

**TASK**

**Exploratory Data Analysis on the Titanic Dataset**

[](http://www.hyperiondev.com/portal/)

**Introduction**

The dataset chosen is the Titanic dataset. The titanic dataset contains information about the passengers aboard the RMS Titanic, which sank in its maiden voyage in 1902. The dataset includes variables such as passenger age, gender, class, fare and whether thay survived the disaster.

**DATA CLEANING**

This dataset did not require much of data cleaning. The steps I took to clean the dataset was I first use “isnull().sum()” to calculate the number of null data in each column. I then proceeded to clean the dataset.

**MISSING DATA**

The dataset contains missing values in the Age, Cabin, and Embarked columns. Here’s how I handled them:

* Age: Imputed with the median age of passengers.
* Cabin: Dropped due to a high percentage of missing values.
* Embarked: Imputed with the mode (most frequent value).

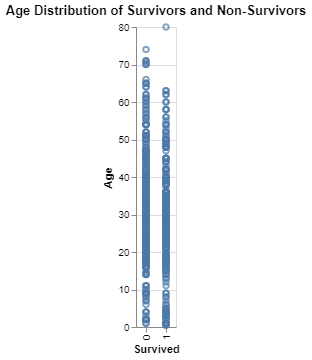
**DATA STORIES AND VISUALISATIONS**

Get survival rate of the passengers by comparing passenger class and gender



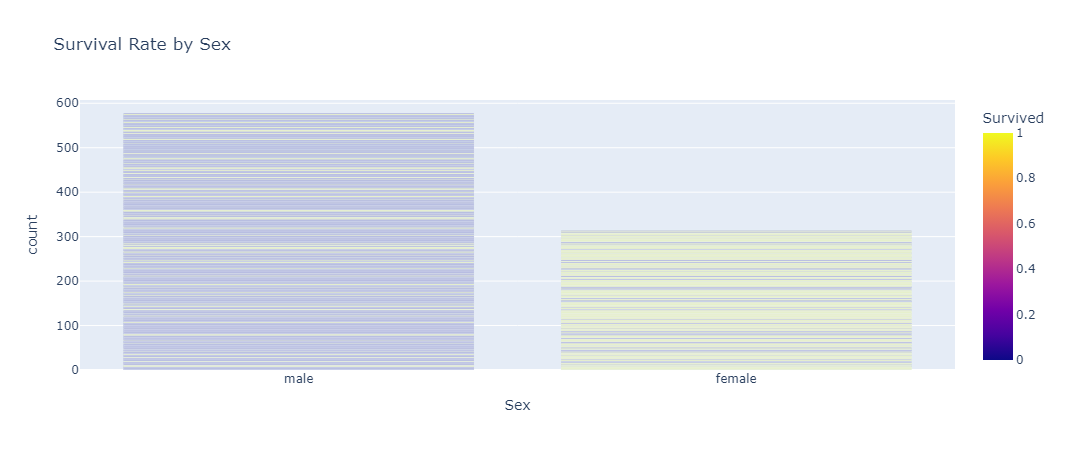
The data visualization highlights that, Females, particularly those in first class, had a much higher chance of survival compared to males and those in lower classes. This suggests that social status and gender played crucial roles in determining the likelihood of survival.

Age Distribution of Survivors and Non-Survivors



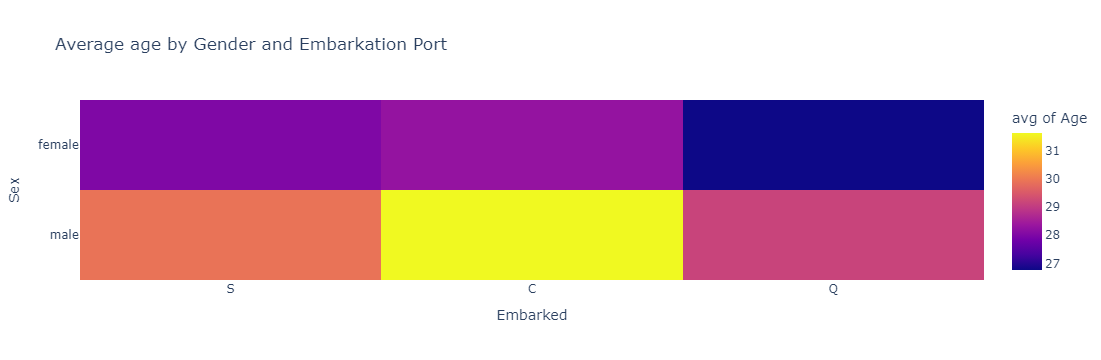
The data visualization highlights the impact of age on survival rates during the Titanic disaster. Younger passengers had a higher chance of survival, which could be attributed to various factors such as physical agility, prioritization during evacuation, or accompanying adults ensuring their safety.

Bar plot for survival rate of the titanic passengers by gender.



The data visualization highlights a significant disparity in survival rates between males and females during the Titanic disaster. Females had a much higher chance of survival compared to males, which could be attributed to the “women and children first” policy during the evacuation.

Highlights differences in the average age of passengers based on their gender and the port from which they embarked.



The data visualization highlights that on average, the old people were from Cherbourg(C), most of young passengers were from Queenstown(Q). and lastly the heatmap suggest that the middle-aged group mostly came from Southampton(S).

**THIS REPORT WAS WRITTEN BY : SPHELELE KHUMALO**

