Predicting High-Risk Passengers for Maritime and Travel Insurance Pricing

Presented by Anny Carella Irumva

Agenda

1. Business Problem
2. Data Understanding
3. Data Preparation
4. Modeling
5. Conclusion

Business Understanding

- Maritime insurance covers risks related to sea travel, including the loss or damage of ships and cargo, as well as passenger-related risks such as injury or death during maritime disasters. Travel insurance, on the other hand, provides coverage for unexpected events like accidents, health issues, and natural disasters during trips, including risks like shipwrecks or extreme weather

- Risk Pricing and Claims Management Insurance companies must set premiums based on the likelihood of risks occurring, such as passenger mortality in maritime disasters. Accurate pricing is essential to balance claims and premiums, ensuring financial sustainability without driving away customers

Business Problem

- While maritime disasters are infrequent, they can result in substantial insurance payouts for injuries or fatalities. Insurers require a more effective method to evaluate passenger risk during these incidents, as conventional approaches relying on broad categories like age or ticket class often lead to inaccurate pricing and ineffective risk management

- By examining historical data, insurers can develop predictive models that predict the risk of every person. High-risk individuals would incur higher premiums, whereas low-risk individuals might benefit from discounts, enhancing customer loyalty and minimizing potential payouts.

Goal: Predict whether a passenger is likely to survive or not in a disaster scenario

Data Understanding

Dataset: Titanic Dataset with 891 entries and 69 features of various individual information

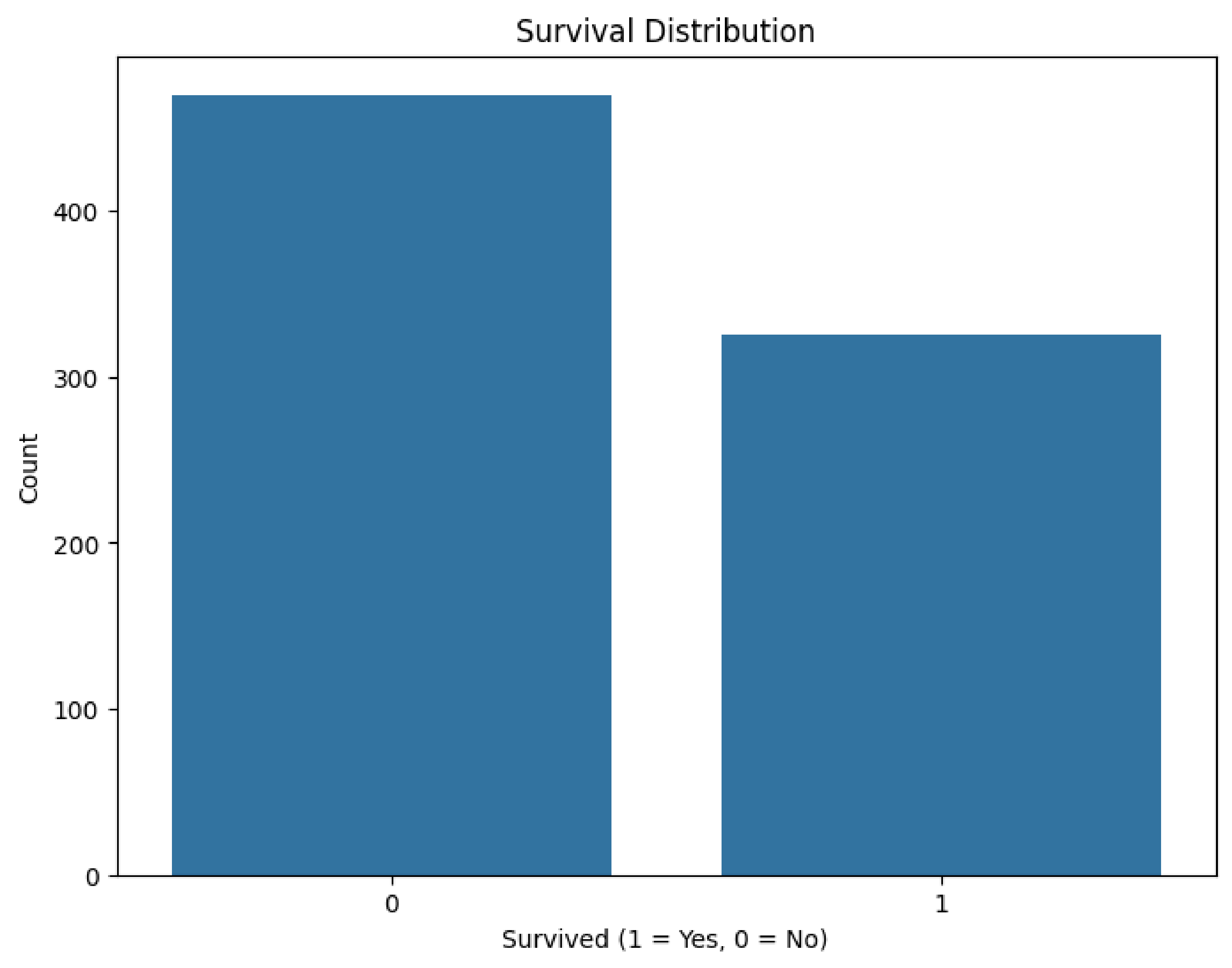
Key Features Used:

