

Titanic Data Analytics Report

1. Introduction

The Titanic dataset is a popular dataset used for exploring machine learning concepts. It contains demographic and travel information for passengers on the ill-fated RMS Titanic, along with whether they survived. This project aims to analyze key aspects of the data to gain insights into survival rates and factors influencing them.

2. Dataset Overview

The dataset contains 891 passengers and 12 attributes. Key variables include:

- Survived: Survival (0 = No, 1 = Yes)
- Pclass: Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd)
- Name, Sex, Age: Passenger details
- SibSp, Parch: Family aboard
- Ticket, Fare, Cabin, Embarked: Travel info

3. Data Cleaning

Missing values were found in the following columns:

PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0

Titanic Data Analytics Report

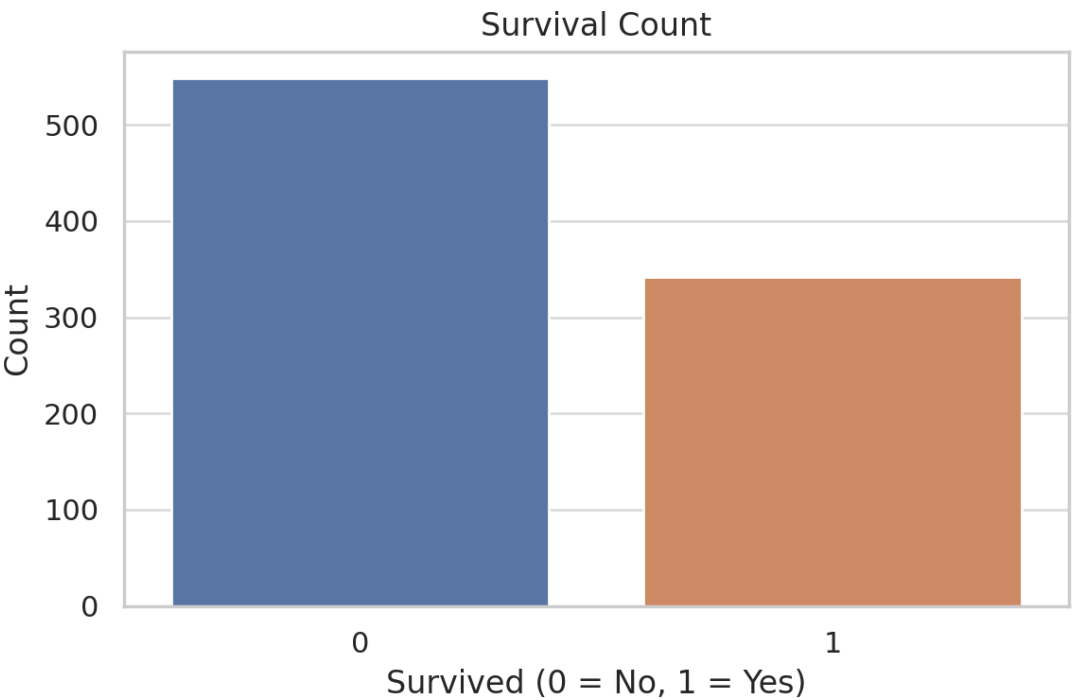
Parch	0
Ticket	0
Fare	0
Cabin	687
Embarked	2

The 'Age' and 'Cabin' fields have significant missing values. 'Cabin' was excluded from analysis due to sparsity. For visualization, rows with missing age were omitted where necessary.

