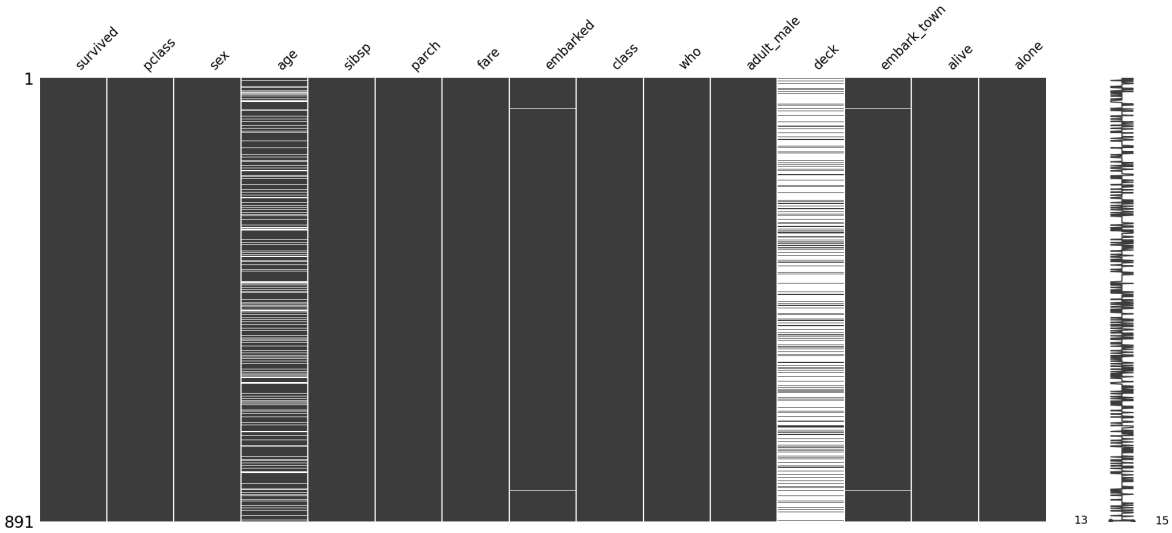
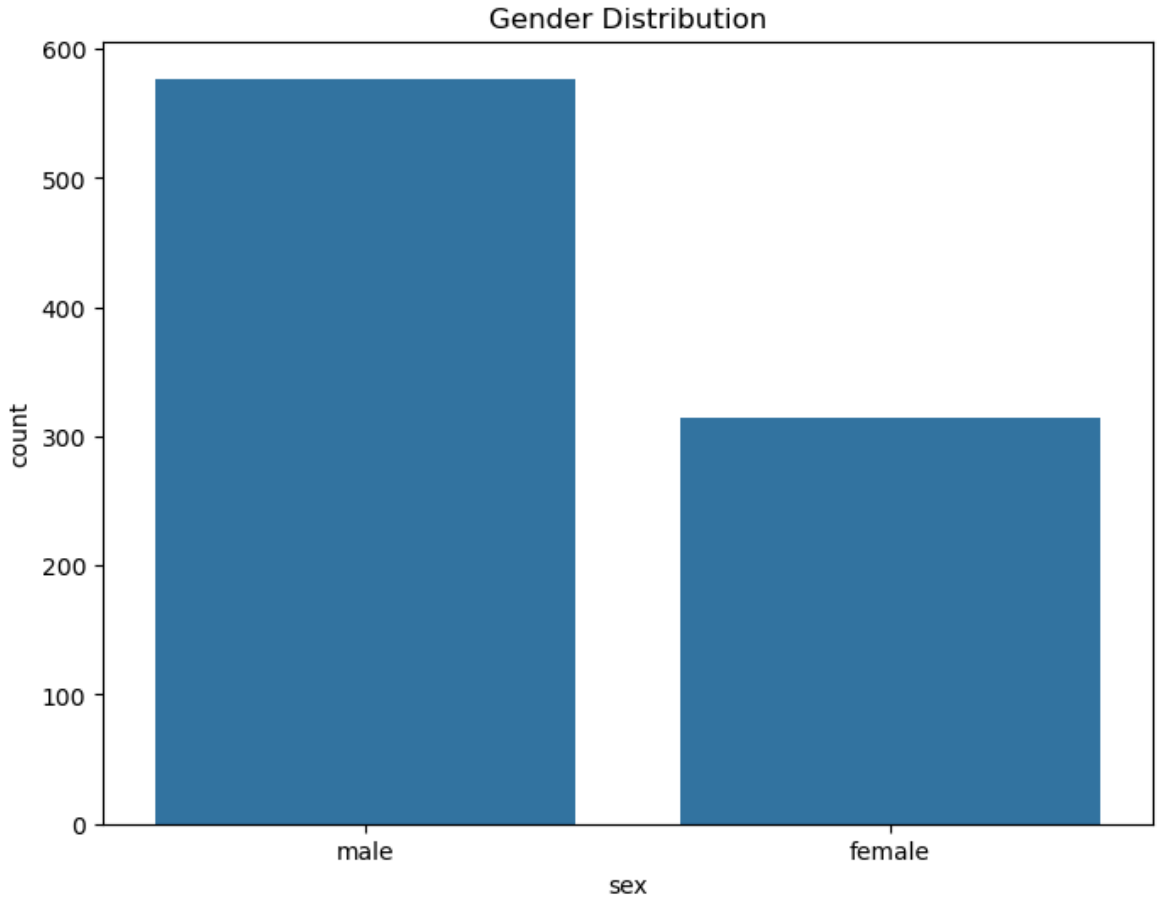
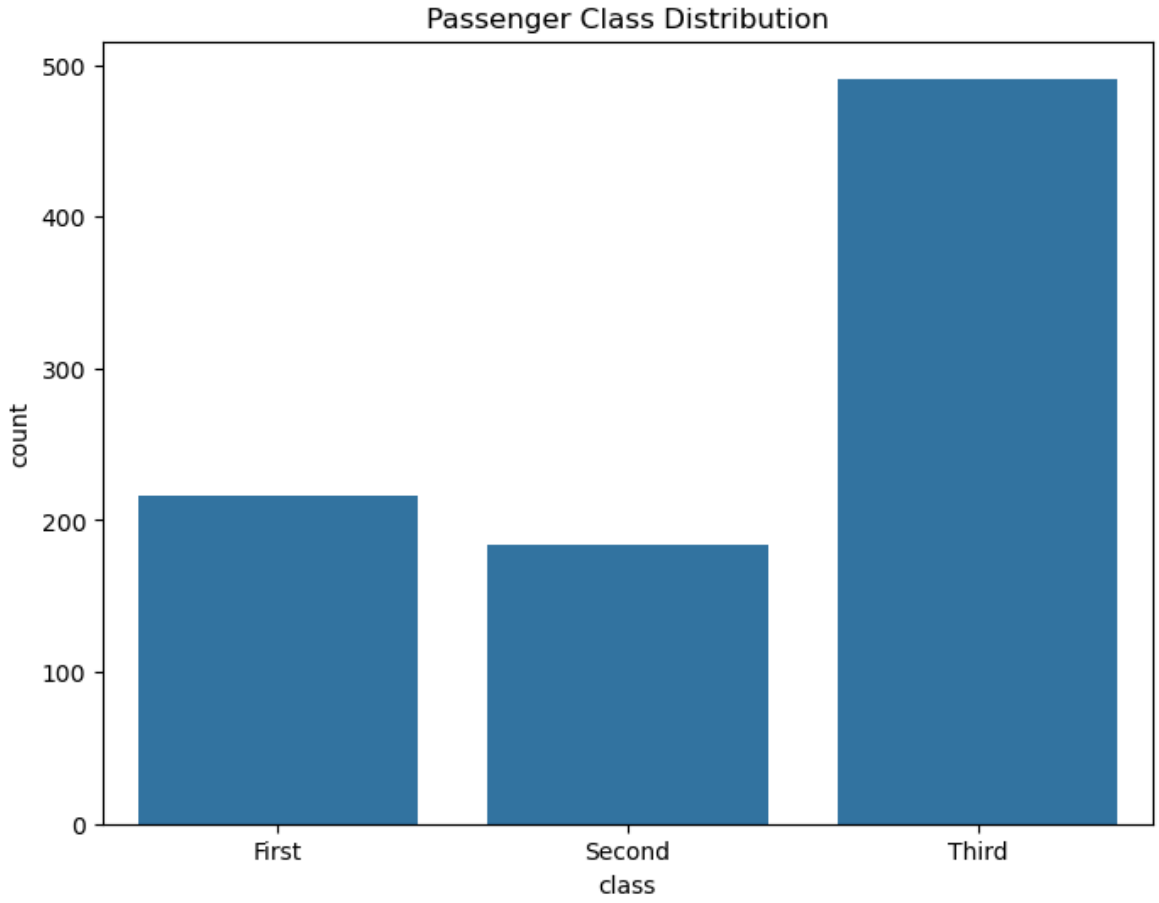
In [1]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load Titanic dataset from seaborn
df = sns.load_dataset('titanic')

