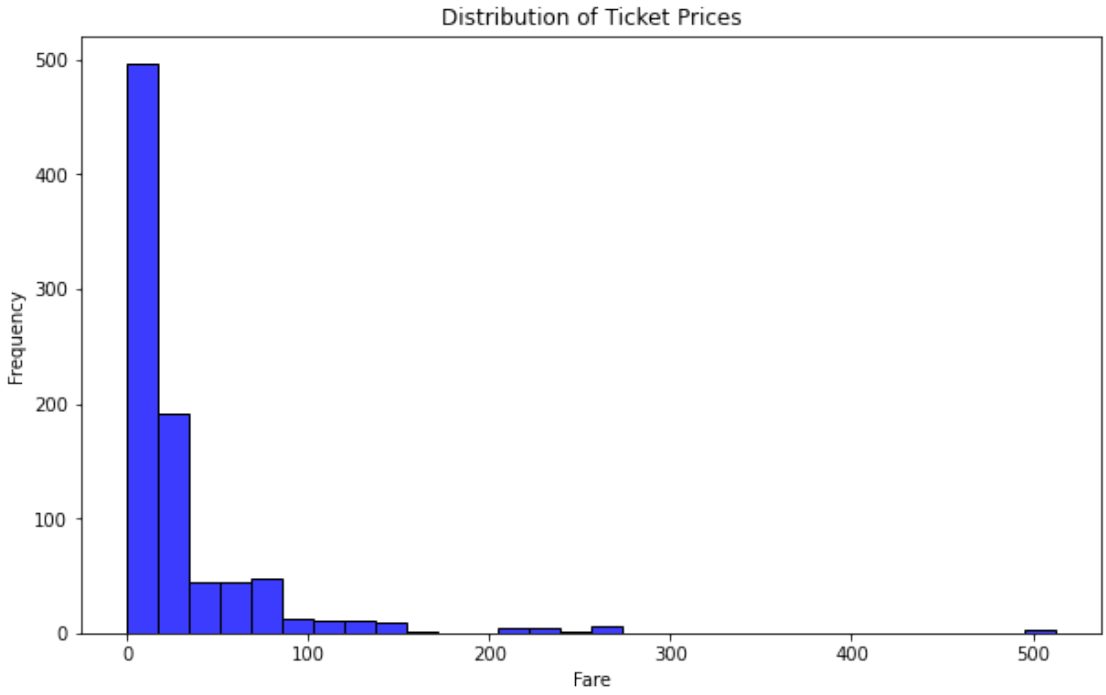


In [57]:
```
sns.countplot(x="Pclass",hue="Embarked",data=train)
```

Out[57]: <Axes: xlabel='Pclass', ylabel='count'>



In [58]:
```
import matplotlib.pyplot as plt
plt.figure(figsize=(10,6))
sns.histplot(train['Fare'],bins=30,color='blue')
plt.title('Distribution of Ticket Prices')
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.show()
```



In [ ]: