

Introduction to the Titanic Dataset

The RMS Titanic was a British passenger liner that sank in the North Atlantic Ocean on April 15, 1912, after striking an iceberg during its maiden voyage from Southampton, England, to New York City. Of the approximately 2,200 passengers and crew on board, more than 1,500 lost their lives, making it one of the deadliest maritime disasters in history.

The Titanic dataset contains information about many of the passengers aboard the ship. For each passenger, variables such as survival status (Alive or Dead), passenger class (Crew, First, Second, Third), sex (Male, Female), age (Adult/Child) were recorded.

In this activity, you will use ChatGPT to create some tables and graphs. Some questions you might investigate include:

- Were women more likely to survive than men?
- Did survival rates differ among passenger classes?
- Was the relationship between passenger class and survival the same for men and women?
- What factors appear to be associated with a passenger's chance of survival?

As you analyze the data, remember that association does not necessarily imply causation. Your goal is to use statistical tools to identify patterns and relationships in the data and to determine whether those patterns are likely due to chance or reflect meaningful differences among groups.

Learning objectives: By the end of this activity, students will be able to:

- Use AI tools to generate statistical tables and graphs
- Revise prompts to improve AI-generated output
- Verify the accuracy of AI-generated results
- Interpret relationships among categorical variables
- Evaluate the strengths and limitations of AI in statistical analysis

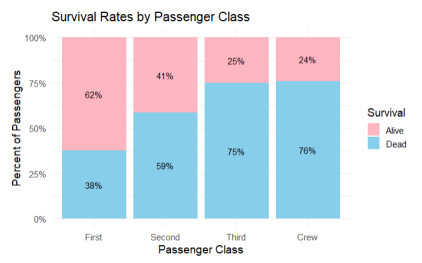
Part 1: Prompt Engineering

1. Write two different prompts to generate the table below. Which prompt produced the more accurate or useful output? What details in the prompt improved the result?

	First	Second	Third	Crew	Total
Alive	203 (28.6%)	118 (16.6%)	178 (25.0%)	212 (29.8%)	711 (100%)

Dead	122 (8.19%)	167 (11.2%)	528 (35.4%)	673 (45.2%)	1490 (100%)
Total	325 (14.8%)	285 (13.0%)	706 (32.1%)	885 (40.2%)	2201 (100%)

- Did you get the exact table above? If not, continue to prompt ChatGPT until the same table above appears in ChatGPT.
- Without asking ChatGPT, explain why the percentages in the table and graph below might differ. Then ask ChatGPT for an explanation and compare its explanation to your own. Comment on any interesting differences.



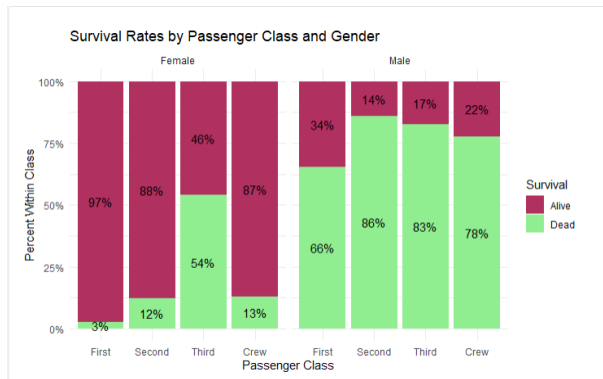
- Ask ChatGPT to create a graph where the percentages match the percentages in the table above (not the graph).

Part 2: Statistical Analysis

- Using the table or the graph answer the following questions.
 - What percent of first-class passengers died?
 - What percentage of first and second class passengers lived?
 - Of the passengers that lived, what percent were first class passengers? What percent were crew?
 - Which class had the highest survival rate?
 - Explain why it is important to look at the percentages rather than just the counts when determining the effects of passenger class on survival status.
- Write a few sentences describing the graph.
- Write a few sentences that summarizes the relationship between survival and ticket class.
- Based on the class-survival graphs and tables above, what conclusion would you draw about survival? What important variable might be missing from the analysis? Explain.

Part 3: Adding Gender

1. Now let's consider the gender variable. Ask ChatGPT to create a table and a graph that summarizes the survival rates of females and males. Report the percent of females that survived and the percent of males that survived.
2. Ask ChatGPT to create two separate tables similar to the table in part 1, one for females and one for males.
3. Do you notice any differences between the relationship between ticket class and survival between males and females? Explain.
4. What percent of First-Class females survived? What percent of first-class males survived?
5. Prompt ChatGPT to make the graph that you see below.



6. After bringing in the gender variable, did you overall conclusion about class and survival change after considering gender? Why might examining a third variable be important in statistical analysis?