Titanic Dataset Analysis Report

Introduction:

This report contains the analysis performed on the Titanic dataset using Jupyter Notebook. The purpose of the analysis is to understand the distribution and characteristics of passengers aboard the Titanic.

Dataset Overview:

- The dataset includes features such as PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, and Embarked.
- The target variable is "Survived" (0 = No, 1 = Yes).

Data Description:

Using the describe() function, we observed the following statistical summary:

- Age:
- Mean age: ~29.7 years
- Minimum age: 0.42 years (infant)
- Maximum age: 80 years

- Fare:
- Minimum fare: 0 (free or missing data)
- Maximum fare: 512.33 (luxury cabins)

- SibSp (Siblings/Spouses aboard):

- Most passengers have 0 siblings/spouses with them.

- Parch (Parents/Children aboard):

- Most passengers have 0 parents/children with them.

Passenger Details:

Sample records displayed include:

- Names like "Cumings, Mrs. John Bradley", "Futrelle, Mrs. Jacques Heath", and "McCarthy, Mr. Timothy J".

- Gender distribution is shown (male/female).

- Embarkation port is included (C = Cherbourg, Q = Queenstown, S = Southampton).

- Ticket numbers and cabin numbers are listed.

Value Counts:

- Pclass (Passenger Class):

- 3rd class has the highest number of passengers.

- 1st class has the least but most luxurious.

- Survival Count:

- The survival count analysis shows how many passengers survived and how many did not.

Observations:

- Higher-class passengers (Pclass = 1) had better survival chances.

- Women had higher survival rates compared to men.
- Younger passengers (children) also had relatively better chances.

Conclusion:

The Titanic dataset provides valuable insights into survival patterns based on passenger demographics and ticket class. The preliminary data exploration gives a strong base for further machine learning model building if needed.

End of Report