Analytical Report

on

Titanic Disaster Dataset

Submitted by:  
ARYAN CHOUDHARY

23115024  
B.Tech CSE 2nd Yr

Univariate Analysis

* Age distribution shows peaks at 20-30 and 0-5 years.
* Fare contains significant right skew even after outlier treatment
* Survival rate: 38.4% (342 survived vs 549 perished)
* Majority embarked at Southampton (72.6%)

Bivariate Analysis

**1. Survival Probability by Pclass and Sex (Bar Plot)**

* + **Gender Influence on Survival**: Females have significantly higher survival probabilities across all classes compared to males, suggestive of the "women and children first" policy during the evacuation.
  + First-class passengers (Pclass = 1) have the highest survival probabilities for both sexes, followed by second-class, and then third-class passengers.
  + The disparity between male and female survival is most pronounced in first and second class.
  + In third class, while females still have higher survival rates than males, the gap is narrower.

**2.  Heatmap**

* + Survival is negatively correlated with Pclass (-0.34), indicating that passengers in lower classes were less likely to survive.
  + Fare has a positive correlation with survival (0.32), suggesting wealthier passengers (who likely paid higher fares) had better chances of survival.
  + The variable Cabin\_Missing (indicating whether cabin information was missing) is negatively correlated with survival (-0.32), implying that passengers without cabin information (likely lower-class passengers) had lower survival rates.
  + Pclass is strongly negatively correlated with Fare (-0.72), as expected since higher-paying passengers were generally in first class.
  + Age has a weak negative correlation with Pclass (-0.42), indicating younger passengers were more likely to be in lower classes.
  + **SibSp** (number of siblings/spouses aboard) and **Parch** (number of parents/children aboard) show weak correlations with other variables, suggesting they **had limited influence on survival**.

Multivariate Analysis

**1. Age vs Fare Scatter Plot**

* + The concentration of survivors (orange dots) is higher at higher fares, especially for those in the 20–40 age range, indicating that passengers who paid higher fares had a higher survival rate. First-class passengers had easier access to lifeboats during the disaster, as stated by the socioeconomic order.
  + There is a correlation between fare and survival, as evidenced by the more uniform distribution of non-survivors (blue dots) across lower fare ranges.
  + Younger passengers (children) seem to have a marginally higher chance of surviving, but age does not directly correlate with survival.

**2. Pclass vs Age Strip Plot**

* + The plot shows clear stratification of age across passenger classes:
    - First-class passengers (Pclass = 1) span a wide age range, including older individuals.
    - Second-class passengers (Pclass = 2) have fewer older individuals compared to first-class.
    - Third-class passengers (Pclass = 3) predominantly consist of younger individuals.
  + The distribution also highlights the socioeconomic divide among the classes. Older individuals were more likely in first-class, while younger individuals were concentrated in third-class.

**3. Pclass vs Fare Density Plot**

First-class passengers paid much higher fares, with a peak around £50-75, according to the density plot, which shows clear peaks for fares across passenger classes.   
The second-class fares range from £10 to $25.   
The majority of third-class tickets are under £10.

**4. Age vs Fare by Pclass and Embarked (Facet Grid)**

* **First-Class Passengers**:
  + Passengers embarking from Cherbourg show a high concentration of survivors across all age groups, especially those paying high fares (£50–£60).
  + Few first-class passengers embarked from Queenstown, but survival rates appear high for this small group.
* **Second-Class Passengers**:
  + - * Passengers from Southampton show moderate survival rates across fare ranges (£10–£30).
      * Cherbourg passengers in second-class also display relatively high survival rates compared to those from Southampton.
* **Third-Class Passengers**:
  + Southampton passengers dominate the third-class group and show low survival rates regardless of fare or age.
  + Cherbourg third-class passengers exhibit slightly better survival rates.
  + Queenstown third-class passengers have very low survival rates, likely due to limited resources and lifeboat access.
* Fare remains a critical factor, with higher-paying passengers generally surviving more often regardless of embarkation port or class.
* Cherbourg passengers across all classes had better survival rates compared to rest.

**Conclusion:**

Survival was influenced by multiple factors including class, fare and age. First-class passengers had better survival chances due to their proximity to lifeboats and preferential treatment during the evacuation, while survival probability was significantly less in lower classes and among passengers who paid lower fares.