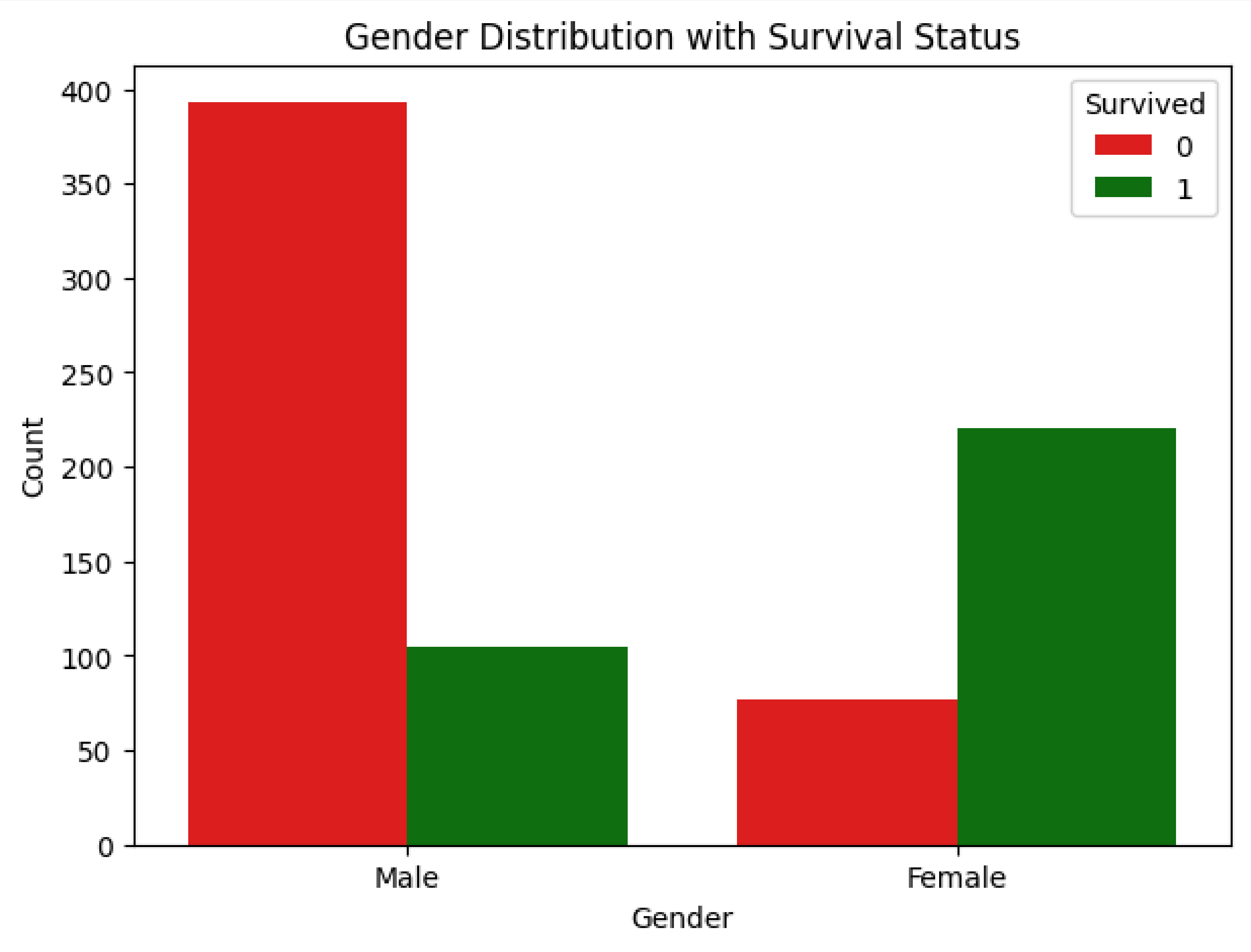
- Sex (1 = Male, 0 = Female)
- Age (Numeric)
- SibSp (Number of siblings or spouses aboard)
- Parch (Number of parents or children aboard)
- Fare (Ticket fare paid)
- Pclass (Ticket class: 1st, 2nd, or 3rd class)
- Family Size (Combination of SibSp and Parch)

Target Variable:

- Survived (1 = Survived, 0 = Did not survive)