4. Exploratory Data Analysis (EDA)

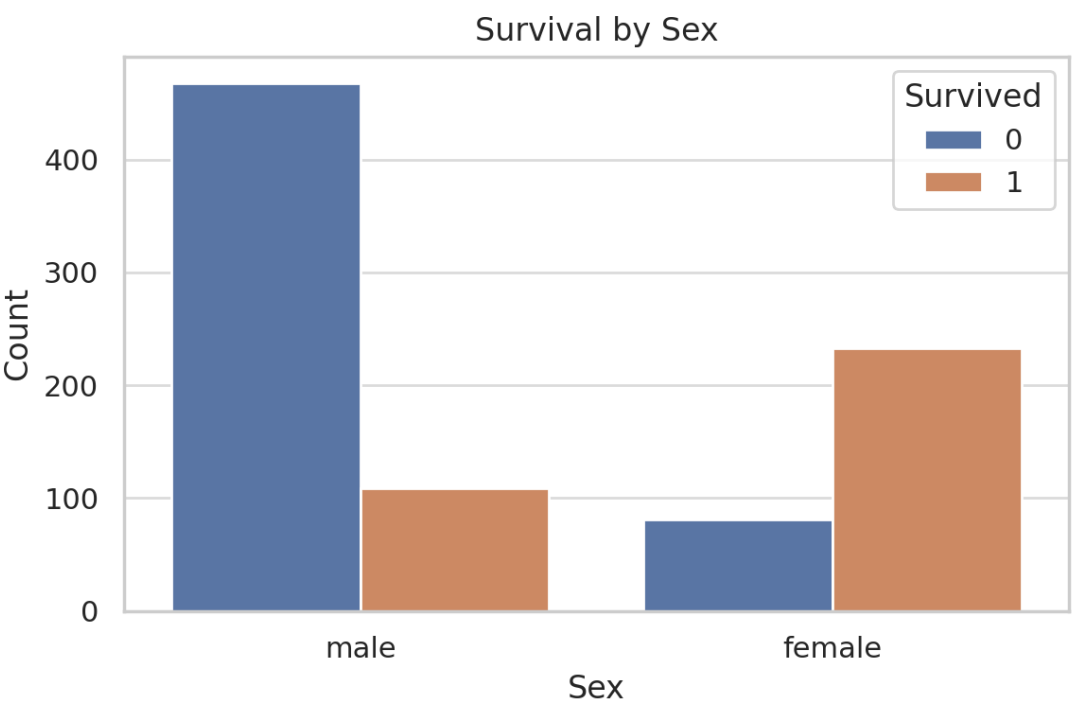
The following visualizations illustrate key trends and patterns in the dataset:

