Titanic data analysis

The Titanic was, at its time, the biggest passenger ship in the world and it sank on April 15, 1912, because it hit an iceberg.

Its story is very popular due to the gravity of the situation, which led to major changes in maritime safety regulations and to multiple forms of media talking about or inspired by the story, which is still relevant today.

One of the things that’s usually known about the disaster is that women and children were prioritized when evacuating the ship, which led to men being most of the deaths. Also, we can assume that those of higher class were also prioritized when evacuating.

So, what can the data tell us about this?

The data analyzed contains information of 156 passengers, so results may vary when compared to analysis of more complete data sets.

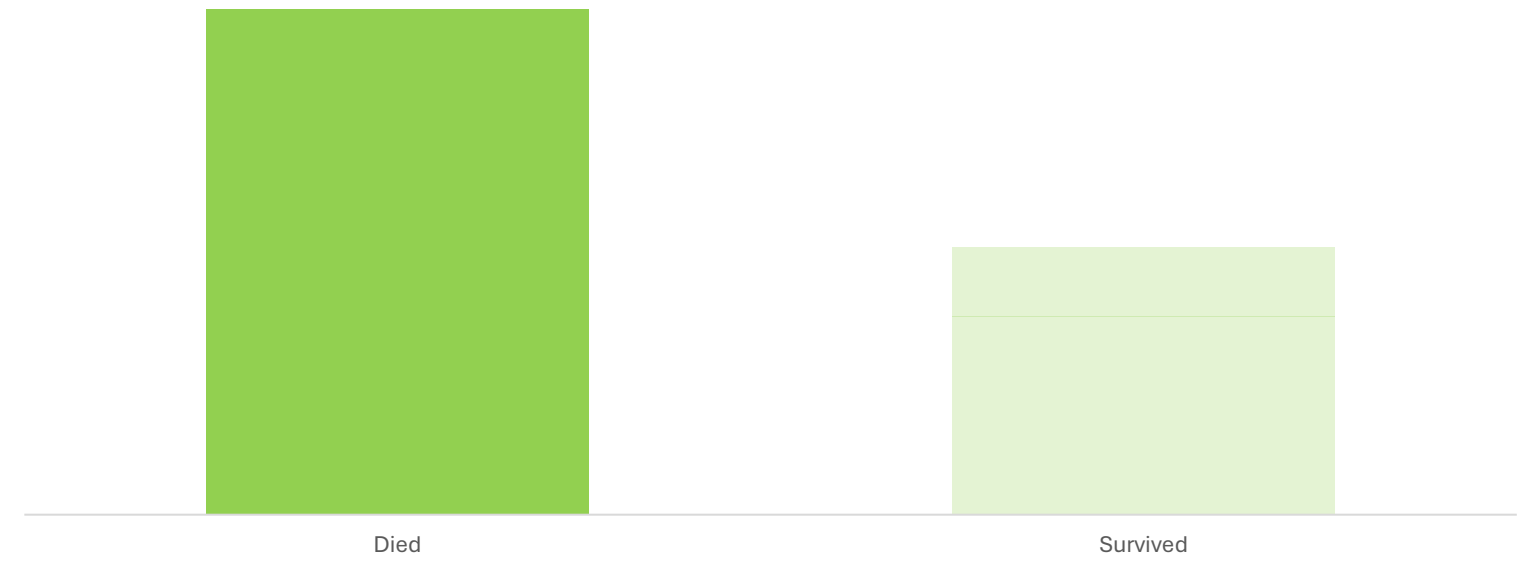
The analysis will be divided in:

- Gender.
- Age group.
- Class.
- Gender, age group and class.

Gender analysis.

Out of 156 passengers analyzed:

