**Exploratory Data Analysis (EDA) on Titanic Dataset Project Report**

**1. Introduction**

The Titanic dataset is one of the most popular datasets for data analysis and machine learning. It contains information about passengers aboard the RMS Titanic, including whether they survived the disaster. This report presents an **Exploratory Data Analysis (EDA)** to uncover insights, patterns, and trends in the data.

**Objectives:**

* Understand the dataset structure.
* Perform data cleaning and preprocessing.
* Analyse survival rates based on different features.
* Visualize key insights using graphs and charts.

**2. Dataset Overview**

The dataset contains **891 entries** with the following features:

| **Feature** | **Description** |
| --- | --- |
| **PassengerId** | Unique ID for each passenger |
| **Survived** | Survival status (0 = No, 1 = Yes) |
| **Pclass** | Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd) |
| **Name** | Passenger's name |
| **Sex** | Gender (Male/Female) |
| **Age** | Age in years |
| **SibSp** | Number of siblings/spouses aboard |
| **Parch** | Number of parents/children aboard |
| **Ticket** | Ticket number |
| **Fare** | Passenger fare |
| **Cabin** | Cabin number |
| **Embarked** | Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton) |

**3. Data Preprocessing**

**3.1 Handling Missing Values**

* **Age**: ~19.8% missing → Filled with median age.
* **Cabin**: ~77% missing → Dropped (too many missing values).
* **Embarked**: 2 missing → Filled with the most frequent value ('S').

**3.2 Feature Engineering**

* **Family Size**: Created by combining SibSp and Parch.
* **Age Groups**: Categorized into Child (0-12), Teen (13-19), Adult (20-59), Senior (60+).
* **Title Extraction**: Extracted titles (Mr, Mrs, Miss, etc.) from names.

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

**4.1 Survival Rate Analysis**

| **Feature** | **Survival Rate (%)** |
| --- | --- |
| **Overall** | 38.38% |
| **Female** | 74.20% |
| **Male** | 18.89% |
| **1st Class** | 62.96% |
| **2nd Class** | 47.28% |
| **3rd Class** | 24.24% |

**Key Insight:**

* **Women and higher-class passengers had significantly higher survival rates.**

**4.2 Age Distribution & Survival**

* **Children (<12)** had a high survival rate (~59%).
* **Seniors (60+)** had the lowest survival rate (~23%).

 *(Example visualization)*

**4.3 Fare Analysis**

* **Higher fares** (1st class) correlated with better survival chances.
* Median fare for survivors: **£26**, vs non-survivors: **£10**.

**4.4 Embarkation Port Impact**

| **Port** | **Survival Rate (%)** |
| --- | --- |
| **Cherbourg (C)** | 55.36% |
| **Queenstown (Q)** | 39.02% |
| **Southampton (S)** | 33.70% |

**Key Insight:** Passengers from **Cherbourg** had the highest survival rate.

**5. Key Findings & Insights**

1. **Women and children were prioritized** during rescue ("Women and children first").
2. **Higher socio-economic status (Pclass 1) had better survival odds** due to cabin location and lifeboat access.
3. **Age played a role**, with children having higher survival rates.
4. **Passengers from Cherbourg** had better survival rates, possibly due to more 1st-class passengers.

**6. Conclusion & Recommendations**

* **Gender and class were the strongest predictors of survival.**
* **Further analysis** could include machine learning models (Logistic Regression, Random Forest) for survival prediction.
* **Feature importance analysis** could help identify key survival factors.

Report Generated By

Sumit Kumar