Titanic EDA Report

1. Introduction

This report presents the findings of an Exploratory Data Analysis (EDA) conducted on the Titanic dataset. The objective was to explore patterns, trends, and relationships among different variables that could influence survival rates. **2. Key Insights**

Key Insights:

1. Women had a significantly higher survival rate than men.
2. Passengers in 1st class (Pclass 1) were more likely to survive.
3. Younger passengers had a slightly higher survival chance.
4. Fare and survival showed a mild positive correlation (wealthier passengers were more likely to survive).
5. The Embarked feature had some missing values filled with the most common port.
6. The Cabin column was dropped due to excessive missing values.

3. EDA Process Summary

EDA Steps Taken:

1. Loaded dataset and checked shape, info, and summary statistics.
2. Identified and handled missing values (Age and Embarked filled, Cabin dropped).
3. Univariate analysis: Histograms and countplots of Age, Fare, Sex, Pclass, Survived.
4. Bivariate analysis: Bar plots and box plots to examine relationships between survival and features.
5. Multivariate analysis: Pairplot and heatmap to analyze correlations and distributions.

Titanic EDA Report

Common EDA Interview Questions (with answers):

1. What is EDA?

   - EDA is the process of visually and statistically exploring data to uncover insights and patterns.

2. Which plots do you use to check correlation?

   - Heatmaps and pairplots.

3. How do you handle skewed data?

   - Apply transformations like log, sqrt, or use robust measures like median.

4. How to detect multicollinearity?

   - Use correlation matrix and Variance Inflation Factor (VIF).

5. What are univariate, bivariate, and multivariate analyses?

   - Univariate: One variable; Bivariate: Two variables; Multivariate: Multiple variables.

6. Difference between heatmap and pairplot?

   - Heatmap shows correlation matrix; pairplot shows pairwise scatter plots.

7. How do you summarize insights?

   - Combine visualizations with clear text-based observations.