Project: Create a Tableau Story - Titanic Data

Final: (After Feedback) - Titanic Tableau Story Project

Summary

Titanic data provided by Udacity is a subset of the demographics and passenger information as well as the crew in the Titanic incident. On April 15, 1912, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. However, the dataset for this only includes 891 rows. The goal of this project to build a Tableau story that illustrates the demographics or passenger information between those passengers who survived and those who did not survive.

Design

- In my initial submission, I mainly focused on the number of survived passengers to investigate the effects of the demographic factors.
- As I shared my first attempt with my friends and some colleagues, and received their feedbacks, I realized that I should find a way to base the comparisons with survival rate or the chance of survival. A simple way was to consider Average of the Survived variable.
- After considering the survival rate variable, I initially looked at individual factors such as Gender, Ticket Class, Age Group, Embarked to see the impact on the survival rates.
- Then, I tried to see the effect of multiple factors on the survival rate. For instance, I could find some interesting insights when I looked at both Gender and Ticket Class effects on Survival rates.
- As the important variables are mostly categorical, I used bar charts for looking at how the factors affect the survival rate.
- For the initial file, I had a dashboard as part of my story. However, as a friend suggested: “The variables are self-explanatory and all the charts are bar-

charts, so there is no need for building a dashboard." Which was sounded right to me.

- I reformatted the font size and colors for the chart labels (X and Y as well as the chart titles) to make them more readable and look nicer.

Feedback

- The most important feedback I received (almost from everyone) was changing the number of survived people to chance of survival in the charts which was a significant change. Then I used AVG for the Survived variable instead of SUM. This gave me an average survival rate for each groups/ factors.
- The next very important feedback was the one a friend suggested: "…Age can be categorized to three groups: Children, Adults and seniors…" visualization Based on his feedback: I categorized ages into three groups : Children (18 and younger), Seniors(50+), and the rest.
- It was not clear for a colleague that what 0 and 1 were representing in the Survived versus Number of Records chart (The first chart) and he asked: "What does 0 and 1 mean in the chart?" Using this feedback, I changed 0 to dead and 1 to Survived.
- Based on the feedback: "The variables are self-explanatory and all the charts are bar-charts, so there is no need for building a dashboard." I removed the dashboard that I had in my initial file.
- There were a few feedbacks regarding the X-axis in a few charts to make them more clear, for instance:
    - ✓ Removed Null in the Embarked from visualization, and put the complete names of the cities (e.g. Queenstown instead of Q)
    - ✓ Replacing Ticket Class instead of Pclass
- Based on the feedback: "…Gender is the most important factor in the data and it deserves its own chart…", Separated Genders from the first chart and added a new chart (the second chart).

- Furthermore, based on the same feedback as above, I ended up including the sex factor into the original Ticket Class cart. As explained in the final version of the story, the result was interesting.
- " Some of the graphs look too thin…You can drag the X-axis chart to make them larger so they look nicer." I did as suggested in the feedback and the charts looked much nicer.
- "Please include more explanations in the story, so that it tells a story!". That was a great feedback from my boss that I benefitted from.

Additional Feedbacks:

- "Slide 1. Using different color to imply dead/alive could be more engaging… Or even better you could present this as a pie chart". Although the instructor in the video lectures of the Udacity is in general hesitant to use Pie-charts, but I believe here could be an exception and found that Pie-chart is useful in my case. Also, another suggestion about assigning colors to imply Dead or Survived was a great feedback and I used it a couple times.
- "Slide #2: consider using different color for male/female…" I used other provided feedback and completely re-designed this slide. I believe there is no problem in the new slide regarding this feedback.
- "Slide #3: this is good…Please be careful about colors. We want to avoid color confusion.) I have made the chart larger and also changed the colors to avoid color confusion. Please note that as I redesigned and rearranged the slides in the new file (named as Titanic Project-Final (After Feedback)) this is on slide 4.
- Slide #4 and #5: "good except color. Consider switching to a different set of colors. Also percentage is more effective than fractions. Please think about it." I have changed the dark blue color to a light blue one. I am not quite sure if the feedback is meant to say set different colors for each age category, or just the dark blue is not the best option and change it to another color. I remember from the video lectures that is saying if a bar chart is not too complicated and self-explanatory by only looking at the bar titles, there's no

need to add color for each category and the associated legends. In my case, there are only 3 categories and I believe it would be the same case as explained in the lecture and it's better to not add legends and multiple colors. I also added the percentages in the y-axis. The age comparison is in the slide 3 of my new file. Also, I have removed slide 5 as explained for the following feedbacks (I explain what I did at the end):

- ✓ "…Please note that this is an explanatory project meaning we assume that the exploration is already done and some key facts have been found…You have already done a good job in finding some facts, just clearly specify a few findings and center your story around them."
- ✓ "Please see the first comment for key finding points, and then communicate them through use of visuals."
- ✓ "The captions of slide need to be more descriptive and precise. For example, in the first slide, "Survived passenger by gender" could be replaced by "More passenger died than female though there were fewer female passenger on board." The same key fact is then depicted in a chart to backup the textual claim.

In general, these comments meaning that the story needed to be more focused on the key findings with better designs and more descriptive and precise caption. One thing about the caption should be mentioned is "Survived passenger by gender" was in the initial file and in the final file I believe I had more descriptive and precise captions compared with the initial file. Regardless, they were not focused to answer the main question with this dataset.

The first thing I did was to re-define new questions, as I found out that I was not answering the best questions in my first submission. For instance, in the first submission I was only answering whether or not some demographic factors affecting the survival rate, but I did not have the followed-up question in mind as which one of these factor have a key role in survival rate. By adding this question in the title of the story, it made it much easier to focus on the key ones and keeping in mind that story is for an Explanatory Analysis and there shouldn't be any chart

with the purpose of Exploratory Analysis. For this reason, I removed the Embarked On chart, as I believe it was not a major factor and it was not as interesting as Age. To me it was an important result to say Age was not a major factor, so I kept that chart with a new design.

Also, the feedbacks about each slide helped me to pick better designs for the new slides. Furthermore, when designing new graphs I kept that in mind that what questions are being answered and tried to pick the best designs to convey the answers. Same thing with the captions, I tried to have the focus of the captions on what's being conveyed in the charts to answer the questions. Overall, these were great feedbacks and I feel I learned a lot more working on this project.

Resources

Udacity Data source

Kaggle Titanic Data