Task 2: Exploratory Data Analysis on Titanic Dataset

# Introduction

This report presents the analysis of the Titanic dataset. The dataset contains information about passengers, including demographic details such as age, gender, class, and survival status. The purpose of this analysis is to explore patterns, clean the data, and build predictive models to understand the factors that influence survival rates.

# Dataset Overview

The dataset consists of 891 passenger records with the following key variables:  
- PassengerId: Unique identifier  
- Survived: Survival status (1 = Survived, 0 = Did not survive)  
- Pclass: Passenger class (1, 2, or 3)  
- Name, Sex, Age  
- SibSp: Number of siblings/spouses aboard  
- Parch: Number of parents/children aboard  
- Ticket, Fare, Cabin, Embarked

# Results and Outcomes

The following observations and results were obtained from the notebook analysis:

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 Braund, Mr. Owen Harris male 22.0 1   
1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 1   
2 Heikkinen, Miss. Laina female 26.0 0   
3 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 1   
4 Allen, Mr. William Henry male 35.0 0   
  
 Parch Ticket Fare Cabin Embarked   
0 0 A/5 21171 7.2500 NaN S   
1 0 PC 17599 71.2833 C85 C   
2 0 STON/O2. 3101282 7.9250 NaN S   
3 0 113803 53.1000 C123 S   
4 0 373450 8.0500 NaN S   
  
<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  
  
  
PassengerId int64  
Survived int64  
Pclass int64  
Name object  
Sex object  
Age float64  
SibSp int64  
Parch int64  
Ticket object  
Fare float64  
Cabin

# Conclusion

The analysis highlights the importance of passenger class, gender, and age in determining survival on the Titanic. Females and children had a higher chance of survival compared to adult males, and passengers in higher classes were more likely to survive. These findings were consistent with expectations based on historical records of the Titanic disaster.