project-4-storytelling

January 20, 2018

0.1 Summary:

Here we explore a kaggle dataset taken from the sinking of the titanic. The dataset contains 800+ data points each recording information about a passenger, including whether or not the passenger survived. We attempt to answer the question of what is the greatest factor predicting passenger survival.

0.2 Design:

Exploratory Data Analysis I gathered the data from the public data source on kaggle, and examined it for quality and tidyness issues according to the udacity EDA curriculum. The main issue was that of several dozen datapoints that had a null value for their age variable. Since one of the main variables I wished to explore was age, I excluded those datapoints from the dataset.

Questions Asked Based on the available variables, I selected several questions to focus on and answer in this visualization:

- What are the age & gender distributions of the passengers?
- What was the survival rate per age group?
- What was the survival rate per gender?
- What was the survival rate per ticket class?
- What was the single greatest factor for survival?

Visualization Tableau was used as the tool to build the visualization. Plot selection was driven by the nature of the variables and the questions asked. A combination of histograms, and solid or stacked bar charts were used to communicate the findings. The plots were displayed in a logical sequence in a story. The color scheme used in the plots is mostly comprised of blue, pink, and gray, and uses hues that are easy on the eye. Chart junk is avoided.

0.3 Feedback:

The first draft [3] of the visualization was shared on the udacity forum [2] for feedback. A response was received and is paraphrased below. Based on some of the feedback in this responses, the second version [4] of the visualizaiton was developed. Specifically, I changed the plot type for the gender proportion from a bar plot to a pie chart.

Response 1:

> I have this suggestions: - 1 first diagram (gender): I see no clear comparison between genders. Isn't it better to put it in a pie chart? - 2 (Gender Proportion of Passengers): looks better in a pie chart. - 3 in second page bins could be narrower. They are very tick now. - 4 for pages 4 and 5, you can compare different clusters with a "side by side bar". It will be more informative. - 5 different pages of your story can be combined in one page. Hence, viewers can compare different graphs and factor simultaneously.

0.4 Resources:

The dataset was downloaded from one of the datasets suggested by udacidy – a public dataset on kaggle [1]. This data was cleaned and provided as part of this presentation in csv format.

- [1] https://www.kaggle.com/c/titanic/data (Data set)
- [2] http://bit.ly/2mWlQpw (Feedback)
- [3] http://tabsoft.co/2DmDpX1 (Tableau Public Story - Version 1)
- [4] http://tabsoft.co/2mVSuIO (Tableau Public Story - Version 2)