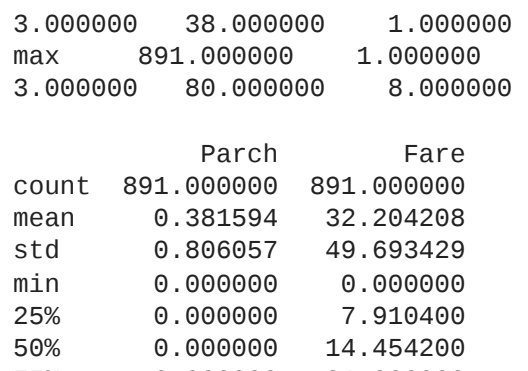
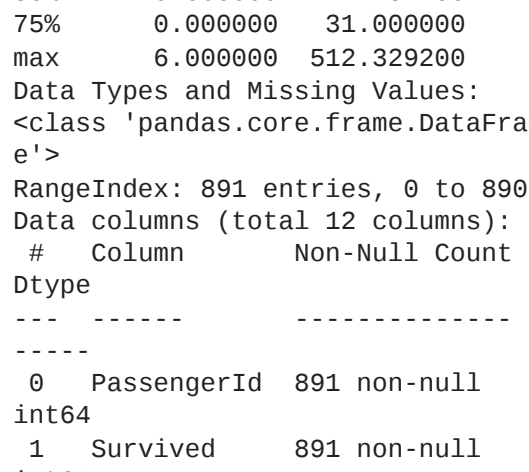
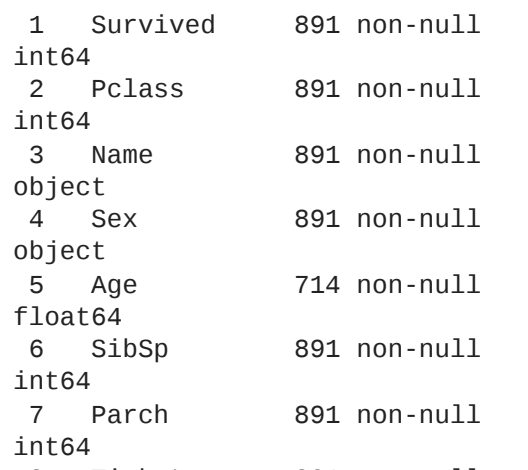
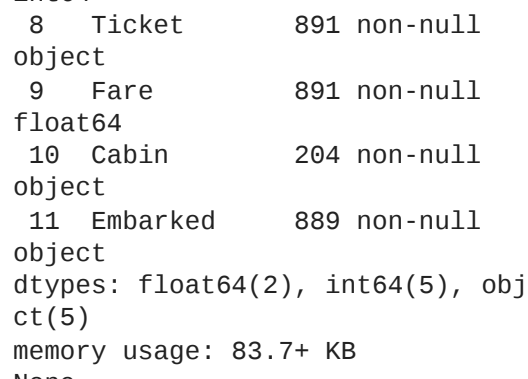
In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#loding the train.csv file
titanic_data = pd.read_csv('C:/Use

# Display the first few rows of th
print(titanic_data.head())

# Basic statistics
print("Basic Statistics:")
print(titanic_data.describe())

# Data types and missing values
print("Data Types and Missing Valu
print(titanic_data.info())

# Survival count
print("Survival Count:")
print(titanic_data['Survived'].val

# Age distribution
plt.figure(figsize=(10, 6))
sns.histplot(titanic_data['Age'].d
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()

# Survival by gender
plt.figure(figsize=(8, 6))
sns.countplot(x='Survived', hue='S
plt.title('Survival by Gender')
plt.xlabel('Survived')
plt.ylabel('Count')
plt.legend(title='Sex', loc='upper
plt.show()

# Survival by passenger class
plt.figure(figsize=(8, 6))
sns.countplot(x='Survived', hue='P
plt.title('Survival by Passenger C
plt.xlabel('Survived')
plt.ylabel('Count')
plt.legend(title='Pclass', loc='up
plt.show()

# Fare distribution by passenger c
plt.figure(figsize=(10, 6))
sns.boxplot(x='Pclass', y='Fare',
plt.title('Fare Distribution by Pa
plt.xlabel('Pclass')
plt.ylabel('Fare')
plt.show()
```

```
   PassengerId  Survived  Pclass
\
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3


Name     Sex   Age  SibSp  \
0                              Braun
d, Mr. Owen Harris    male  22.0
1
1  Cumings, Mrs. John Bradley (Flo
rence Briggs Th...  female  38.0
1
2                               Heik
kinen, Miss. Laina  female  26.0
0
3       Futrelle, Mrs. Jacques Hea
th (Lily May Peel)  female  35.0
1
4                             Allen,
Mr. William Henry    male  35.0
0

   Parch            Ticket     Far
e Cabin Embarked
0      0         A/5 21171   7.250
0   NaN        S
1      0          PC 17599  71.283
3   C85        C
2      0  STON/O2. 3101282   7.925
0   NaN        S
3      0            113803  53.100
0  C123        S
4      0            373450   8.050
0   NaN        S
Basic Statistics:
       PassengerId    Survived
Pclass         Age       SibSp  \
count   891.000000  891.000000  89
1.000000  714.000000  891.000000
mean    446.000000    0.383838
2.308642   29.699118    0.523008
std     257.353842    0.486592
0.836071   14.526497    1.102743
min       1.000000    0.000000
1.000000    0.420000    0.000000
25%     223.500000    0.000000
2.000000   20.125000    0.000000
50%     446.000000    0.000000
3.000000   28.000000    0.000000
75%     668.500000    1.000000
3.000000   38.000000    1.000000
max     891.000000    1.000000
3.000000   80.000000    8.000000

            Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200
Data Types and Missing Values:
<class 'pandas.core.frame.DataFram
e'>
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), obje
ct(5)
memory usage: 83.7+ KB
None
Survival Count:
Survived
0    549
1    342
Name: count, dtype: int64
```









In [ ]: