# CASE REPORT: TITANIC DATASET EDA (EXPLORATORY DATA ANALYSIS)

This report presents the results of an Exploratory Data Analysis (EDA) conducted on the Titanic dataset. The primary objective was to determine the factors influencing the survival rate of passengers. Three key variables were investigated: gender, passenger class, and age. The analysis used count plots, chi-squared tests, histograms, and box plots to draw meaningful conclusions regarding the relationships between these variables and survival rates.

1. Gender and Survival Rate:
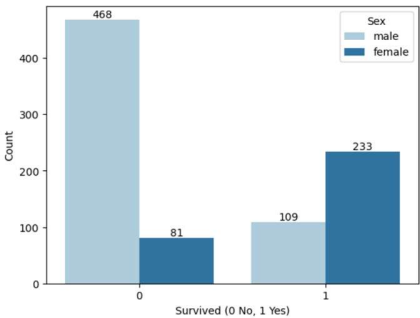
H_0: *The survival rate and gender are indepedent*

H_a: *The survival rate and gender are not indepedent*

$Chi-Squared$: 102.88898875696056

$P-Value$: 4.549251711298793e − 23

$Degrees\ of\ Freedom$: 2

$Decision\ Point\ (0,05)$: 5.99

- **Data Exploration:** A count plot was used to visualize the relationship between gender and survival rates. The plot showed that women had a significantly higher survival rate than men.

- **Statistical Test (Chi-squared):** The chi-squared statistic (260.72) exceeded the Decision Point (3.84), leading to the rejection of the null hypothesis (H0) and the acceptance of the alternative (Ha). Indicating that survival rate and gender are not independent.

- **Conclusion:** Based on the count plot and chi-squared test, it can be concluded that survival rates are significantly associated with gender, and women had a greater chance of surviving.

2. Passenger Class and Survival Rate:
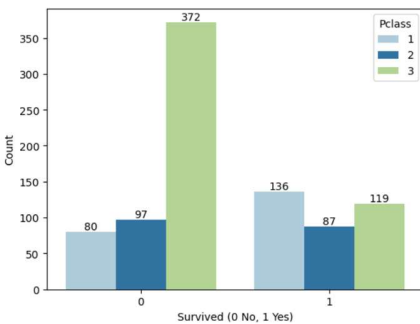
H_0: *The survival rate and passenger class are indepedent*

H_a: *The survival rate and passenger class are not indepedent*

$Chi-Squared$: 260.71702016732104

$P-Value$: 1.1973570627755645e − 58

$Degrees\ of\ Freedom$: 1

$Decision\ Point\ (0,05)$: 3.84

- **Data Exploration:** Similar to the gender analysis, a count plot was employed to visualize the relationship between passenger class and survival rates. The plot revealed an inverse relationship between class and survival rate.

- **Statistical Test (Chi-squared):** The chi-squared statistic 102.89) exceeded the Decision Point (5.99), leading to the rejection of H0 and the acceptance of Ha. This indicates that survival rate and passenger class are not independent.

- **Conclusion:** The analysis found a clear association between passenger class and survival rate. The survival rate was highest in the 1st class (62.9%), followed by the 2nd class (47.3%), and was lowest in the 3rd class (24.2%). Passengers in 3rd class had a significantly lower chance of surviving.

3. Age and Survival Rate:
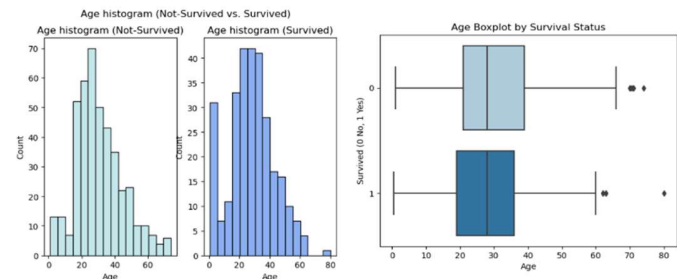
- **Data Exploration:** Histograms and box plots were used to explore the relationship between age and survival rates. These plots highlighted that the median age of survivors and non-survivors is similar. However, the analysis revealed that the survival rate was significantly higher for passengers aged 0 to 5 and notably lower for passengers older than 60.

- **Conclusion:** While the median age of survivors and non-survivors is comparable, the survival rate is noticeably higher for passengers aged 0 to 5. On the other hand, passengers older than 60 had a considerably low survival rate, with very few counts, except for some outliers. This suggests that age is related to survival rate, possibly due to the "Women and Children First" policy.