Data Understanding

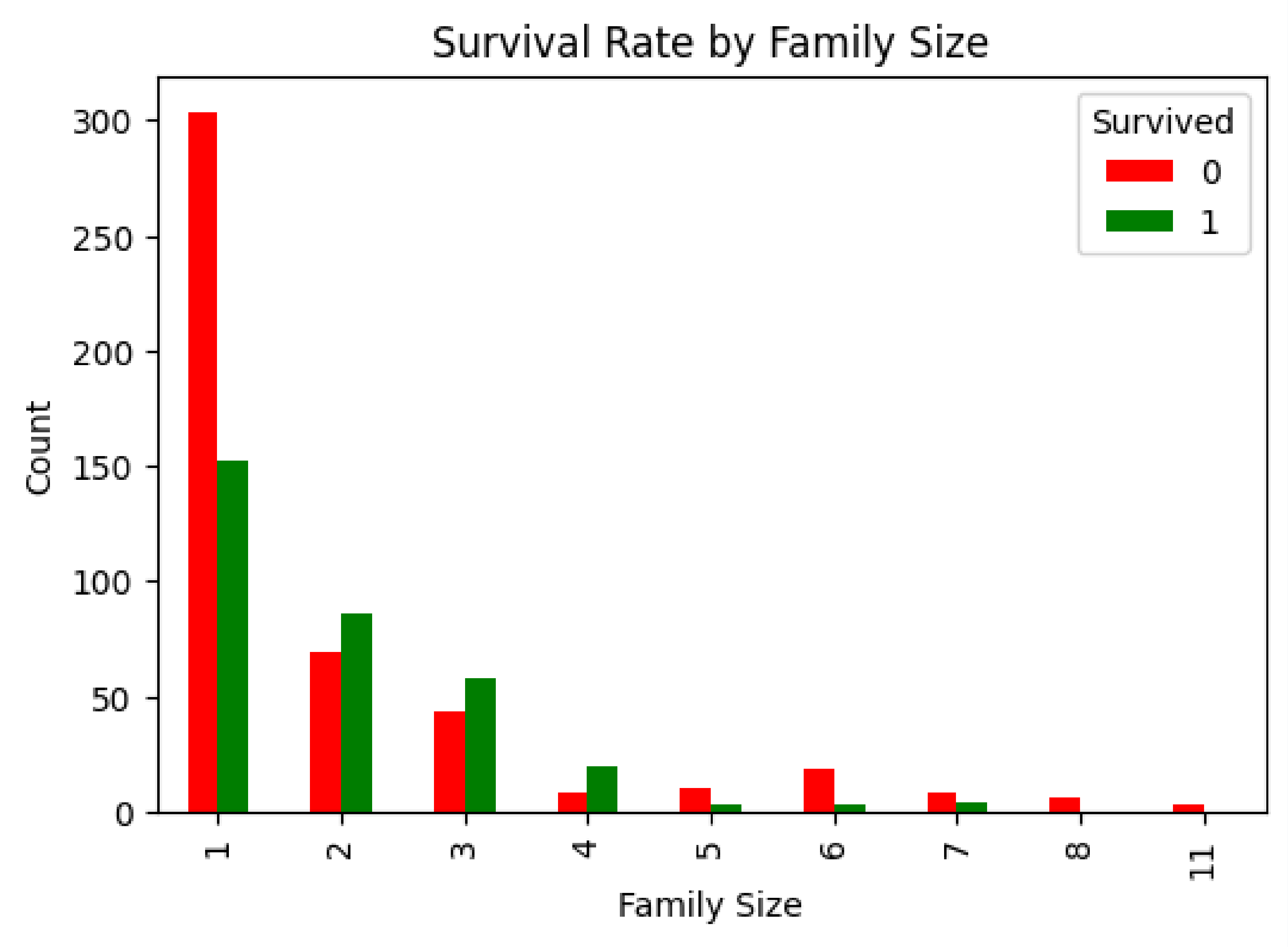




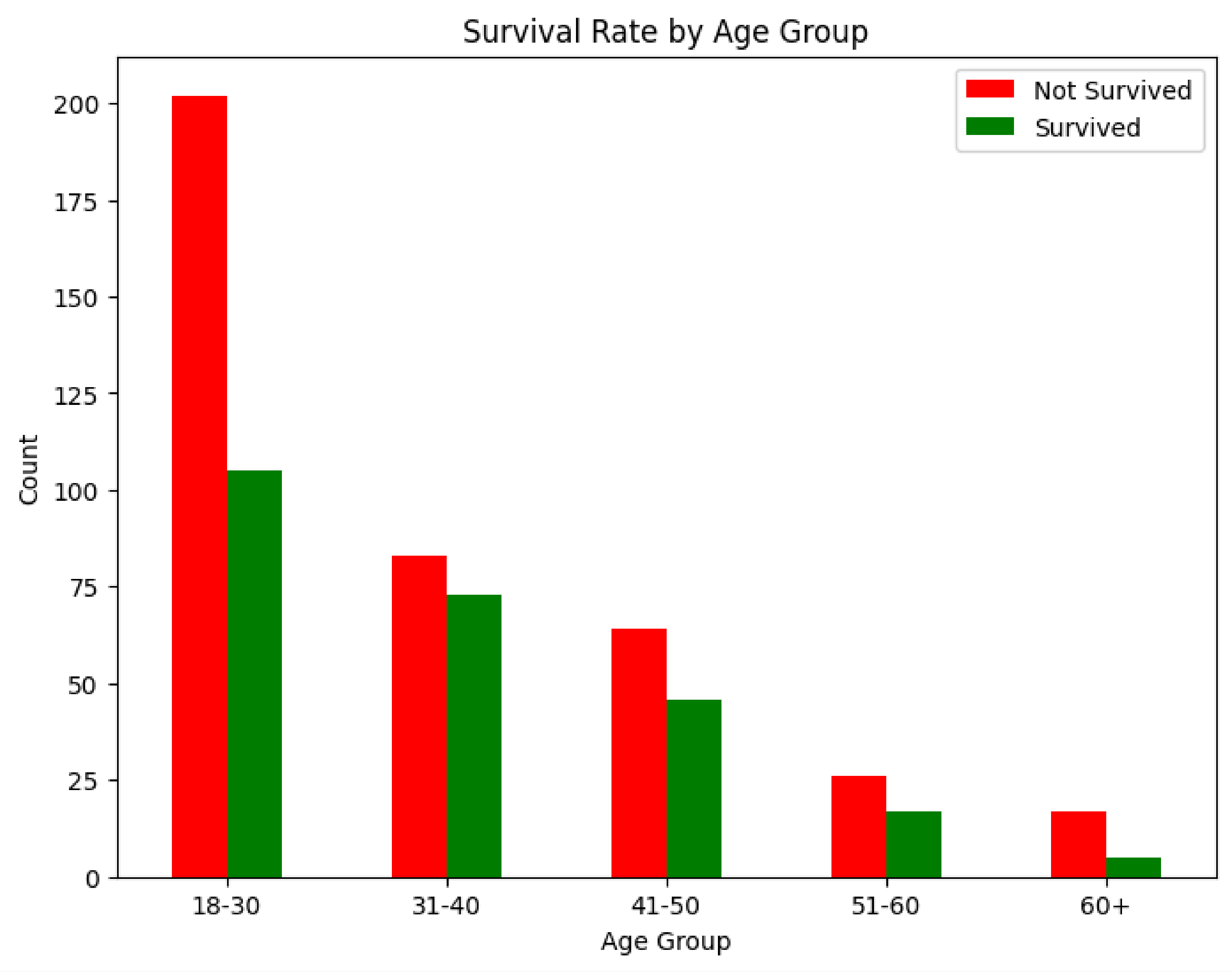
More passengers did not survive (class 0) than those who did survive (class 1)

