

## Task 5: Exploratory Data Analysis (EDA) Report

### 1. Introduction

This report details the Exploratory Data Analysis (EDA) performed on the Titanic dataset. The objective is to extract meaningful insights from the data using statistical and visual methods with Python libraries: Pandas, Matplotlib, and Seaborn.

### 2. Data Loading and Cleaning

The train.csv file was loaded into a Pandas DataFrame. An initial inspection using `.info()` revealed that the Age, Cabin, and Embarked columns had missing values.

- Age: Missing values were imputed with the median age.
- Embarked: The two missing values were filled with the mode (most frequent port).
- Cabin: The Cabin column was dropped due to a high percentage of missing values.

### 3. Statistical Exploration

Key statistical findings were identified using `.value_counts()` and `.describe()`:

- Survival Rate: Out of 891 passengers, 342 survived, giving a survival rate of approximately 38.4%.
- Demographics: The majority of passengers were male, in the 3rd class, and embarked from Southampton.

### 4. Visual Exploration and Observations

Various plots were generated to visualize the data and identify relationships.

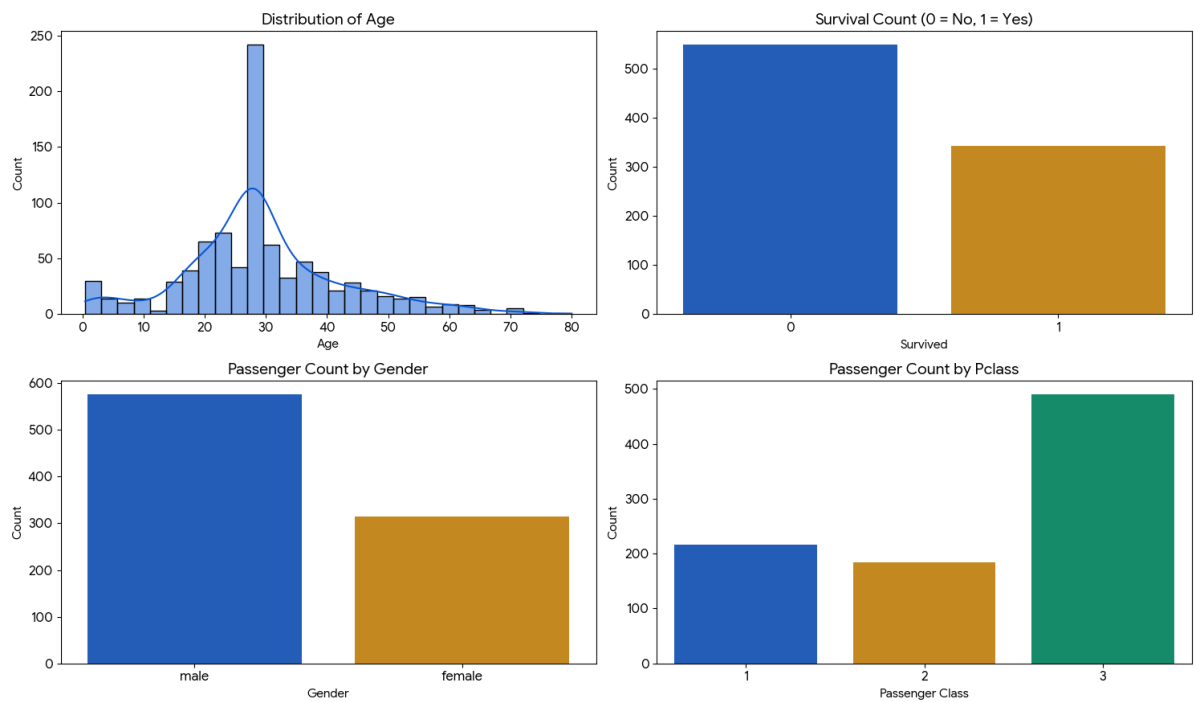
- Age Distribution: The histogram of passenger ages showed that the majority were young adults (20-30 years old).
- Survival Rates: Bar plots revealed a significant difference in survival rates based on both gender and passenger class.
  - Gender: Female passengers had a much higher survival rate than male passengers.
  - Passenger Class: First-class passengers had the highest survival rate, while third-class passengers had the lowest.
- Correlation Heatmap: The heatmap of numerical features showed a positive correlation between Fare and Survived, which aligns with the finding that higher-class passengers (who paid more) had a better chance of survival.
- Age and Survival: A boxplot comparing age and survival suggested that younger passengers had a better chance of surviving.

### 5. Summary of Findings

The EDA concluded that gender, passenger class, and age were the most influential factors in a passenger's survival. These insights can be used for further analysis or to build a predictive

model. The document's purpose is to help you gain skills in finding patterns, trends, and anomalies.

Univariate Analysis of Key Features



Bivariate and Multivariate Analysis

