Task 02:

Exploratory Data Analysis (EDA) on Titanic Dataset

Objective

The goal of this task was to perform data cleaning and exploratory data analysis (EDA) on the Titanic dataset. The focus was to explore relationships between variables and identify patterns and trends in the data, which could provide insights into the factors affecting passenger survival.

Dataset

The dataset used for this analysis is the Titanic dataset from Kaggle, specifically the train.csv file. This dataset contains the following columns:

* PassengerId: Unique ID for each passenger.
* Survived: Survival status (0 = No, 1 = Yes).
* Pclass: Ticket class (1st, 2nd, or 3rd).
* Name: Name of the passenger.
* Sex: Gender of the passenger.
* Age: Age of the passenger.
* SibSp: Number of siblings/spouses aboard.
* Parch: Number of parents/children aboard.
* Ticket: Ticket number.
* Fare: Ticket fare.
* Cabin: Cabin number (if available).
* Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).
* Process

Data Loading:

The dataset was loaded using pandas to facilitate data manipulation.

Data Cleaning:

* Column Removal: Removed irrelevant columns (Cabin, Ticket) due to high levels of missing data.
* Missing Value Treatment: Filled missing values in the Age column with the median age and the Embarked column with the mode to maintain data integrity.

Exploratory Data Analysis (EDA):

Descriptive Statistics: Generated descriptive statistics to understand the dataset's characteristics.

Visualizations:

* Countplot: Displayed the count of survived vs. not survived passengers.
* Barplot: Analyzed survival rates based on gender.
* Correlation Heatmap: Explored correlations between numerical variables to identify potential relationships.

Findings

* Survival Rates: The analysis revealed notable differences in survival rates based on gender, with a higher survival rate for females compared to males.
* Age Impact: Additional exploration could reveal more insights regarding how age affected survival rates.

Conclusion

This task provided hands-on experience with data cleaning and EDA, essential skills for data analysis. The insights gathered from the Titanic dataset can serve as a foundation for further analysis, modeling, or machine learning applications.