Predicting Titanic Survivors with Data Analytics

On April 15, 1912, the largest passenger liner ever made collided with an iceberg during her maiden voyage. When the Titanic sank it killed 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships. One of the reasons that the shipwreck resulted in such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others.

The TitanicData.csv file contains data for 2201 of the real Titanic passengers. Each row represents one person. The columns describe different attributes about the person including whether they survived (survived), their age (Age), their passenger-class (Class), their sex (Sex).

Use your data analytical skillz (erm.. skills) to mine association rules for survival/no survival using Apriori, with support 0.005, confidence 0.8, and minimum item length = 2, Please remove the redundant rules. Report all the rules sorted by lift

Write a report (e.g., ½ page) and explain the **factors that had the most impact on the survival of Titanic passengers**. Also, **please reflect on how the lift measure** helped / didn't help you interpret the data in terms of factors of passenger survival.

Turn ins:

1. Your calculations of lift and the report in a separate Word file.

2. The script (R or Python) that was used to load, correct, and impute the dataset.