## Summary:

For the Data Visualization project, I have chosen to explore Titanic Data Set.

Questions I am looking to analyze are:

1. What is the passenger's age frequency?
2. Which age groups survived better?
3. How does survival rate compare between Gender?
4. How does survival rate compare between PClass?
5. Is there any relation between Fare and Survival?

I have used Tableau Public 10.4.2 for this project. The final published story version is available @ https://public.tableau.com/shared/6BK89XZ8Z?:display_count=yes

The initial iteration is available for reference @ https://public.tableau.com/views/TitanicStoryV1/TitanicSurvivalStory?:embed=y&:display_count=yes

## Design:

I chose the Tableau Story to take the viewers through the series of visualization as the Viz attempts to answer through the series of questions.

**Viz #1:** The idea was to understand the spread of the passengers age and I went with the histogram for this viz. I bucketed the Age measure into Age bin using Create Bin option in Tableau

**Viz #2:** Further, I attempted to understand the survival rate by age of the passengers. In order to achieve this, I created a calculated field called Percentage of Survived Passengers as Survived / Total Number of Records.

I then plotted the Age Bin Vs agg(Percentage of Survived Passengers) to study the same.

**Viz #3:** I then sliced the data by gender within the Age bin and attempted to understand if within the same age, one gender survived better than the other.

In order to get the viz, I used the measures Age bin, Gender and plotted against the Total number of records. I sliced the Total number of records by Survived measure using the color Mark.

**Viz #4:** Next, I wanted to understand if there is any relation between PClass and the Survival rate. I used the Percentage of Survived Passengers and the absolute numbers (Survived / Total number of records) to give both the percentage and number of passengers under each PClass category.

**Viz #5:** Finally, I wanted to understand if there is any relation between Fare bin and the survival rate. To achieve this, I chose the scatter plot of Fare Bin Vs Percentage of Survival rate with the Total number of passengers as the size Mark.

Please note that I have noted my inference within the published Tableau story.

## **Feedback:**

1) Is the Age bin discrete?

   Action Taken: I had increased the bar width to give the impression of continuous dataset.

2) Feedback on Viz 2: Would it be possible to show the nums and dens as well?

   Action Taken: I realized that Percentage of the Survival rate may only reveal part of the story and hence included the Survived passenger count (Numerator) and Number of Records (Denominator) in the tool tip so that the absolute numbers are less intrusive and available to users on hovering action.

3) Feedback on Viz 3: Is there a way to know which color or value represents survived?

   Action Taken: I edited the Legend title and explicitly added the description as (0 – No, 1- Yes) with corresponding color as the Stacked bar chart

4) Feedback on Viz 4: Please include the numbers so it is easy to pick up the details.

   Action Taken:
   A) I added the Percentage of the Survival Rate in the chart 1 & the Suvived, Total Number of Records count on chart 2
   B) Also, I moved the percentage chart from bottom to top as I realized that that was primary message that I wanted to convey from the viz. the Survived / Total number of records chart was supplementary to provide insight into the PClass analysis
   C) Also, added the Gender and Age bin filter to compare if the generic Pclass trend is applicable when broken down by Age bin / Gender. It offered interesting insight.

Resources:

https://community.tableau.com/

https://www.kaggle.com/c/titanic

**Data Files:**