A greater percentage of females survived in comparison to males

Data Understanding



Passengers traveling alone had the highest mortality rate, while those with family sizes of 2 or 3 had better survival chances. Larger families (size 4 and above) showed significantly lower survival rates



Younger passengers (especially those between 18-30) had the largest number of non-survivors, and while survival improves slightly with older age groups, survival rates drop again for passengers over 60

Data Preparation

01

Duplicates Treatment:

- Dropped 95 Duplicates rows

02

Feature Engineering:

- Combined P_Class with Sex columns
- Created IsMan Column
- Created FareClassRatio
- Add polynomial features (age^2, and fare^2)
- Dropped multicollinear columns
- Did a min max scale of some columns

03

Outliers Treatment:

- Dropped the outliers in FamilySize and Age columns

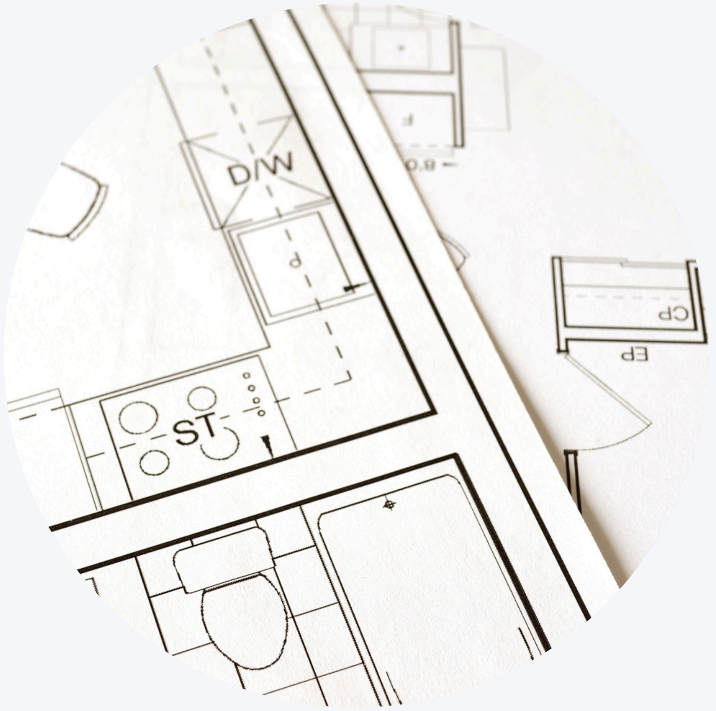
Modeling: Classification

Algorithm used:

- Binary Logistic Regression
- Two-Layer Feed Forward Perception
- Artificial Neural Net with Back Propagation

	Binary Logistic Regression	Two-Layer Feed Forward Perception	ANN with Back Propagation
Accuracy	84%	62.5%	83.8%

Value to Business



Risk-Based Pricing:

Insurance premiums are tailored to individual passenger risk, ensuring high-risk passengers pay more, while low-risk passengers enjoy lower premiums. This leads to fairer and more competitive pricing

Improved Risk Management:

The model helps insurers identify high-risk passengers, allowing them to better manage financial exposure by adjusting premiums or limiting coverage for high-risk individuals, thus reducing the risk of large payouts

Custom Insurance Plans:

