**Titanic Dataset Analysis Report**

**Objective**

The goal of this project was to explore the Titanic dataset and uncover patterns and factors influencing passenger survival rates. The analysis utilized data visualizations and machine learning techniques to identify trends and relationships among various passenger features, which were critical in understanding survival outcomes.

**Data Sources**

**Data Set :** [**titanic**](https://www.canva.com/link?target=https%3A%2F%2Fwww.kaggle.com%2Fc%2Ftitanic%2Fdata&design=DAFpRxy47kU&accessRole=viewer&linkSource=document)

The analysis was carried out on the following datasets:

* **train.csv**: The training dataset containing survival information and passenger details (used for model training).
* **test.csv**: The testing dataset containing passenger details used for survival prediction.
* **gender\_submission.csv**: A sample of gender-based predictions for the test set.

**1. Data Preprocessing**

The Titanic dataset was preprocessed with the following steps:

* **Handling Missing Data**:
  + **Age**: Missing values were imputed with the median age.
  + **Embarked**: Missing values were filled with the most frequent embarkation port (mode).
* **Categorical Encoding**:
  + **Sex**: Converted to binary encoding (male = 0, female = 1).
  + **Embarked**: Categorical values (C, Q, S) were encoded as numerical values (C = 1, Q = 2, S = 3).

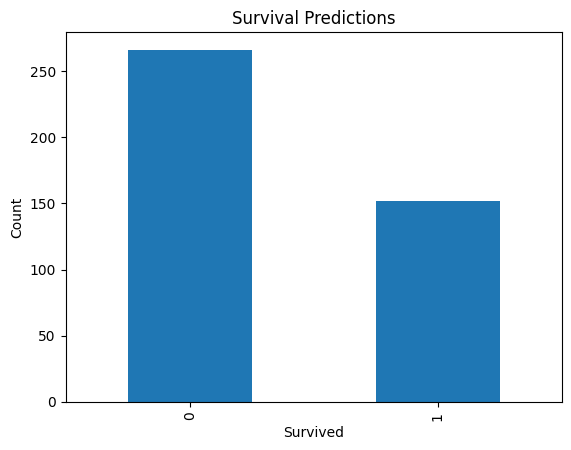
**2. Exploratory Data Analysis (EDA)**

EDA was performed to understand the relationships between variables and identify key survival patterns.

Visualizations and Findings

1. **Survival Distribution:**

* Insight: A higher number of passengers did not survive compared to those who did. This indicated an overall lower survival rate.



1. **Correlation Heatmap**

* Insight:

Sex is highly correlated with survival, indicating women had higher survival rates.

Pclass (Passenger Class) negatively correlates with survival, with lower-class passengers less likely to survive.

Fare positively correlates with survival, showing that passengers who paid higher fares had better chances of survival.

A diagram of heatmap

Description automatically generated

**4. Survival Rate by Gender**

Insight:

Male passengers (label 0) had significantly lower survival rates compared to female passengers (label 1).

A graph with blue squares

Description automatically generated

**5. Survival Rate by Passenger Class**

Insight:

First-class passengers had the highest survival rates, followed by second-class, with third-class passengers experiencing the lowest survival rates.

A graph of a graph with numbers and a bar

Description automatically generated with medium confidence

**Machine Learning Model**

**Algorithm**

A classification model was built using Logistic Regression to predict survival outcomes.

**Steps**

Splitted the data into training and test sets.

Applied feature scaling to normalize numeric values.

Evaluated model performance using accuracy metrics on the test set.

**Conclusions**

Key Factors for Survival:

Gender (females had higher chances of survival).

Passenger class (first-class passengers had the highest survival rates).

Fare (higher fares correlated with higher survival rates).

The dataset highlights socioeconomic and demographic disparities in survival rates during the Titanic disaster.

Suggestions

Balance the dataset for better model training (e.g., using oversampling techniques).

Explore additional machine learning models (e.g., Random Forest, XGBoost) to improve prediction accuracy.

**Visualizations**

The analysis included the following visualizations:

Survival Distribution: Countplot showing survival counts.

Correlation Heatmap: Heatmap highlighting relationships between variables.

Survival by Gender: Barplot comparing male and female survival rates.

Survival by Passenger Class: Barplot comparing **survival rates across classes.**

Predicted Survival (gender\_submission.csv): Bar chart showing the distribution of gender-based predictions.

**Insights**

Gender: Female passengers had a significantly higher survival rate, reflecting priority given to women during rescue efforts.

Passenger Class: Survival probability decreased with passenger class, indicating better access to lifeboats for higher-class passengers.

Fare: Higher fare correlates with higher survival chances, potentially linked to passenger class.

Embarkation Port: Passengers who embarked from Cherbourg (C) had a slightly higher survival rate, possibly due to higher-class demographics.

**Concluding Remarks**

The analysis highlights the influence of gender, socioeconomic status, and passenger class on survival rates during the Titanic disaster. These findings align with historical accounts of the event, emphasizing the priority of women and children in rescue efforts and the advantages of wealthier passengers.