

# Dataset Analysis Report – Titanic Dataset

## Dataset Overview

The Titanic dataset contains information about passengers aboard the Titanic, including demographic details and travel-related attributes. Each row represents an individual passenger, and the dataset is commonly used for supervised machine learning classification tasks.

## Feature Types Identified

The dataset includes multiple data types:

- **Numerical features** such as Age, Fare, SibSp, and Parch.
- **Categorical features** including Sex, Embarked, Ticket, Cabin, and Name.
- **Ordinal feature** Pclass, where passenger class has an inherent order.
- **Binary feature** Survived, which indicates survival status.

## Target Variable

The target variable is **Survived**, a binary variable where 0 represents non-survival and 1 represents survival. This makes the problem a binary classification task.

## Data Quality Observations

The dataset contains missing values in several columns. Age has missing values for multiple passengers, Cabin has a large number of null values, and Embarked has a small number of missing entries. The target variable also shows slight class imbalance, with more non-survivors than survivors.

## Machine Learning Readiness

The dataset is suitable for classical machine learning models after preprocessing. Required steps include handling missing values, encoding categorical features, and potentially scaling numerical features. Due to its size, the dataset is not suitable for deep learning models.

## Conclusion

The Titanic dataset is well-structured and appropriate for understanding data exploration, feature identification, and machine learning readiness. Proper preprocessing is essential before model training.