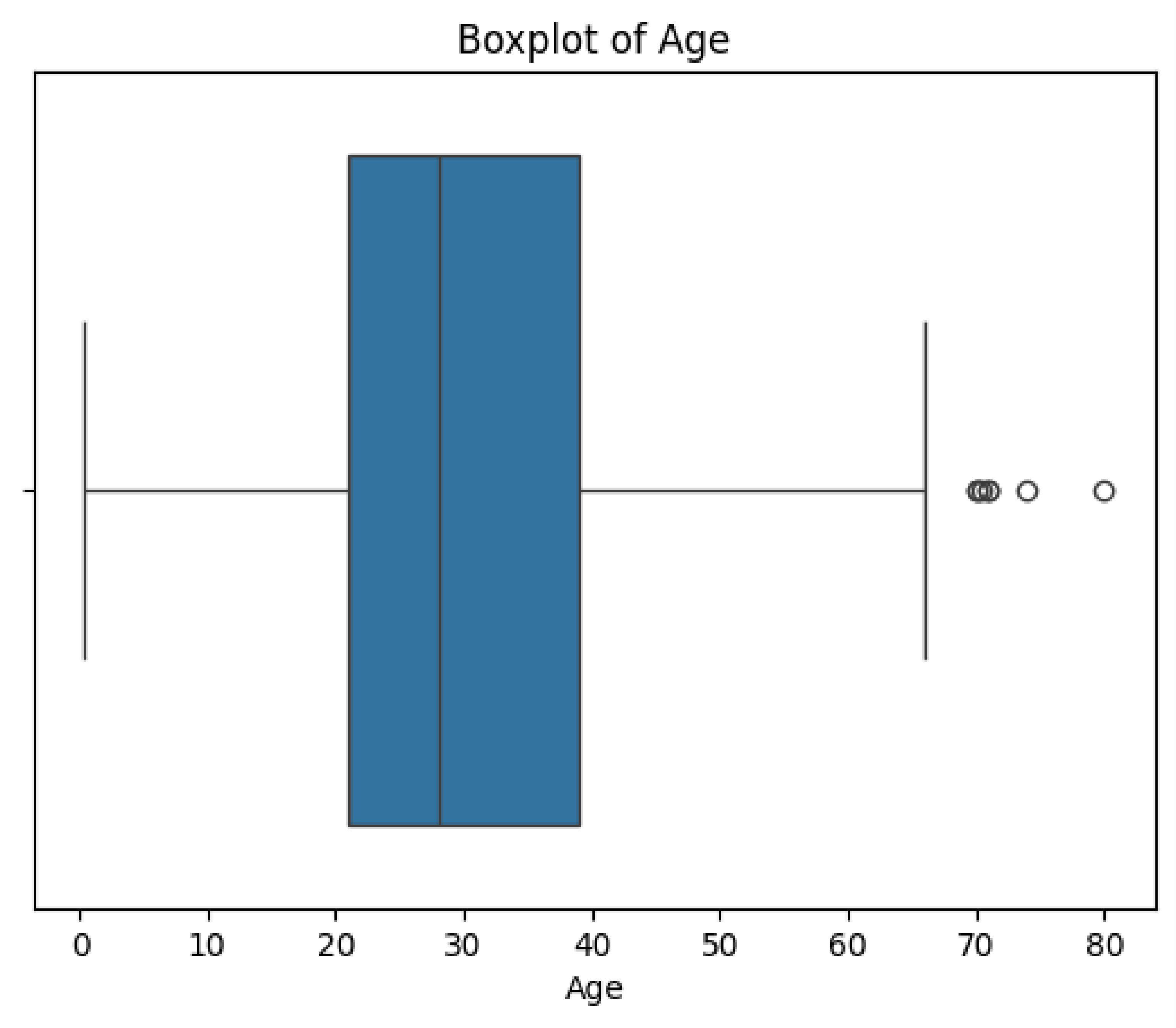
Insurers can offer specialized insurance packages for high-risk passengers, such as disaster recovery or medical evacuation coverage, meeting diverse customer needs while generating additional revenue

