Titanic Data Analysis and Visualisation

Initial Version:
https://public.tableau.com/profile/arvind.raj5951#!/vizhome/Titanic_451/TitanicVisualizationI

Final Version:
https://public.tableau.com/profile/arvind.raj5951#!/vizhome/Titanic_451/TitanicDataVisualisation

Summary

The RMS Titanic ship sank in the North Atlantic Ocean in 1912 after colliding with an iceberg, killing more than 1,500 of the 2,224 passengers aboard. The given data consists of information about 891 passengers who travelled in Titanic ship.

The main aim of the visualisation is to find the likely factors which may have influenced the death or survival of the passengers.

Design

Cleaning

The given data had some issues. The age of passengers weren't available to all the passengers. So as to differentiate null from other values I created a calculated field named 'Is Adult' which separated the passengers in terms of their age. If a passenger's age is below 18, they were referred to as child and if their age is above 18, they were referred to as adult. The passengers without the information on their age were just categorized as Null. And for other calculation purposes, I treated the null values as zero.

The data in survived column were in binary format. The values were zero for dead and one for survived. I created a calculated field named 'Survival' which changed the values 0 and 1 into died and survived respectively.

A bin was created for the age column without the null values and the bin value was set as 10.

Visualisation

The story had 3 slides.

1. The first slide had a bar chart which gave a general information about the list of the passengers who died and survived.
2. The second slide had a dashboard. The first bar chart in it separated the passengers according to their age and gender. The second one made a slight change to the above graph. It just modified the above graph such that it showed the list of passengers who died or survived according to their age group and gender. Finally two filters were added. 'Sex', so as to find data about a particular sex. 'Is Adult', so that we could find the data about those passengers with age as null.
3. The third slide had a dashboard too. It had the general information about the list of passengers who died and survived and then it had the same information with an added variable 'gender'. Then the passengers were categorized according to their classes. Then the information from the previous graph was added to it i.e. the passengers were

categorized by the variable ‘Survival’ and the list of men and women in each class were added to it. So we get a graph which showed the number and men and women from each class who survived or died.

Feedback

Initial version received the following feedback.

It was brought to my attention that I could explore more with the age and gender i.e. categorizing them into child and adult according to their age and their gender and add a visualisation accordingly.

Design

Cleaning

I wanted to create a new calculated field that included both the gender and age of the passengers. So, I made use of the already created field ‘Is Adult’ and the existing field ‘Sex’. I created a set of conditional statements which categorized male to man and boy if their corresponding ‘Is Adult’ value is Adult and Child. If the value was Null, the Age-Gender value was set as Male. A similar conditional statement was created for female category. The values of Age-Gender were set as woman, girl and female if their corresponding ‘Is Adult’ was Adult, Child and Null.

Visualisation

The story had 4 slides. The first 3 slides were same as the initial visualisation.

4. The final slide was also a dashboard. The first bar chart consisted of the count of each category of passengers namely man, woman, boy, girl. The next chart consisted of the same data with the numbers of people who survived and died in each category. The final chart consists of the details of the people who survived and died according to their Age-Gender and the class in which they were travelling in. I added the filters on the variable ‘Age-Gender’ variables so that we could check the details for each and every category of ‘Age-Gender’.