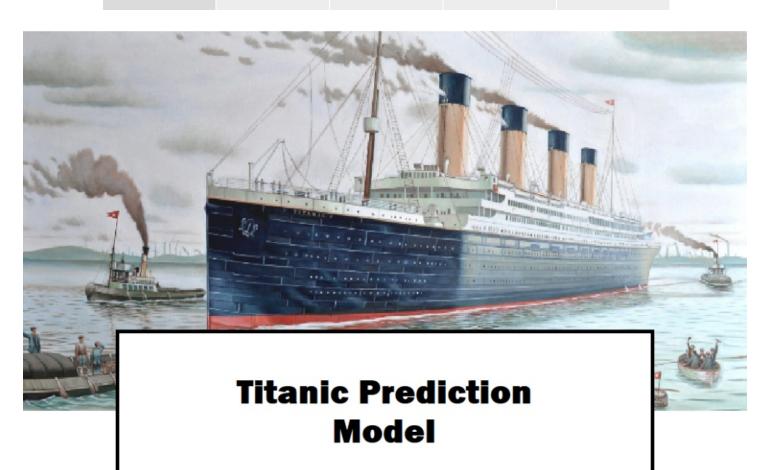
Overview
 Project Goal & Flow
 Heatmap Correlations
 Relative Correlation
 Classification Models



Overview Project Goal & Flow Heatmap Correlations Relative Correlation Classification Models

Project Goal: Find who is more likely to survive. **Project Flow** Create an EDA Compare & Clean and and model: Collect data: evaluate review data: Created a Heat Download results: Clean and Map to find Titanic Dataset Compare Extract correlations models to from Kaggle Relevant Data and build confirm results models Southampton 10 April 1912 Queenstown 11 April 1912 Cherbourg New York 41°43.5'N 49°56.8'W 15 April 1912

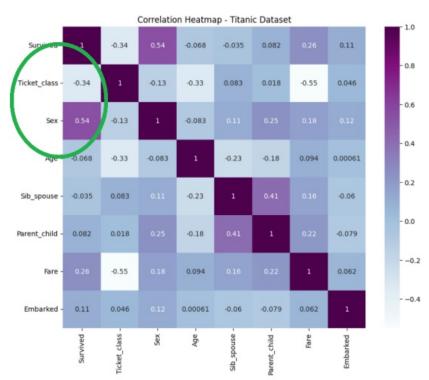
Overview	Project Goal & Flow	Heatmap Correlations	Relative Correlation	Classification Models

Heatmap Correlations

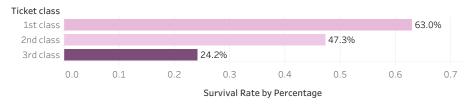
Survival Rate by Male and Female

Sex / Age Group female male 100 Age Group 90 20s 30s 40s 80 50s child old 70 60 Survived 50 40 30 39.7% 68.8% 88.9% 20 %9.02 10 20s 30s 40s 50s child old 20s 30s 40s 50s

Based on this heatmap we can see there is a high correlation in both the Ticket class and Sex features.



Survival Rate by Ticket Class

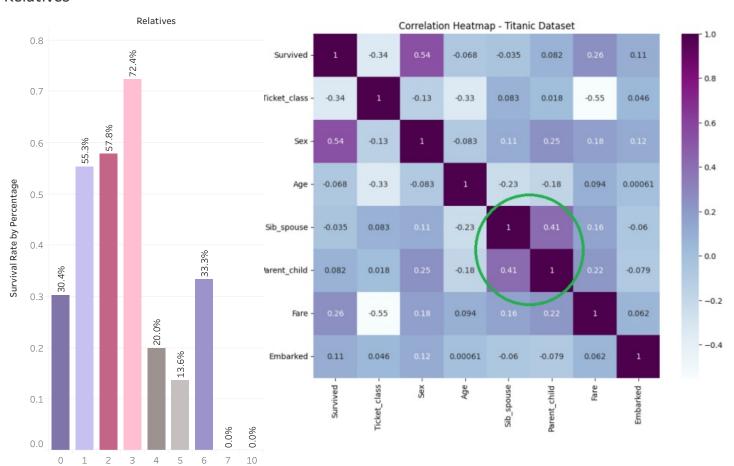


Overview	Project Goal & Flow	Heatmap Correlations	Relative Correlation	Classification Models

Relative Correlation

In addition to the correlation between Sex and Ticket Class, there was a relative correlation between Sibling-Spouse and Parent-Child.

Survival Rate by Number of Relatives



Overview	Project Goal & Flow	Heatmap Correlations	Relative Correlation	Classification Models

Classification Models

For this dataset, I tried to use a **Linear Model**, however, the outputs were skewed. I realized classification models would work better for this dataset.

I also used a **Logistical Regression Model** to confirm accuracy. I used Survived as my dependent variable and Ticket class, Sex, Age, Relatives, and Embarked as my independent variables.

However, I also used the **Random Forest model** on the Titanic dataset to predict survival based on the same variables.

