Task 5: Exploratory Data Analysis (EDA) Report – Titanic Dataset

Objective:

The objective of this task is to perform Exploratory Data Analysis (EDA) on the Titanic dataset using Python.
The goal is to understand patterns, trends, and relationships within the dataset and generate meaningful insights through data visualization.

Dataset Description:

The Titanic dataset contains passenger information such as:

- Passenger class
- Age
- Gender
- Fare
- Survival status
- Family details

This dataset is widely used for data analysis and machine learning practice to understand survival patterns.

Tools and Technologies Used:

- Python
- Pandas
- Matplotlib
- Seaborn
- Jupyter Notebook

Data Analysis Performed:

1. Data Understanding

- Loaded dataset using Pandas
- Displayed first few rows of data
- Checked dataset information and statistical summary

2. Data Cleaning

- Checked for missing values
- Handled missing values appropriately
- Removed unnecessary columns where required

3. Univariate Analysis

- Analyzed age distribution using histogram
- Visualized survival count of passengers

4. Bivariate Analysis

- Survival analysis based on gender
- Survival analysis based on passenger class
- Relationship between age and fare

5. Multivariate Analysis

- Generated correlation heatmap
- Created pairplot to understand relationships between features

Key Findings:

1. Majority of passengers were between the ages of 20–40 years.
2. Female passengers had a significantly higher survival rate compared to male passengers.
3. First-class passengers showed higher survival chances compared to second and third class.
4. Passengers who paid higher fares generally had better survival probability.
5. Passenger class and gender were the most influential factors affecting survival.

6. Positive correlation observed between fare and survival, while negative correlation observed between passenger class and survival.

Conclusion:

Exploratory Data Analysis was successfully performed on the Titanic dataset using Python. The analysis helped identify key survival patterns and relationships between different variables.

Visualization techniques such as histograms, count plots, scatter plots, heatmaps, and pairplots made it easier to understand trends and patterns in the dataset.

This project demonstrates the effective use of data analysis and visualization techniques to extract meaningful insights from raw data and support data-driven decision making.