**Exploratory Analysis on Titanic Dataset**

The objective of this report is to perform exploratory data analysis and to answer 3 hypotheses based on the Titanic dataset.

Exploration and Hypotheses

Titanic dataset was downloaded for the purpose of this analysis with 12 attributes (7 numerical & 5 categorical) and 891 occurrences. Missing values were found in 3 of these attributes: ‘Age’, ‘Cabin’ and ‘Embarked’ (177, 687 & 2 respectively). A statistical summary was created for numerical values, a histogram for normal deviation of ‘Age’ and boxplots for survival by age including gender and class variables to create the following hypotheses: survival rate is correlated with 1) passenger class, 2) gender and 3) age. For these visualizations, please refer to the jupyter notebook file.[[1]](#footnote-1)

Conclusion

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedThe associations between passengers’ class, age and gender to survival was validated through bar graphs showing each hypothesis. These were the conclusions:

1. Table

   Description automatically generatedThere is a statistically significant correlation between survival and the socio-economic level of the passengers determined by the cabin class.
2. Chart

   Description automatically generatedThere is a statistically significant correlation between survival and the gender of the passengers.
3. Table

   Description automatically generated with medium confidenceThere is a statistically significant correlation between survival and the socio-economic level of the passengers determined by the cabin class.

1. Git Hub, Paulina Vaca/Titanic; <https://github.com/PaulinaLangara/Titanic> [↑](#footnote-ref-1)