Data Visualization Project – Communicate Data Findings

This project uses the titanic data set to analyse the survival rates of passengers, and the influence of the different parameters on the survival rates. The data set has been obtained from https://github.com/bcko/Ud-DA-Tableau-Titanic.

Structure of the dataset:

There are 891 details of passengers in the titanic dataset; each having 12 features. The features are related to Passenger Id, Name, Class of travel, whether the passenger survived or not, sex, age, siblings/spouse, parents/childrem, cabin, embarked etc. Most of the features are numeric in nature, with a couple identified as string. The columns of Passenger ID, name, ticket and cabin were dropped as they would not be used in the analysis.

Features of interest in the dataset:

My main feature of interest to explore will be the relationship between the passengers who survived (or died) with the class of travel, age, fare and if they had siblings, spouse or children.

The key findings from the analysis were:

- More males died in Class 3, while more females survived in class 1. All the classes had more male deaths than survival, and class 1 and 2 had more females surviving.
- The multivariate analysis shows that category of class travel does have an impact on survival rates between males and females.
- More passengers who paid a low fare died, whereas those passengers who paid a higher fare had a higher chance of survival.
- Females were more likely to survive than males.