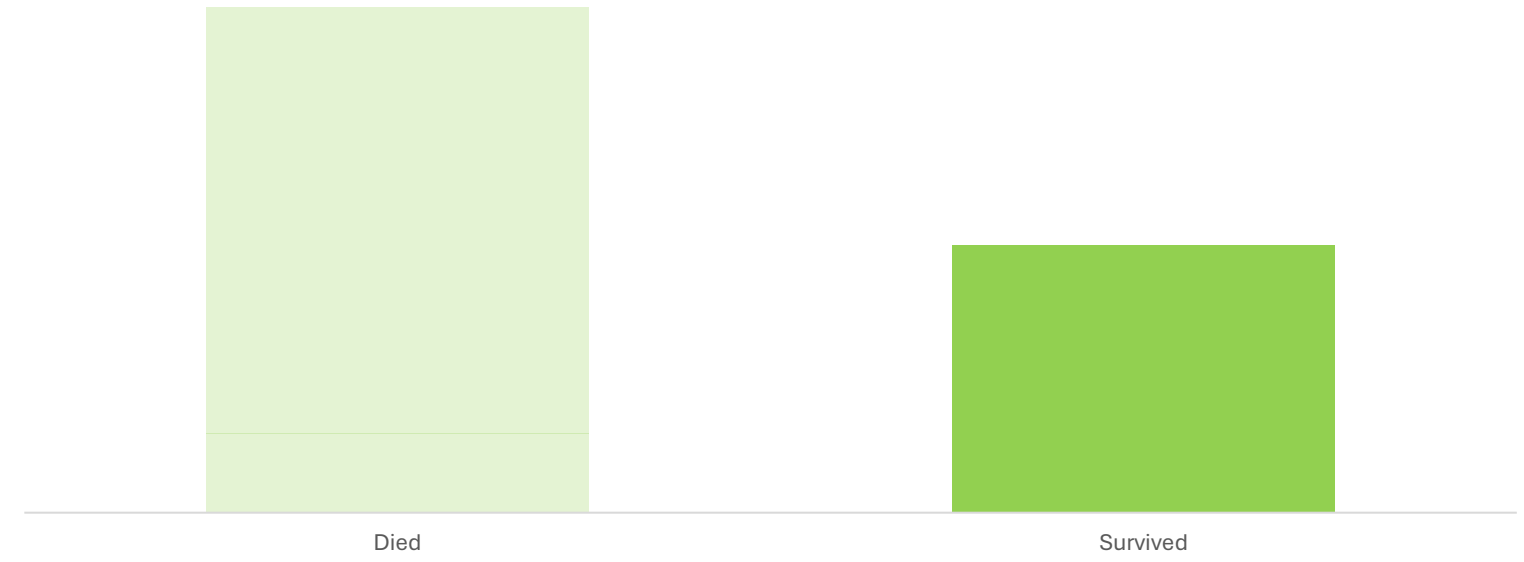
- 102 died.



Gender analysis.

Out of 156 passengers analyzed:

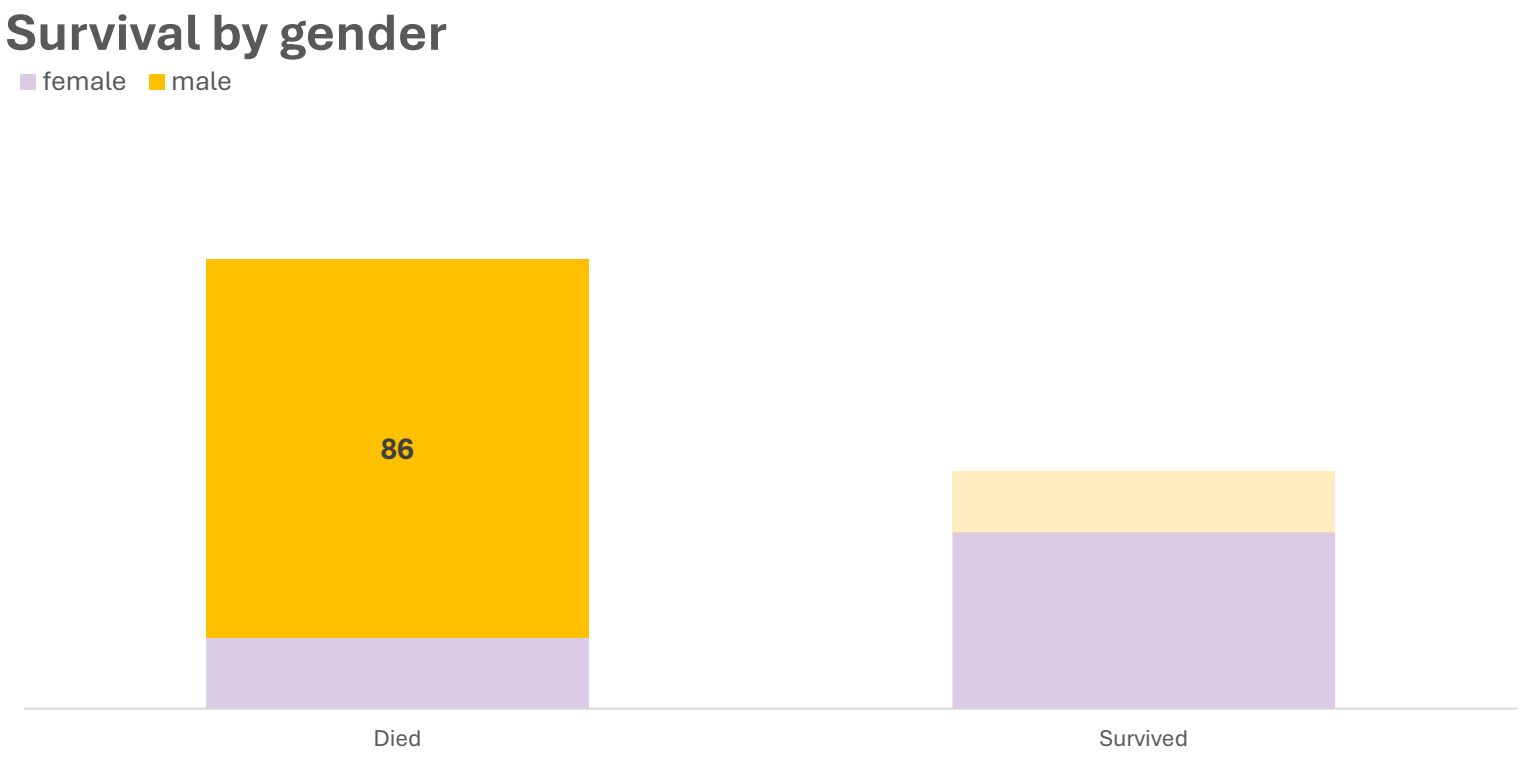
- 54 survived died.



