Exploratory Data Analysis Report - Titanic Dataset

1. Objective

The purpose of this analysis is to explore and extract meaningful insights from the Titanic dataset using Explorat

2. Tools Used

Python
Jupyter Notebook
Pandas
Matplotlib
Seaborn

3. Dataset Overview

The dataset used is the 'train.csv' file from the Titanic Kaggle competition. It includes features such as Passenge

4. Data Cleaning

Checked for null values using df.isnull().sum()
Noted missing values in Age, Cabin, and Embarked
Suggested basic imputation methods like mean/median for Age and mode for Embarked.

5. Univariate Analysis

Used histograms and boxplots to analyze distributions of individual features such as Age, Fare, and Pclass.

6. Bivariate Analysis

Explored survival rates across Sex, Pclass, and Age using barplots and violin plots.

7. Multivariate Analysis

Generated pairplots and correlation heatmaps to examine relationships among multiple variables and detect mul

8. Key Insights

Females had a higher survival rate than males
Passengers in 1st class were more likely to survive
Higher fare-paying passengers had better survival chances
Cabin feature has a high number of missing values

9. Summary

This EDA provided valuable insights into factors influencing survival on the Titanic. The analysis revealed strong

10. Conclusion

The findings can be useful for feature engineering in machine learning models and deepen our understanding of