Titanic EDA Report

Task 5: Data Analyst Internship
Author: N Nithish Kumar
Date: 30-06-2025

1. Introduction

This report explores the Titanic dataset to uncover insights through Exploratory Data Analysis (EDA). The objective is to understand patterns in passenger survival based on demographic and ticket information using Python (Pandas, Seaborn, Matplotlib).

2. Dataset Overview

- **Source: Kaggle Titanic Dataset**
- **Records: 891 rows × 12 columns**
- **Target Variable: Survived (0 = No, 1 = Yes)**

Missing Values:

- **Age: 177 missing**
- **Cabin: 687 missing**
- **Embarked: 2 missing**

3. Univariate Analysis

Survival Count

- **About 38% survived, 62% did not.**

Age Distribution

- **Majority aged between 20–40.**
- **Outliers exist (e.g., age > 70).**

Fa-re Distribution

- **Right-skewed with very high values paid by a few.**

Passenger Class Distribution

- **Most were in 3rd class, followed by 1st and 2nd.**

4. Bivariate Analysis

Survival by Sex

- **Most females survived, most males did not.**

Age by Class

- **1st class passengers generally older.**

Fare vs Survival

- **Higher fare passengers had higher survival rates.**

5. Multivariate Analysis

Correlation Heatmap

- **Positive correlation between Fare and Survived.**
- **Negative correlation between Pclass and Survived.**

Pairplot Insights

- **Survivors cluster in high fare + low Pclass.**
- **Age and Fare affect survival with visible groupings.**

6. Summary of Insights

1. **Gender: Females had much higher survival rate.**
2. **Age: Younger passengers slightly more likely to survive.**
3. **Class: First-class passengers had higher survival chances.**
4. **Fare: High fares correlate with better survival.**
5. **Outliers: Skewness observed in Fare and Age.**