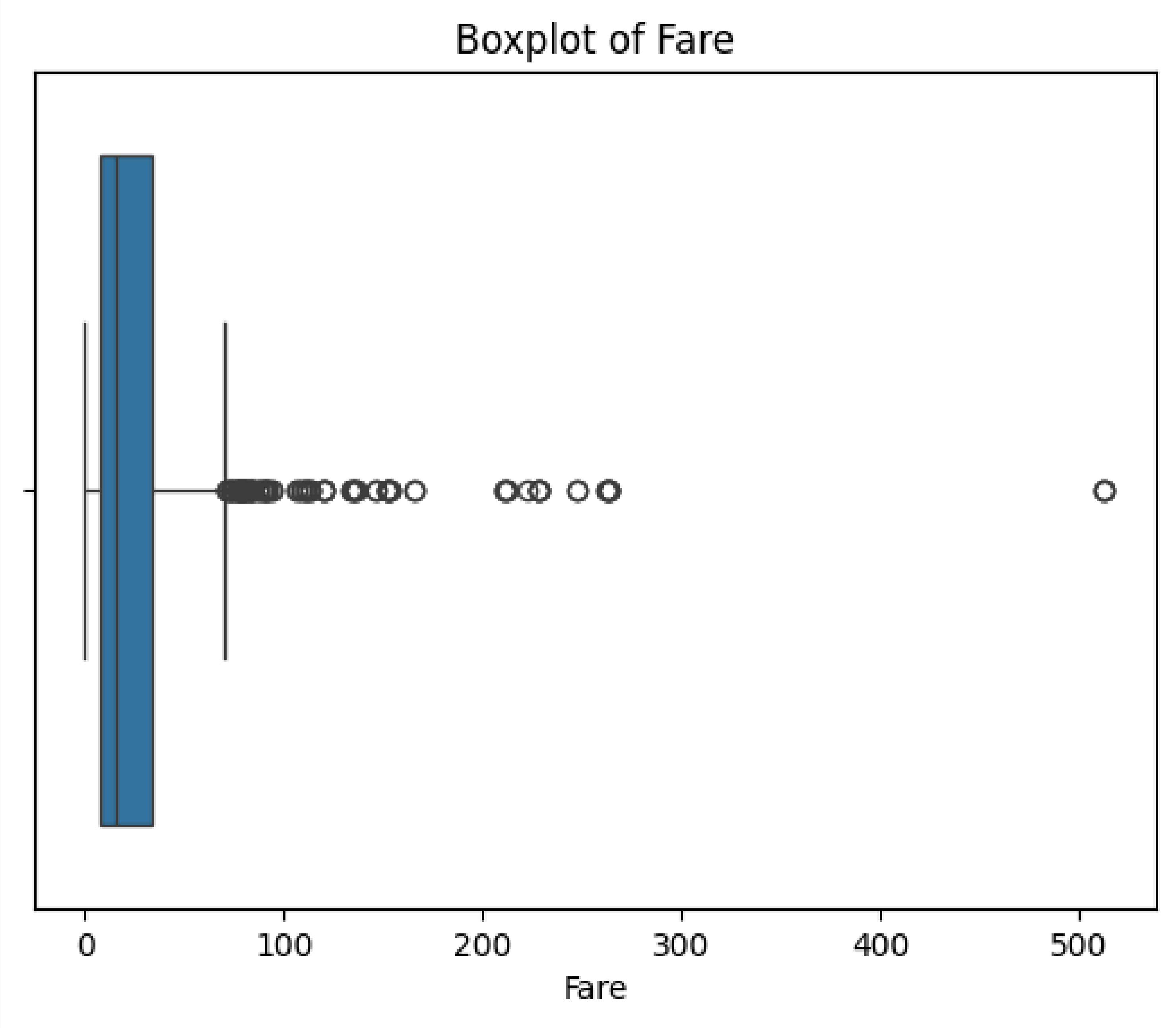
Conclusion

- The objective was to build a machine learning model that predict the likelihood of survival during a maritime disaster
- Utilized the Titanic data
- Tried 3 classification models
- **Binary Logistic Regression** achieved the highest accuracy and use low computing time, making it the best choice for this classification
- This solution is useful for maritime and travel insurance to identify high-risk passengers which will lead to a fairer and more competitive pricing