# Display first few rows
df.head()

# Check Data Info
df.info()

# Check for Null Values
df.isnull().sum()

# Handle Missing Data (optional - here just view it)
# For EDA, it's enough to know where missing values exist.

# Visualize Missing Data (Optional)
import missingno as msno
msno.matrix(df)
plt.show()

# Basic Statistics
df.describe()

# Univariate Analysis
plt.figure(figsize=(8,6))
sns.countplot(x='sex', data=df)
plt.title('Gender Distribution')
plt.show()

plt.figure(figsize=(8,6))
sns.countplot(x='class', data=df)
plt.title('Passenger Class Distribution')
plt.show()

# Bivariate Analysis
plt.figure(figsize=(8,6))
sns.barplot(x='sex', y='survived', data=df)
plt.title('Survival Rate by Gender')
plt.show()

plt.figure(figsize=(8,6))
sns.barplot(x='class', y='survived', data=df)
plt.title('Survival Rate by Class')
plt.show()

# Correlation Heatmap
plt.figure(figsize=(10,8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
plt.title('Feature Correlation')
plt.show()

# Outliers detection (boxplot example)
plt.figure(figsize=(8,6))
sns.boxplot(x='age', data=df)
```

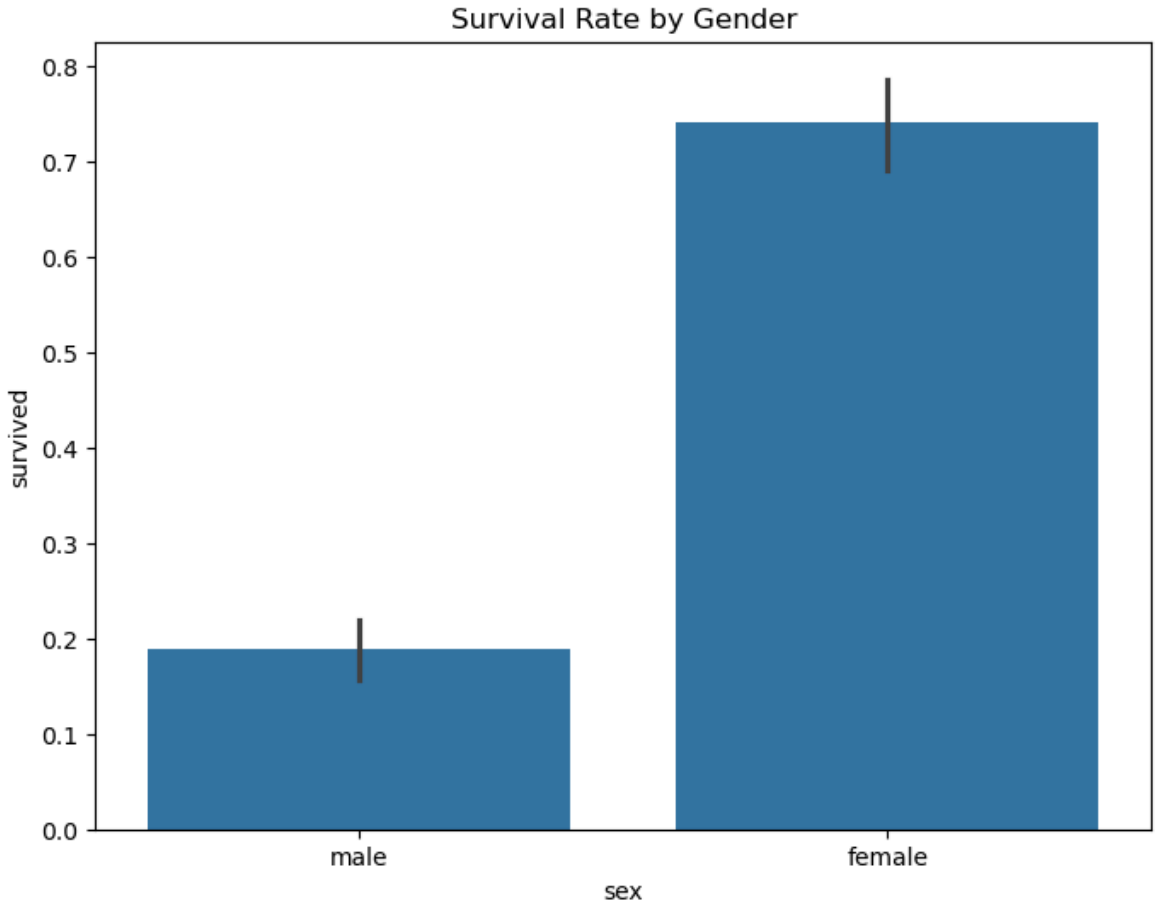
```
plt.title('Age Outliers')
plt.show()
```

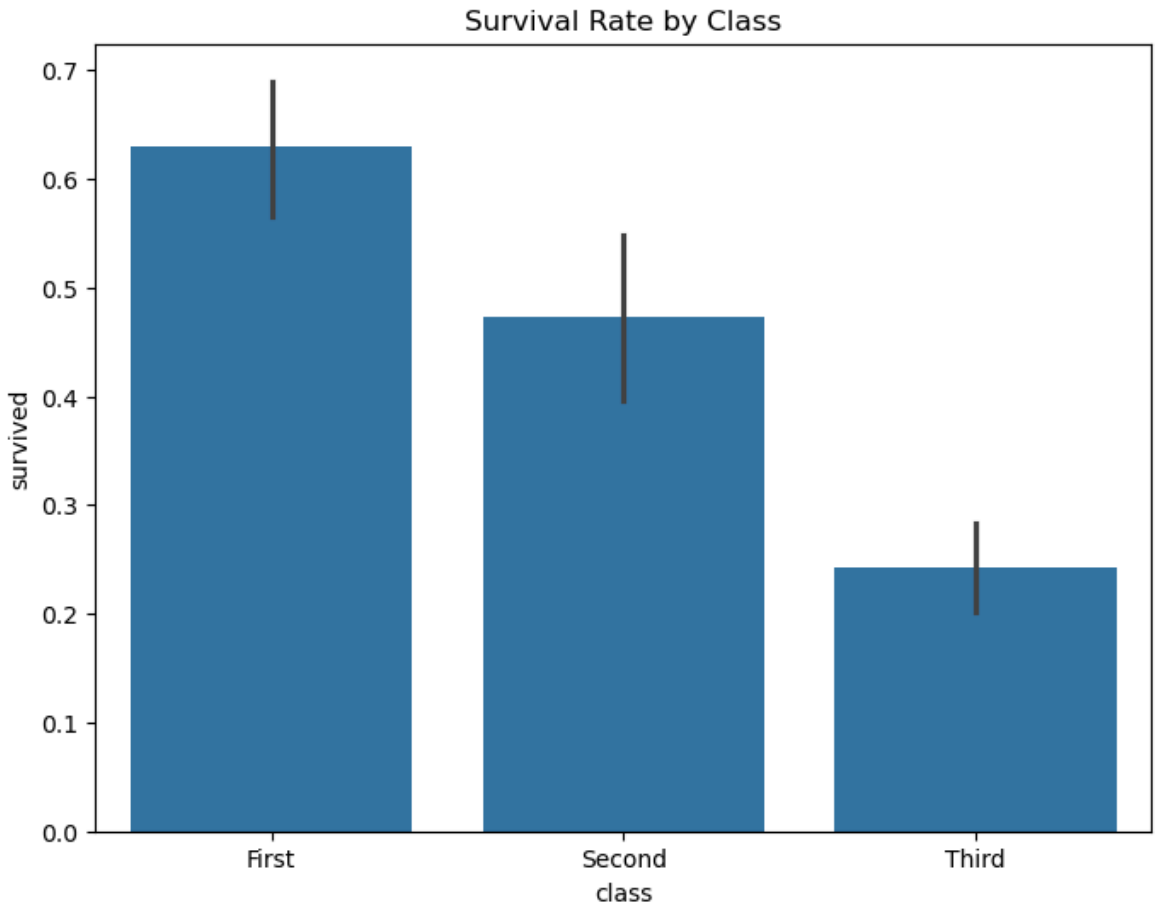
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   survived     891 non-null    int64
 1   pclass       891 non-null    int64
 2   sex          891 non-null    object
 3   age          714 non-null    float64
 4   sibsp        891 non-null    int64
 5   parch        891 non-null    int64
 6   fare         891 non-null    float64
 7   embarked     889 non-null    object
 8   class        891 non-null    category
 9   who          891 non-null    object
 10  adult_male   891 non-null    bool
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object
 13  alive        891 non-null    object
 14  alone        891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
```




