1. Problem Space (Clarification of the Problem Space)  
**What**: The goal of this project is to analyze the Titanic dataset to identify factors that influenced passenger survival during the tragic sinking of the ship in 1912.

**Why**: Understanding these factors provides insights into historical decisions and survival patterns, which can have educational value and relevance for disaster management studies.

**Who**: This analysis is useful for historians, sociologists, and educators interested in understanding human behavior in disaster scenarios.

**Where**: The data focuses on passengers aboard the RMS Titanic, a transatlantic ocean liner.

**When**: The dataset covers the period of the ship's voyage in 1912.

2. Questions to Address through Data Analysis  
What was the survival rate of passengers on the Titanic?

How did gender affect survival rates?

Was there a significant difference in survival rates among different passenger classes (Pclass)?

Did the age of passengers influence their chances of survival?

Were passengers who embarked from different ports (Embarked) more or less likely to survive?

What was the correlation between having family members on board (SibSp and Parch) and survival rates?

3. Understanding of Dataset  
Data Source: The Titanic dataset is publicly available on Kaggle and contains detailed information about the passengers aboard the RMS Titanic.

Dataset Description: The dataset includes fields such as:

Survived: Survival status (1 = Survived, 0 = Did not survive).

Pclass: Passenger class (1st, 2nd, or 3rd).

Sex: Gender of the passenger.

Age: Age of the passenger.

SibSp and Parch: Number of family members onboard.

Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).

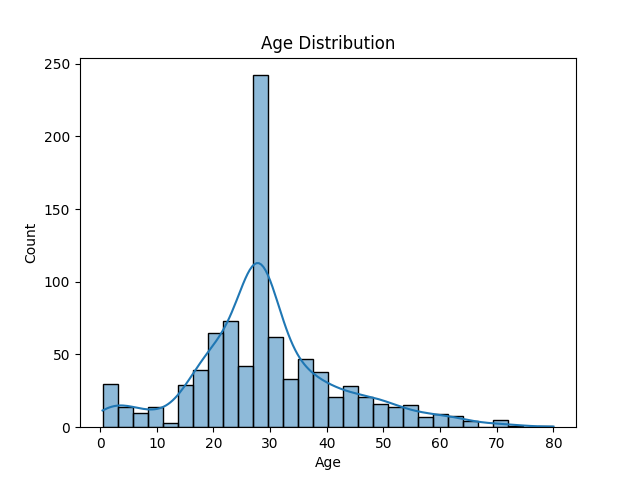
Dataset Relevance:  
The dataset provides all the necessary variables to analyze factors influencing survival rates, such as gender, age, passenger class, and family relationships.

A graph of survival count

Description automatically generated  
A graph of a passenger class distribution

Description automatically generated  
A graph of a number of people

Description automatically generated with medium confidenceA bar graph with blue squares

Description automatically generatedA graph of a number of blue lines

Description automatically generated with medium confidenceA graph with blue rectangular bars

Description automatically generatedA graph of a graph with blue bars

Description automatically generated with medium confidenceA graph of a survival rate

Description automatically generated A graph of a graph of a family size

Description automatically generated with medium confidence**Questions and Answers**

1. **What is the overall survival rate of Titanic passengers?**

**Answer:** Based on the Survived column, the survival rate is approximately 38%,

1. **How does gender affect the survival rate?**

**Answer:** Grouped analysis shows that female passengers had a significantly higher survival rate than male passengers, indicating a priority for rescuing women during the evacuation. This indicates that the priority of rescue for women is higher than that for men

1. **Is there a significant difference in survival rates across different passenger classes (Pclass)?**

**Answer:** Survival rates are notably higher for passengers in higher classes (especially 1st class). This suggests that passenger class influenced rescue priority. It could be proximity to the lifeboats or social norms.

1. **Does age influence survival rate?**

**Answer:** Boxplots or age distribution analysis reveal that younger passengers (especially children) had a higher survival rate.

1. **What is the impact of embarkation ports (Embarked) on survival rates?**

**Answer:** Grouped statistics show that passengers who embarked from Cherbourg (C) had the highest survival rate, possibly due to a higher concentration of first-class passengers from this port.

1. **Did traveling with family affect survival rates?**

**Answer:** Grouped analysis based on the combined values of SibSp and Parch indicates that passengers with a moderate number of family members had a higher survival rate, while those traveling alone or with many family members had lower survival rates. Medium-sized families may be better helped during the evacuation process or because family members help each other.

[[1]](#endnote-1)

1. **Titanic data analysis summary report**

   Data background

   Titanic data set contains the basic information of passengers and survival labels for analyzing the factors affecting the survival rate.

   **Key finding**

   Female passengers have a significantly higher survival rate than male passengers.

   The survival rate of first class passengers is significantly higher than that of other classes.

   The survival rate for child passengers is better than average.

   The port of embarkation (Cherbourg) has the highest survival rate for passengers.

   **Data support**

   Charts and data presentations (e.g., survival charts, sex and survival charts, etc.).

   **Conclusion**

   The main influencing factors of survival rate include sex, cabin and age. These factors may play an important role in prioritizing rescue decisions in a disaster.

   Gender and passenger class were the strongest predictors of survival.  
     
     
     
   The accuracy of the model is 81%, which can explain the key problems reasonably [↑](#endnote-ref-1)