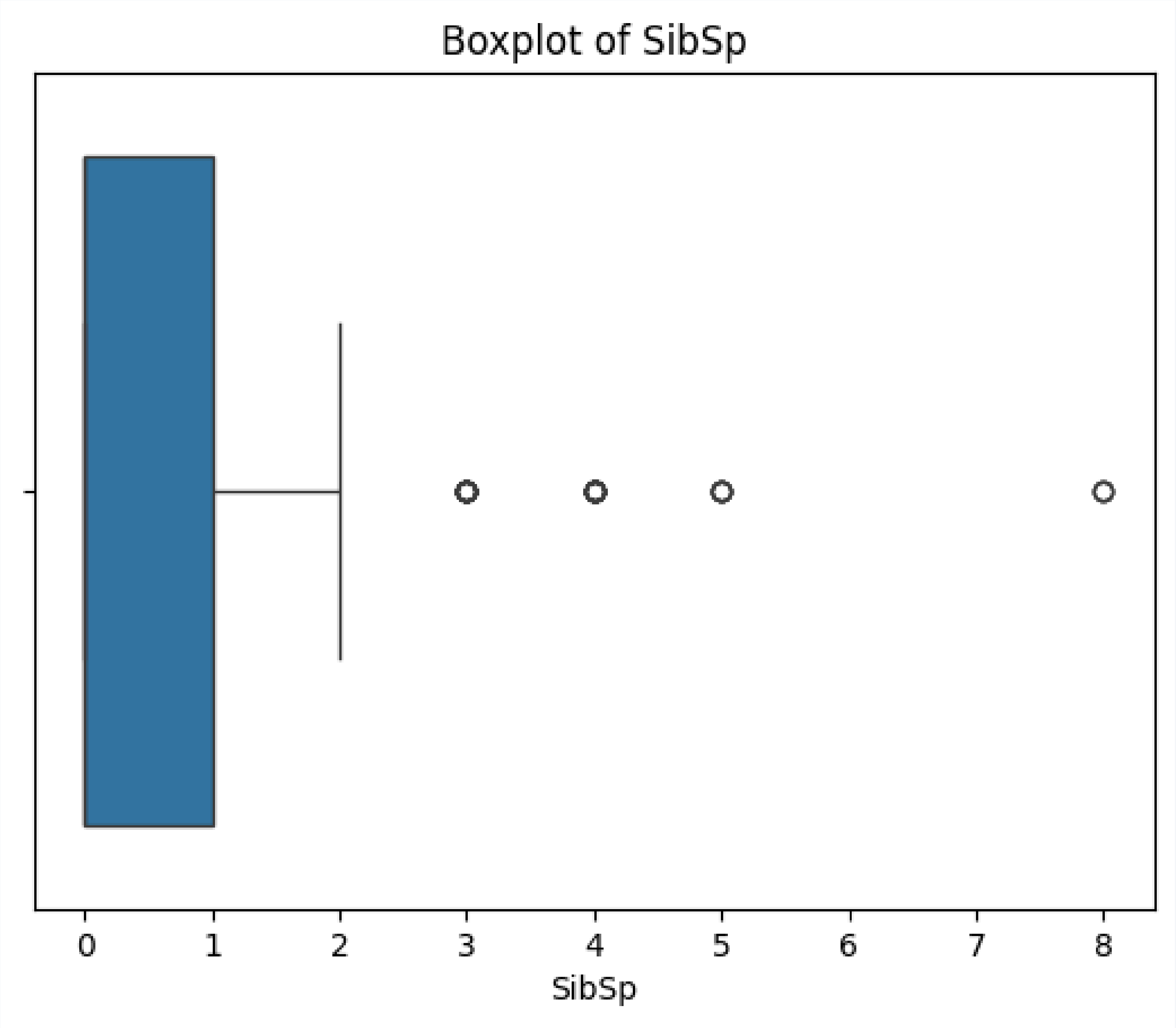
Thank You

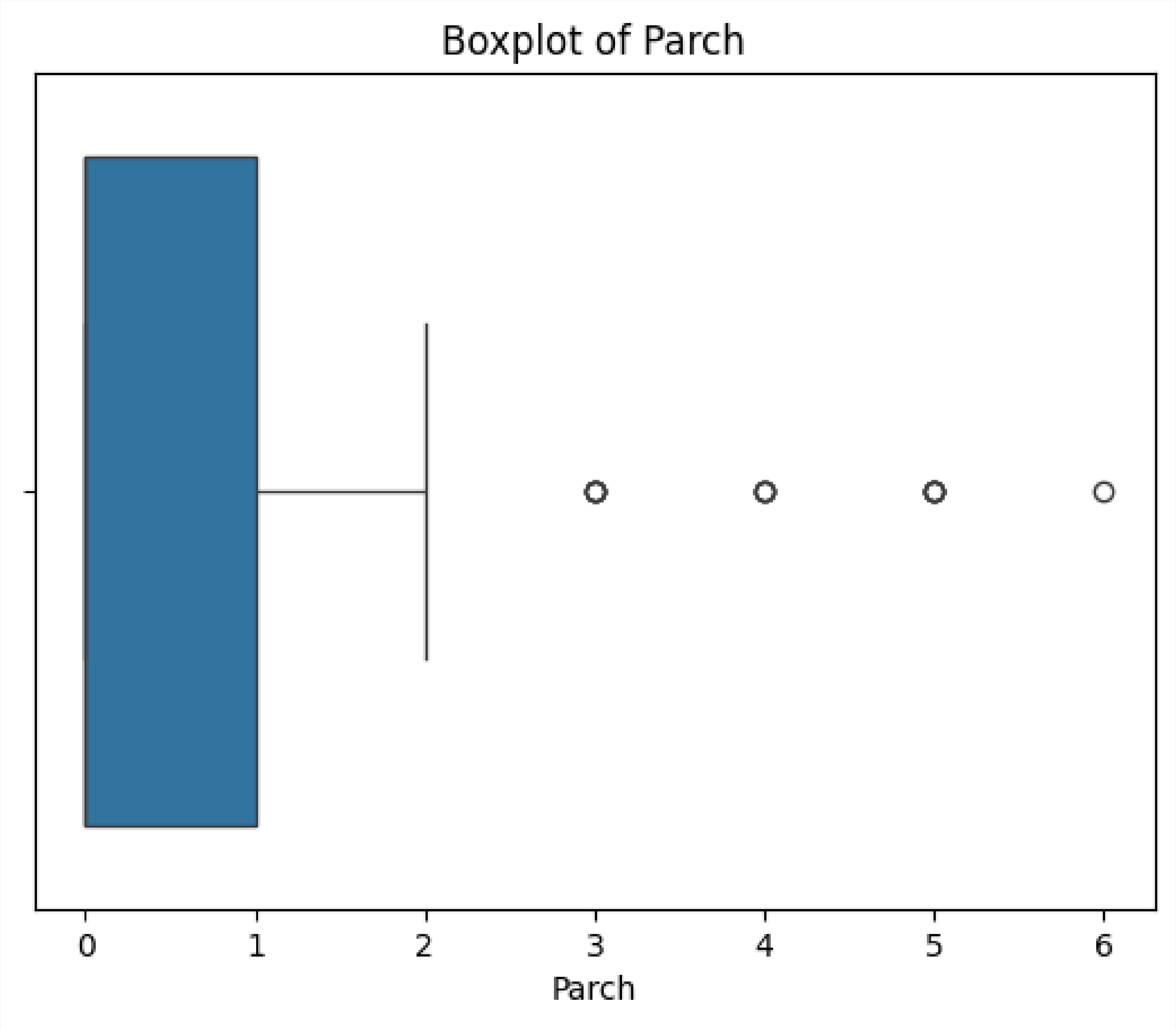
Appendix


Boxplot of Age
0
10
20
30
40
50
60
70
80
Age

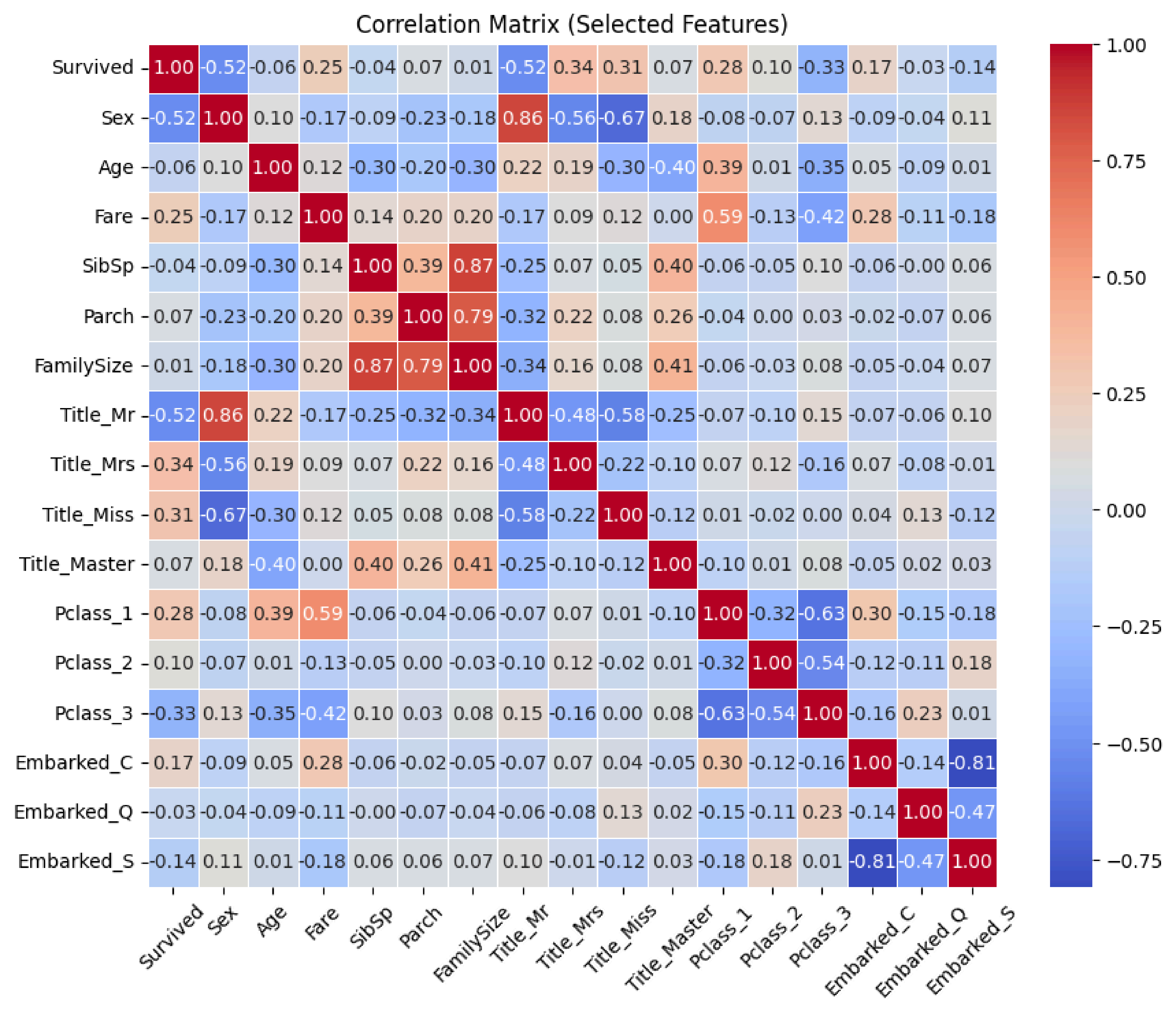

Boxplot of Fare
0
100
200
300
400
500
Fare

Appendix


Boxplot of SibSp
0
1
2
3
4
5
6
7
8
SibSp


Boxplot of Parch
0
1
2
3
4
5
6
Parch

Appendix


Correlation Matrix (Selected Features)
Survived 1.00 -0.52 -0.06 0.25 -0.04 0.07 0.01 -0.52 0.34 0.31 0.07 0.28 0.10 -0.33 0.17 -0.03 -0.14
Sex -0.52 1.00 0.10 -0.17 -0.09 -0.23 -0.18 0.86 -0.56 -0.67 0.18 -0.08 -0.07 0.13 -0.09 -0.04 0.11
Age -0.06 0.10 1.00 0.12 -0.30 -0.20 -0.30 0.22 0.19 -0.30 -0.40 0.39 0.01 -0.35 0.05 -0.09 0.01
Fare 0.25 -0.17 0.12 1.00 0.14 0.20 0.20 -0.17 0.09 0.12 0.00 0.59 -0.13 -0.42 0.28 -0.11 -0.18
SibSp -0.04 -0.09 -0.30 0.14 1.00 0.39 0.87 -0.25 0.07 0.05 0.40 -0.06 -0.05 0.10 -0.06 -0.00 0.06
Parch 0.07 -0.23 -0.20 0.20 0.39 1.00 0.79 -0.32 0.22 0.08 0.26 -0.04 0.00 0.03 -0.02 -0.07 0.06
FamilySize 0.01 -0.18 -0.30 0.20 0.87 0.79 1.00 -0.34 0.16 0.08 0.41 -0.06 -0.03 0.08 -0.05 -0.04 0.07
Title_Mr -0.52 0.86 0.22 -0.17 -0.25 -0.32 -0.34 1.00 -0.48 -0.58 -0.25 -0.07 -0.10 0.15 -0.07 -0.06 0.10
