

Titanic Analysis using Tableau

Summary

The goal of this project is to highlight the reasons using plots which may have led people to either survive or die in the Titanic disaster. Out of the various features available like Pclass, Embarked, Family members on board, Sex, Ages, etc we will look some particular ones that will lead us to conclusion.

Link-

<https://public.tableau.com/profile/prathamesh8818#!/vizhome/TITANAICANALYSISv1/TitanicAnalysisv1?publish=yes>

<https://public.tableau.com/profile/prathamesh8818#!/vizhome/TITANAICANALYSISv2/TitanicAnalysisv2?publish=yes>

Design

While going through the tableau it will be evident that I have used bar graphs mostly to present my data. There are two underlying reasons that support this kind of usage.

1. Understanding When I first started out with the lesson, being curious I immediately searched for some awesome tableau projects and I found that, even though the plots were nice and beautiful, it was a lot of information and lot of minute details in the form of marks. So, if I want to communicate this to a wider audience who are not used to visualizations, I keep it simple.
2. Data which I picked up is very easy dataset and with my level of exposure to the tool, I felt constructing bar graphs was far easier and useful.

Details of each of the story points:

1. The first story point is actually a dashboard that has data about males and females who boarded the Titanic. The reason I split the y axis into two to show survival and class because, I wanted them to fall in the same plot but with a differentiating line which lead me to a split plot within a plot
2. The ages story was ok with a distribution in terms of a histogram. But, I felt it would be better to show to the people how many of a certain age group survived or died. The red and green color are chosen to show danger and safe zones in general and hence green to encode survival and red to encode death.

3. The bubble chart to show the group was a experimental choice for me as I could show this using a bar graph as well. But, I think the measure of size fits well here as there is significant change in sizes of the bubbles per class.
4. Since there were 3 classes of passengers, we would want to see if places they boarded from showed anything about the affordability of the passengers and hence the classes. It seems, max number of people embarked from Southampton and were placed in class3. However, there can be an argument that, the population of people from Southampton was more in other classes as well. But I would say fairly higher in class3.
5. It was expected that, since maximum people boarded from Southampton, we would see more casualties from there and that is clearly seen in the graph that mentions survival for C or Q or S.
6. We go a step further to deduce that, the death rate for people from class3 and from Southampton was considerable higher, whereas the death rate for people from class1 was lower or proportionally less. The people from class1 fared well irrespective of the places they embarked from.

Calculated fields

1. **Ages:** Since there were some null values, I converted them to 0's and created a new field.
2. **Family members:** Since there were Sib Sp and Parch as different fields, I decided to better keep one field which had all the data about the family members, hence this field is, Sib Sp + Parch
3. **Has Dependents:** In case a person has number of dependents > 0, then the field is 'Yes' else 'No', almost like a boolean
4. **Has Family Members:** In case a person has number of Family Members > 0, then the field is 'Yes' else 'No', almost like a boolean

Feedback

I received feedback about my viz from 2 people. Below were some of points they mentioned and I incorporated the changes accordingly

Feedback

I feel below questions should be modified for better grasp of the chart that follows:

- Q1. A family always supports you, even on Titanic
- Q2. Did people from certain class belong to certain place

Q3. Also, chart explaining 'how many people had how many dependents' is a bit misleading when you hover anywhere inside a circle as it shows 'Dependents 0 Family members 161'.

Below are the modifications:

Q1. A family always supports you, even on Titanic: This has now been changed to: **Does having a family member actually increase your chances of survival?**

Q2. Did people from certain class belong to certain place This has been modified to: **Was the Pclass of the people in any way related to places people boarded the Titanic from?**

Q3. Also, chart explaining 'how many people had how many dependents' is a bit misleading when you hover anywhere inside a circle as it shows 'Dependents 0 Family members 161'. This was by far a very important the reader pointed out. When I cross checked, I almost changed 2 of my visualizations and also modified a calculated field. Also, I changed the count variable to *total records* which made more sense rather than *sum(Family Members)*.

Prior to feedback: I renamed the *Sib Sp* field to *Dependents*. I plotted 2 charts which showed whether having a dependent on board increases chances of survival.

After the feedback: What I actually wanted to point out was that, whether having a family member (parents, children, spouse, siblings) affect the survival. I created another field and named it *Family Members* which is a sum of *Sib Sp* and *Parch*. Also, I had created field called *is Dependents* which has true or false depending on whether a person has a dependent. Now the new field I created is named *Has Family Members* which sounds correct rather than *is Family Members*. This feedback point actually helped me change my plots which were not that meaningful.

Resources

<https://www.tableau.com/learn/training>