Gender analysis.

Out of 156 passengers analyzed:

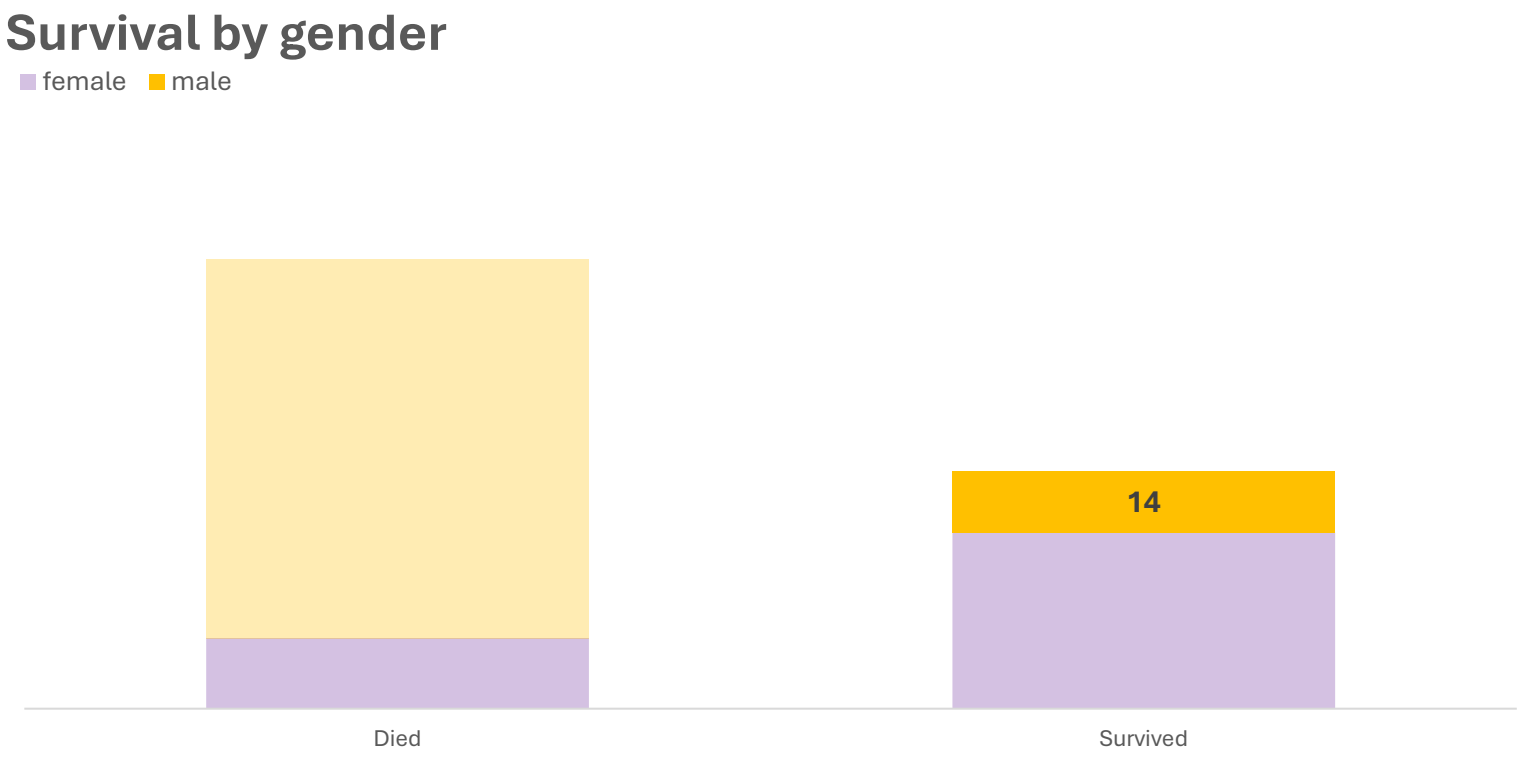
- There was 100 men, and 86 of them died.



