

Dataset Understanding & ML Readiness Report

Titanic Dataset:

The Titanic dataset consists of 891 records with 12 attributes describing passenger details. The target variable is **Survived**, which is binary in nature. Numerical features include Age, Fare, SibSp, and Parch, while categorical features include Sex and Embarked. Pclass is an ordinal feature representing passenger class.

The dataset contains missing values in Age, Cabin, and Embarked columns, with Cabin having a large number of nulls. There is also a noticeable class imbalance between survived and non-survived passengers. Data preprocessing such as missing value handling, encoding categorical variables, and feature scaling is required before modeling.

Students Performance Dataset:

This dataset contains 1000 student records and 8 attributes related to academic performance and background. There are no missing values, making it clean and easy to work with. Scores in math, reading, and writing are numerical, while other columns are categorical or binary.

The dataset is well-balanced and suitable for regression or classification tasks with minimal preprocessing.

Conclusion:

Both datasets are suitable for machine learning. The Titanic dataset requires more preprocessing, while the Students Performance dataset is ML-ready.