Survival Count

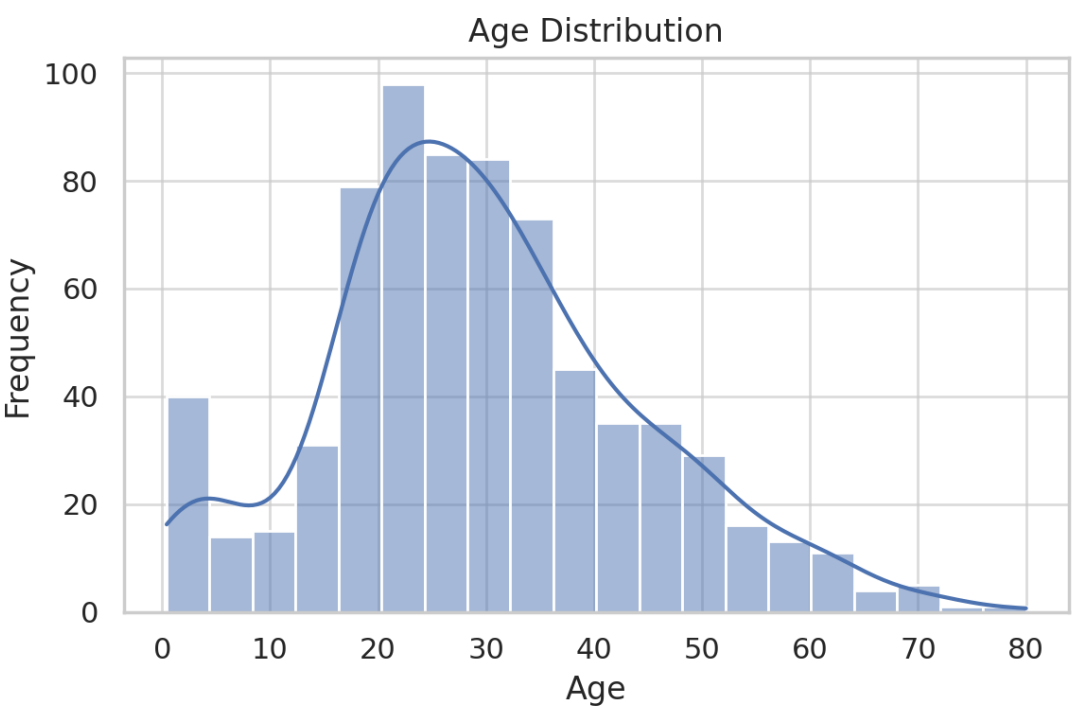


Titanic Data Analytics Report

Survival by Sex

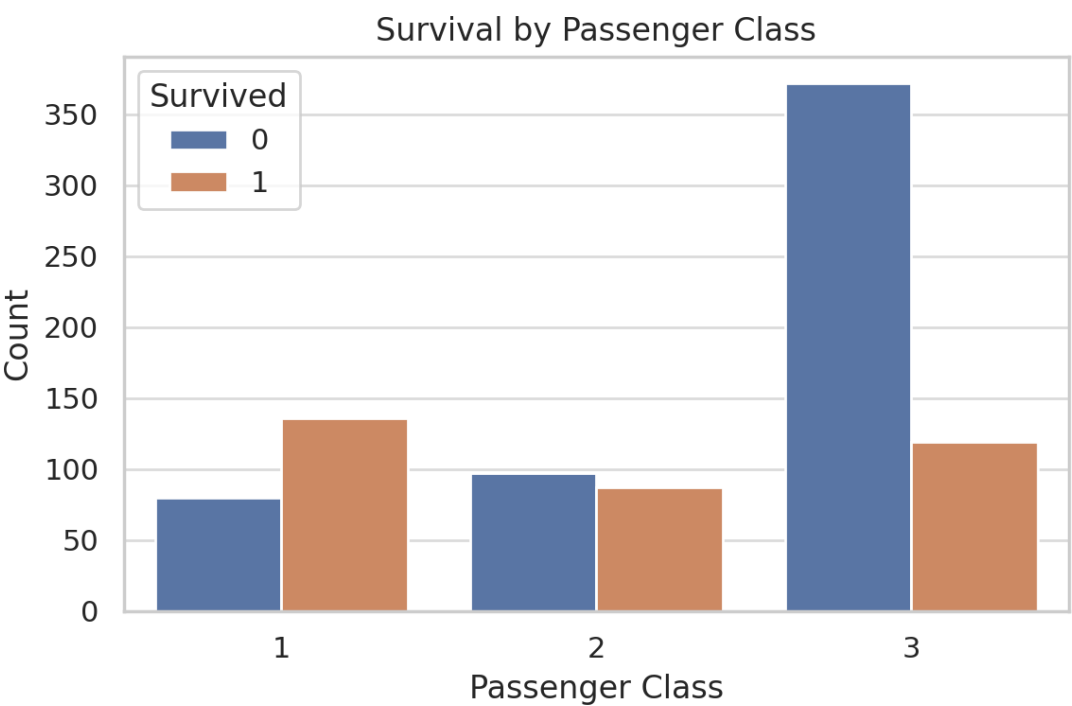


Age Distribution



Titanic Data Analytics Report

Survival by Passenger Class



5. Conclusion

This analysis of the Titanic dataset reveals that gender and class were significant factors in survival. Females had a higher survival rate, as did passengers in 1st class. These insights can help frame further studies, including predictive modeling or deeper demographic segmentation.