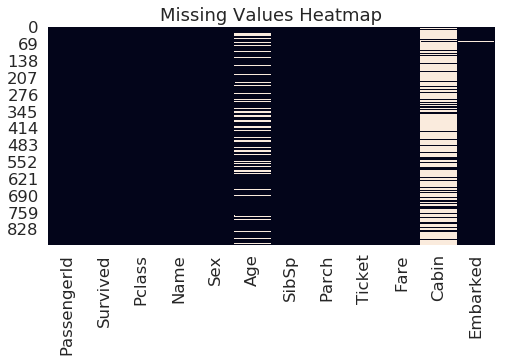
# Data Wrangling Methods used in Capstone

## Missing Value Treatment

I used heatmap to analyze the missing values in the data. There were only 3 features with missing values.



Since more than 77% of the cabin values and 19% of the age values were missing, removing those rows from the data was not an option. I had to come up with other means to fill the missing data.

For cabin, I filled the missing values with a new variable “U”. For age, I filled the missing values with mode of the age data since there were large outliers and mean would have inflated our age data.

## Outlier Value Treatment

Since there were outliers in age and Fare per head, they had to be dealt with. Since there were only few outliers, we could have deleted those rows entirely. But for representation purposes, we found the entries for which Z score was more than 3 and replaced them with values for which Z score was around 3.