Gender Distribution
600
500
400
300
200
100
0
count
male
female
sex


Passenger Class Distribution
500
400
300
200
100
0
count
First
Second
Third
class


Survival Rate by Gender
survived
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
male
female
sex


Survival Rate by Class
survived
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
First
Second
Third
class

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[1], line 52
     50 # Correlation Heatmap
     51 plt.figure(figsize=(10,8))
---> 52 sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
     53 plt.title('Feature Correlation')
     54 plt.show()

File c:\Users\shreyas wakhare\anaconda3\envs\pandas_prof\Lib\site-packages\pandas
\core\frame.py:11036, in DataFrame.corr(self, method, min_periods, numeric_only)
  11034 cols = data.columns
  11035 idx = cols.copy()
> 11036 mat = data.to_numpy(dtype=float, na_value=np.nan, copy=False)
  11038 if method == "pearson":
  11039     correl = libalgos.nancorr(mat, minp=min_periods)

File c:\Users\shreyas wakhare\anaconda3\envs\pandas_prof\Lib\site-packages\pandas
\core\frame.py:1981, in DataFrame.to_numpy(self, dtype, copy, na_value)
   1979 if dtype is not None:
   1980     dtype = np.dtype(dtype)
-> 1981 result = self._mgr.as_array(dtype=dtype, copy=copy, na_value=na_value)
   1982 if result.dtype is not dtype:
   1983     result = np.array(result, dtype=dtype, copy=False)

File c:\Users\shreyas wakhare\anaconda3\envs\pandas_prof\Lib\site-packages\pandas
\core\internals\managers.py:1692, in BlockManager.as_array(self, dtype, copy, na_
value)
   1690         arr.flags.writeable = False
   1691 else:
-> 1692     arr = self._interleave(dtype=dtype, na_value=na_value)
   1693     # The underlying data was copied within _interleave, so no need
   1694     # to further copy if copy=True or setting na_value
   1696 if na_value is lib.no_default:

File c:\Users\shreyas wakhare\anaconda3\envs\pandas_prof\Lib\site-packages\pandas
\core\internals\managers.py:1751, in BlockManager._interleave(self, dtype, na_val
ue)
   1749     else:
   1750         arr = blk.get_values(dtype)
-> 1751     result[rl.indexer] = arr
   1752     itemmask[rl.indexer] = 1
   1754 if not itemmask.all():

ValueError: could not convert string to float: 'male'
```