Title_Mrs 0.34 -0.56 0.19 0.09 0.07 0.22 0.16 -0.48 1.00 -0.22 -0.10 0.07 0.12 -0.16 0.07 -0.08 -0.01
Title_Miss 0.31 -0.67 -0.30 0.12 0.05 0.08 0.08 -0.58 -0.22 1.00 -0.12 0.01 -0.02 0.00 0.04 0.13 -0.12
Title_Master 0.07 0.18 -0.40 0.00 0.40 0.26 0.41 -0.25 -0.10 -0.12 1.00 -0.10 0.01 0.08 -0.05 0.02 0.03
Pclass_1 0.28 -0.08 0.39 0.59 -0.06 -0.04 -0.06 -0.07 0.07 0.01 -0.10 1.00 -0.32 -0.63 0.30 -0.15 -0.18
Pclass_2 0.10 -0.07 0.01 -0.13 -0.05 0.00 -0.03 -0.10 0.12 -0.02 0.01 -0.32 1.00 -0.54 -0.12 -0.11 0.18
Pclass_3 -0.33 0.13 -0.35 -0.42 0.10 0.03 0.08 0.15 -0.16 0.00 0.08 -0.63 -0.54 1.00 -0.16 0.23 0.01
Embarked_C 0.17 -0.09 0.05 0.28 -0.06 -0.02 -0.05 -0.07 0.07 0.04 -0.05 0.30 -0.12 -0.16 1.00 -0.14 -0.81
Embarked_Q -0.03 -0.04 -0.09 -0.11 -0.00 -0.07 -0.04 -0.06 -0.08 0.13 0.02 -0.15 -0.11 0.23 -0.14 1.00 -0.47
Embarked_S -0.14 0.11 0.01 -0.18 0.06 0.06 0.07 0.10 -0.01 -0.12 0.03 -0.18 0.18 0.01 -0.81 -0.47 1.00
Survived Sex Age Fare SibSp Parch FamilySize Title_Mr Title_Mrs Title_Miss Title_Master Pclass_1 Pclass_2 Pclass_3 Embarked_C Embarked_Q Embarked_S
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75