

# Titanic Data Analysis Report

**Assignment:** Statistics and Trends (20%); **Module:** Applied Data Science 1

**Name:** Deepak Raj Manickam

**Student ID:** 23070139; **Student Email ID:** [dm24aav@herts.ac.uk](mailto:dm24aav@herts.ac.uk)

**GitHub Repository:** [https://github.com/deepakraj-04/Applied\\_Data\\_Science\\_1](https://github.com/deepakraj-04/Applied_Data_Science_1)

**Introduction:** We are analyzing the Titanic dataset to learn more about the passengers and what might have influenced their chances of survival. To ensure accuracy, the data must be cleaned. Basic statistics methods are used to gain an understanding of the dataset. We will create visualizations to see important patterns and trends.

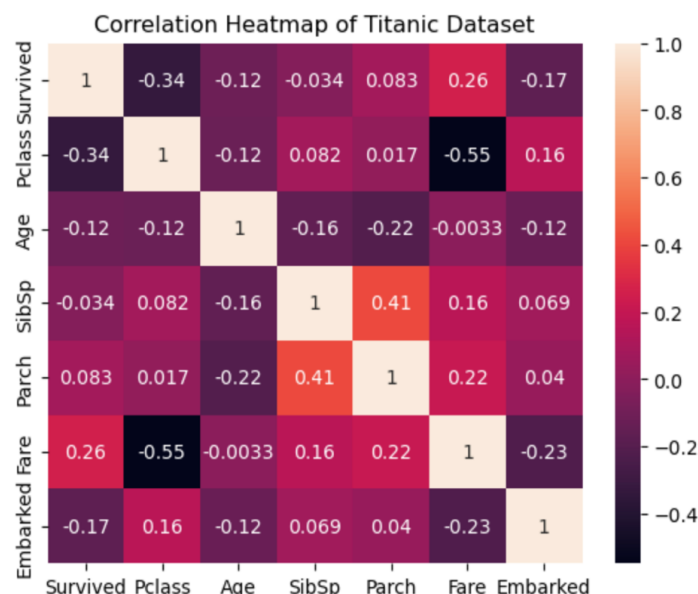
## Statistical Methods:

**Descriptive Statistics:** The describe() method gives us a summary of the dataset, including count, mean, standard deviation, minimum, maximum, 25<sup>th</sup> percentile, 50<sup>th</sup> percentile and 75<sup>th</sup> percentile values for numerical columns. The summary shows significant variation in 'Fare' and 'Age' which may influence survival rates.

**Skewness:** The skew() method is used to calculate the skewness of data. Some columns like 'SibSp' and 'Parch' are highly skewed, which indicates that most of the passengers had few or no companions, with a few having many.

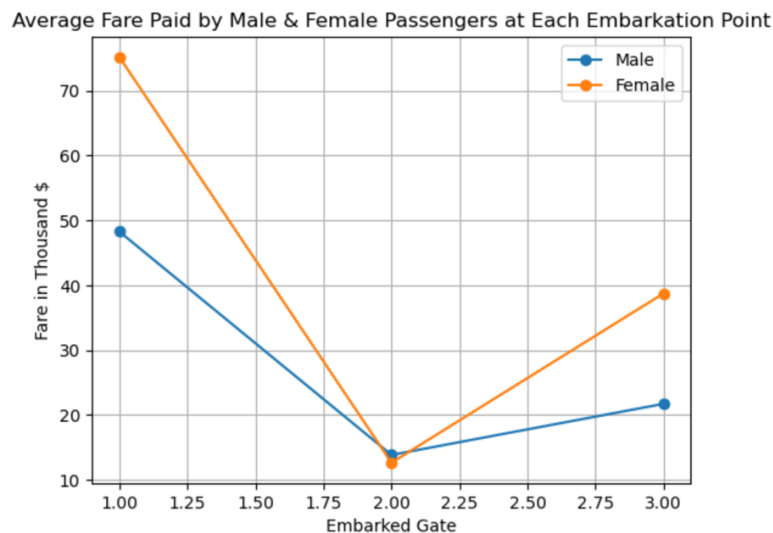
**Kurtosis:** The kurt() method is used to calculate kurtosis for each numerical column, in which it shows high kurtosis values for the columns 'SibSp' and 'Parch', indicating a peaked distribution.

**Correlation Analysis:** The corr() method is used to calculate the Pearson correlation between numerical columns to calculate how they are related to each other.

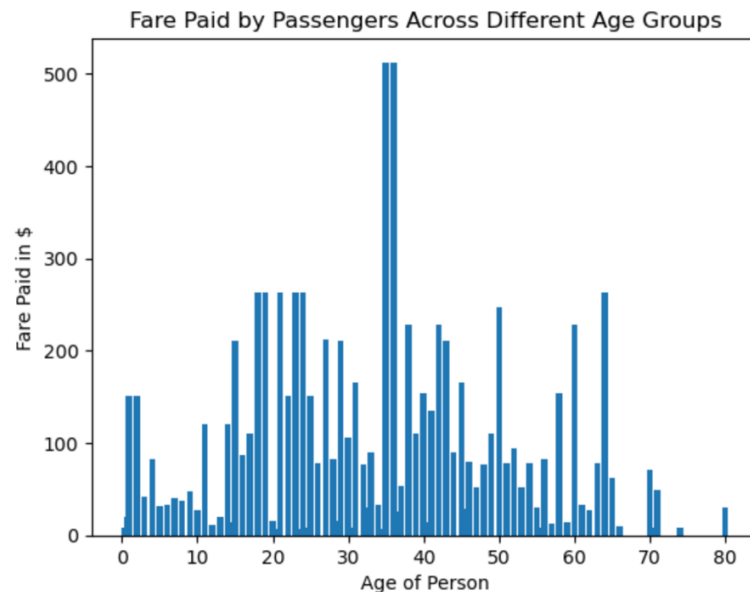


A heatmap is created using seaborn to visualize correlations with dark colors indicating strong relationships, such that there is strong negative correlation between 'Pclass' and 'Fare', which indicates that higher fares were paid by passengers in lower classes.

**Line Graph:** This plot compares the average fares paid by male and female passengers at different embarkation points on the titanic, which achieves by plotting two lines on graph for male passengers and female passengers paid at each embarkation point.



**Bar Chart:** The bar chart shows how prices changed with age by visualizing the relationship between the passengers ages and the fares they paid.



**Conclusion:** The connections between passenger demographics and survival odds are shown by our analysis, which offers insightful information about the Titanic dataset.