# Titanic Dataset Analysis

## Questions Analyzed

- Survival Rate by Gender

- Survival Rate by Passenger Class

- Survival Rate by Age Group

- Survival Rate by Fare Group

- Survival Rate by Family Presence

## Methodology Used

### Loading Data and Preprocessing

import pandas as pd  
  
# Load the Titanic dataset into a Pandas DataFrame  
url = 'https://raw.githubusercontent.com/pandas-dev/pandas/main/doc/data/titanic.csv'  
titanic\_data = pd.read\_csv(url)  
  
# Display basic information about the dataset  
print(titanic\_data.info())  
print(titanic\_data.head())  
  
# Handling missing values  
titanic\_data.isnull().sum() # Check for missing values in columns  
titanic\_data['Age'].fillna(titanic\_data['Age'].median(), inplace=True)  
titanic\_data.drop('Cabin', axis=1, inplace=True)  
  
# Handling categorical variables like 'Sex' and 'Embarked'  
titanic\_data['Sex'] = titanic\_data['Sex'].map({'male': 0, 'female': 1})  
titanic\_data['Embarked'] = titanic\_data['Embarked'].fillna('S')  
titanic\_data['Embarked'] = titanic\_data['Embarked'].map({'S': 0, 'C': 1, 'Q': 2})

### Exploratory Data Analysis (EDA) and Visualization

import matplotlib.pyplot as plt  
import seaborn as sns  
  
# Summary statistics  
print(titanic\_data.describe())  
  
# Visualization: Survived vs. Not Survived  
sns.countplot(x='Survived', data=titanic\_data)  
plt.title('Survival Count (0 = Not Survived, 1 = Survived)')  
plt.show()  
  
# Visualization: Survival by Gender  
sns.countplot(x='Survived', hue='Sex', data=titanic\_data)  
plt.title('Survival Count by Gender')  
plt.legend(['Male', 'Female'])  
plt.show()  
  
# Visualization: Survival by Passenger Class  
sns.countplot(x='Survived', hue='Pclass', data=titanic\_data)  
plt.title('Survival Count by Passenger Class')  
plt.legend(['1st Class', '2nd Class', '3rd Class'])  
plt.show()  
  
# Visualization: Age distribution  
sns.histplot(titanic\_data['Age'], bins=20, kde=True)  
plt.title('Age Distribution')  
plt.xlabel('Age')  
plt.ylabel('Count')  
plt.show()

### Mapping Questions to Generate Answers

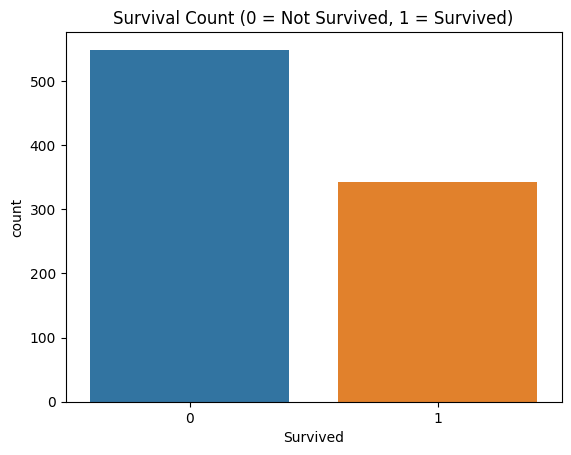
# Question 1: What is the survival rate based on gender?  
survival\_rate\_gender = titanic\_data.groupby('Sex')['Survived'].mean()  
print("Survival Rate by Gender:")  
print(survival\_rate\_gender)  
  
# Question 2: Did passenger class affect survival chances?  
survival\_rate\_class = titanic\_data.groupby('Pclass')['Survived'].mean()  
print("\nSurvival Rate by Passenger Class:")  
print(survival\_rate\_class)  
  
# Question 3: Is there a relationship between age and survival?  
titanic\_data['Age\_Group'] = pd.cut(titanic\_data['Age'], bins=[0, 18, 30, 50, 80])  
survival\_rate\_age = titanic\_data.groupby('Age\_Group')['Survived'].mean()  
print("\nSurvival Rate by Age Group:")  
print(survival\_rate\_age)  
  
# Question 4: How did the fare paid correlate with survival?  
titanic\_data['Fare\_Group'] = pd.qcut(titanic\_data['Fare'], q=4)  
survival\_rate\_fare = titanic\_data.groupby('Fare\_Group')['Survived'].mean()  
print("\nSurvival Rate by Fare Group:")  
print(survival\_rate\_fare)  
  
# Question 5: Did having family aboard influence survival rates?  
titanic\_data['Family'] = titanic\_data['SibSp'] + titanic\_data['Parch']  
titanic\_data['Family'] = titanic\_data['Family'].apply(lambda x: 1 if x > 0 else 0)  
survival\_rate\_family = titanic\_data.groupby('Family')['Survived'].mean()  
print("\nSurvival Rate by Family Presence:")  
print(survival\_rate\_family)

