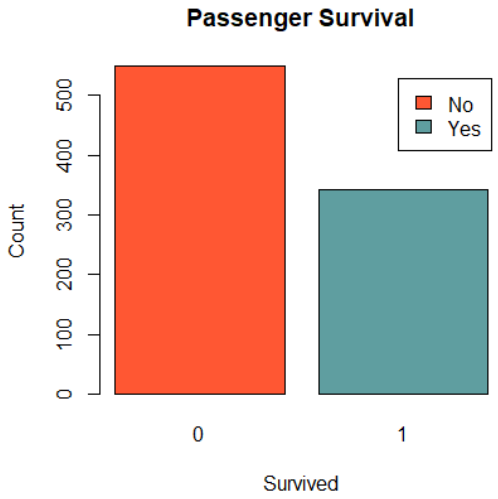
Predicting Survival on the Titanic

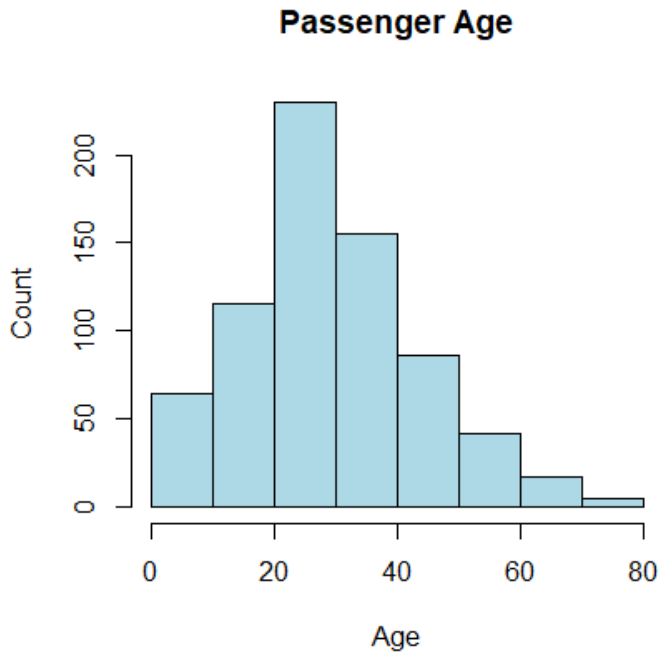
Name: هادي احمد عبدالسلام عبدالحميد
Id: 20191700728
Dpt.: CS
Section: 6

Data Exploration:

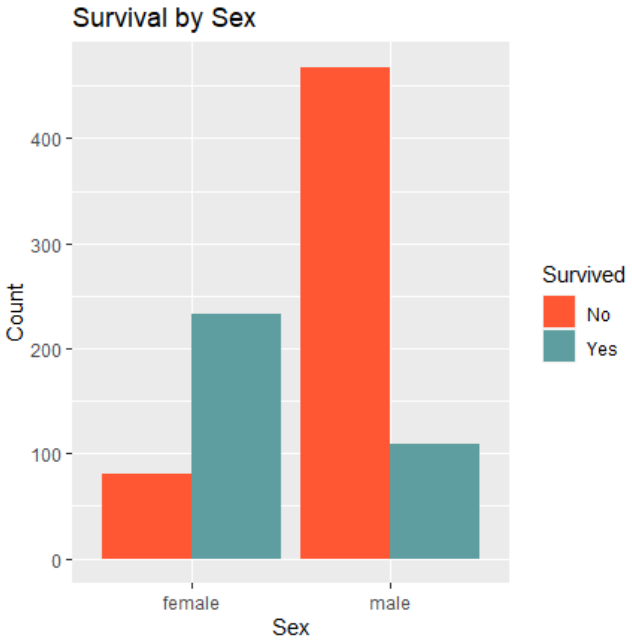
1. Count the number of passengers by survival status.



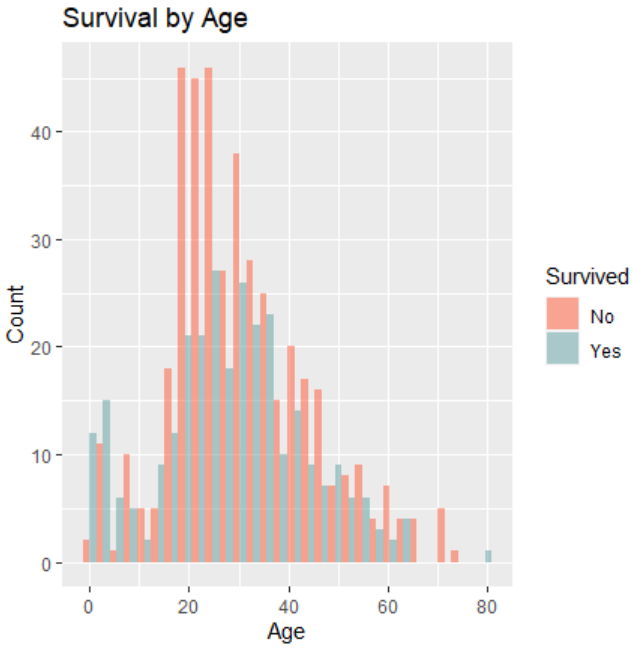
2. Create a histogram of passenger age.



3. Bar plot for categorical variables: For variables like "Sex" and "Embarked", you can create a bar plot to see the proportion of survivors for each category. Here's an example code for the "Sex" variable:



4. Histogram for continuous variables: For variables like "Age" and "Fare", you can create a histogram to see the distribution of survivors and non-survivors. Here's an example code for the "Age" variable:



5. Box plot for continuous variables and categorical variables: For variables like "Pclass" and "Embarked", you can create a box plot to see the distribution of survivors and non-survivors for each category. Here's an example code for the "Pclass" variable:

Preprocessing

Based on the Data Exploration this preprocessing is done:

- Dropped unnecessary columns like "PassengerId", "Name", "Ticket", "Cabin", and "take-off".
- Applied one-hot encoding for the "Sex" column.
- Changed some columns into factors like "tack. off" and "Pclass".
- Filled the null values in both the "Age" and "Fear" columns with the mean.

Used Models

- Random Forest Model
- Logistic Regression
- Naive Bayes Model

Results

This Results is Based on the Kaggle summations of the Test data.

The **Random Forest** Model achieved the **highest accuracy** with a score of **79%,** followed by **Logistic Regression** with a score of **76%**, and **Naive Bayes** Model with a score of **74%**.