

CASE STUDY 006

[R]

Titanic Passengers

Difficulty Level: 3 of 3

Disclaimer: SuperDataScience has no affiliation with the Titanic or Kaggle. The scenario is made up for educational purposes.

You are a Data Analyst working for White Star Line (the owners of RMS Titanic). The company is trying to better understand the passengers who were on board the ship when it sank. You have been supplied with a data set of passengers and head office would like to understand the following

1. How many passengers are included in the data?
2. Who bought the most expensive tickets?
3. How many men and how many women survived? (please provide a bar chart)
4. How many passengers are included in the data by class? (please provide a bar chart)
5. How many people had the title 'Mrs'?
6. What is the correlation between Age and Fare? (please provide a scatter plot)
7. What is the distribution of fares? (please provide a histogram)

Be sure to read the Hints page to get the code to access the data

Good luck! The business depends on you.

Difficulty note: this is a difficult assignment. Do not be surprised that there will be lots of nuances we have not covered off in the courses. But just like in the Real Life – there will be things training has not prepared you for and you will need to do research to find how to solve the problems at hand. If you get stuck, check the clues file.