# Approach Document

The dataset was already divided into two parts: one for training a model and one for testing its effectiveness. I ensured text data was handled properly, then extracted surnames to create unique group IDs for passengers traveling together. Next, I categorized passengers based on gender and title, focusing on "Master" for young boys. For non-male passengers, I created groups based on their IDs, likely to study their survival rates as a unit.

Using ggplot2, I explored the data, visualizing how survivors and non-survivors were distributed across various categories like class, age, and embarkation point. To thoroughly assess the model's performance, I ran it through a 10-fold cross-validation process over 25 trials. This meant randomly dividing the data into 10 parts, training and testing the model 10 times, each time with a different part acting as the test set.

Finally, I developed a prediction logic that combined gender and group survival rates. This was key to determining the model's predictions based on the features I created. By calculating the survival rates of these groups, I aimed to improve the accuracy of individual predictions, especially for family members or those traveling together. Based on sex, title, and calculated group survival rates, I made predictions for each passenger in the test dataset, printing specific results for both males and females.