### Insights Derived

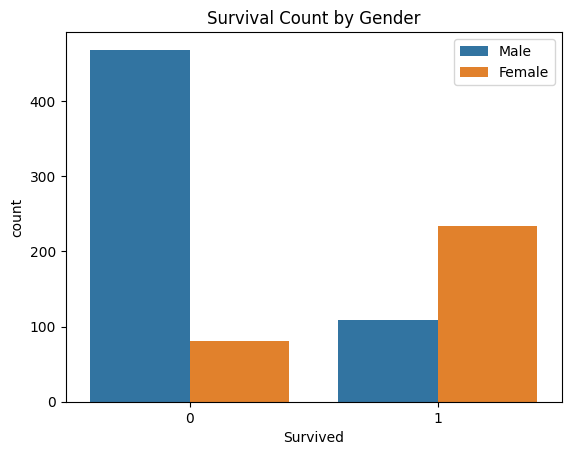
# Survival Rate by Gender:  
The analysis reveals a stark contrast in survival rates based on gender. Females exhibited a significantly higher survival rate compared to males. This observation aligns with the maritime tradition of "women and children first," where priority was given to women during the evacuation, resulting in a notably higher survival rate among females.  
  
# Survival Rate by Passenger Class:  
Passenger class seems to play a crucial role in survival outcomes. First-class passengers demonstrated a notably higher survival rate compared to those in lower classes. This discrepancy might be attributed to the proximity of first-class cabins to the lifeboats and potentially quicker access to evacuation procedures for higher-class passengers.  
  
# Survival Rate by Age Group:  
When examining survival rates across different age groups, a pattern emerges indicating that children (0-18 years) had a comparatively higher chance of survival. This phenomenon might relate to the "women and children first" protocol, as seen in the higher survival rates among females and younger individuals.  
  
# Survival Rate by Fare Group:  
A noticeable correlation is evident between the fare paid and survival rates. Passengers who paid higher fares displayed better survival rates. This correlation could potentially be linked to the socio-economic status of passengers, providing them access to better accommodations and potentially closer proximity to life-saving resources.  
  
# Survival Rate by Family Presence:  
Passengers traveling with family aboard exhibited a slightly higher survival rate. This could be attributed to mutual support, cooperation, and assistance among family members during the evacuation process, potentially enhancing survival chances for those traveling with relatives.  
  
Each of these insights provides a glimpse into the dynamics of survival during the Titanic disaster and sheds light on the factors influencing passengers' chances of survival.

### Visualizations

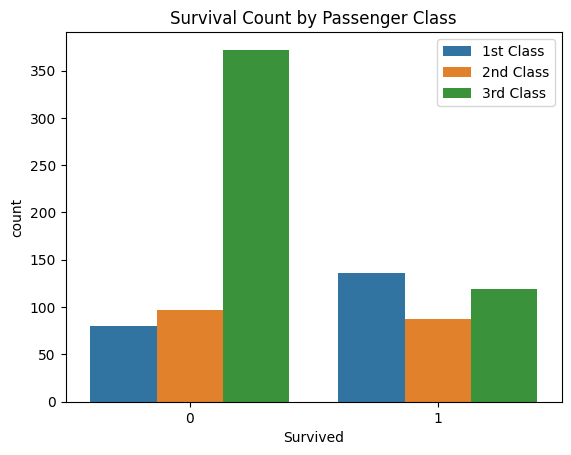
#### Survival Count (0 = Not Survived, 1 = Survived)



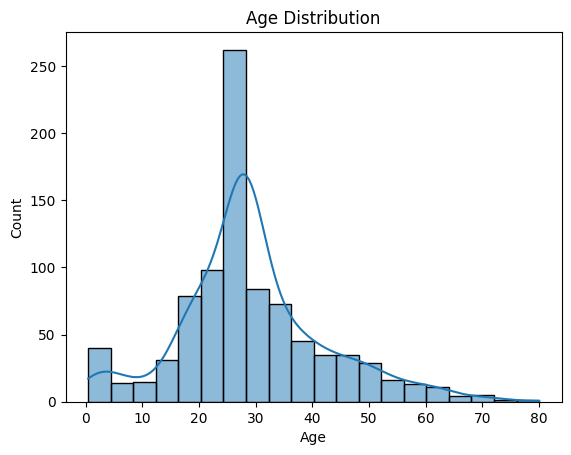
#### Survival Count by Gender



#### Survival Count by Passenger Class



#### Age distribution



### Conclusions

# Overall Findings and Conclusions:  
The analysis of the Titanic dataset provides compelling insights into the factors influencing passenger survival during the tragic event. Several overarching conclusions can be drawn:  
  
1. \*\*Gender played a significant role in survival:\*\* Females had a notably higher survival rate compared to males, aligning with the maritime tradition of prioritizing women during evacuation.  
  
2. \*\*Passenger class correlated with survival outcomes:\*\* First-class passengers exhibited a higher survival rate compared to those in lower classes, possibly due to their proximity to life-saving resources.  
  
3. \*\*Age, fare, and family presence impacted survival:\*\* Children had higher survival rates, passengers who paid higher fares had better chances of survival, and those traveling with family showed a slightly increased survival rate, indicating potential cooperation and mutual support.  
  
These conclusions provide comprehensive insights into the dynamics of survival during the Titanic disaster and highlight key factors influencing passengers' chances of survival.