Gender analysis.

Out of 156 passengers analyzed:

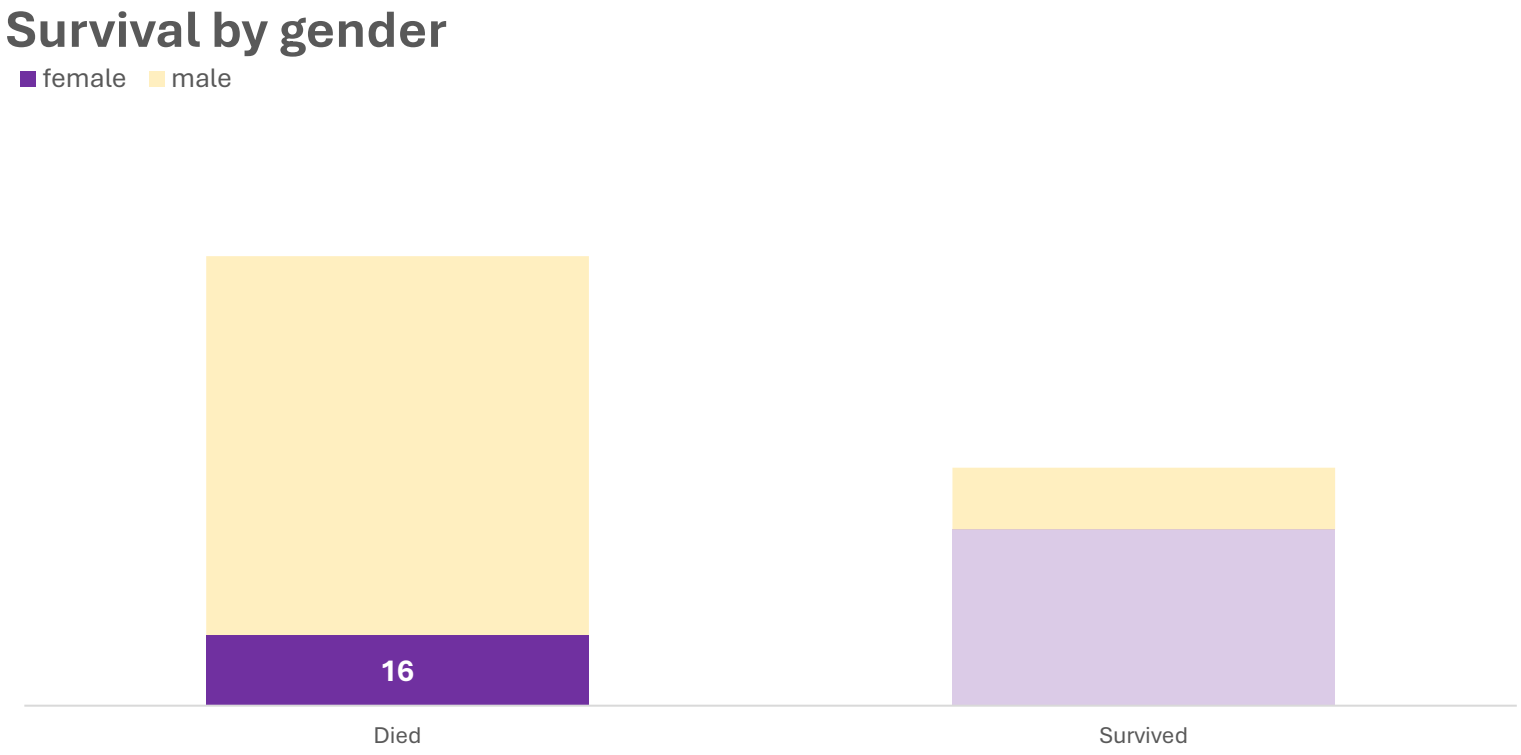
- Only 14 men survived.



Gender analysis.

Out of 156 passengers analyzed:

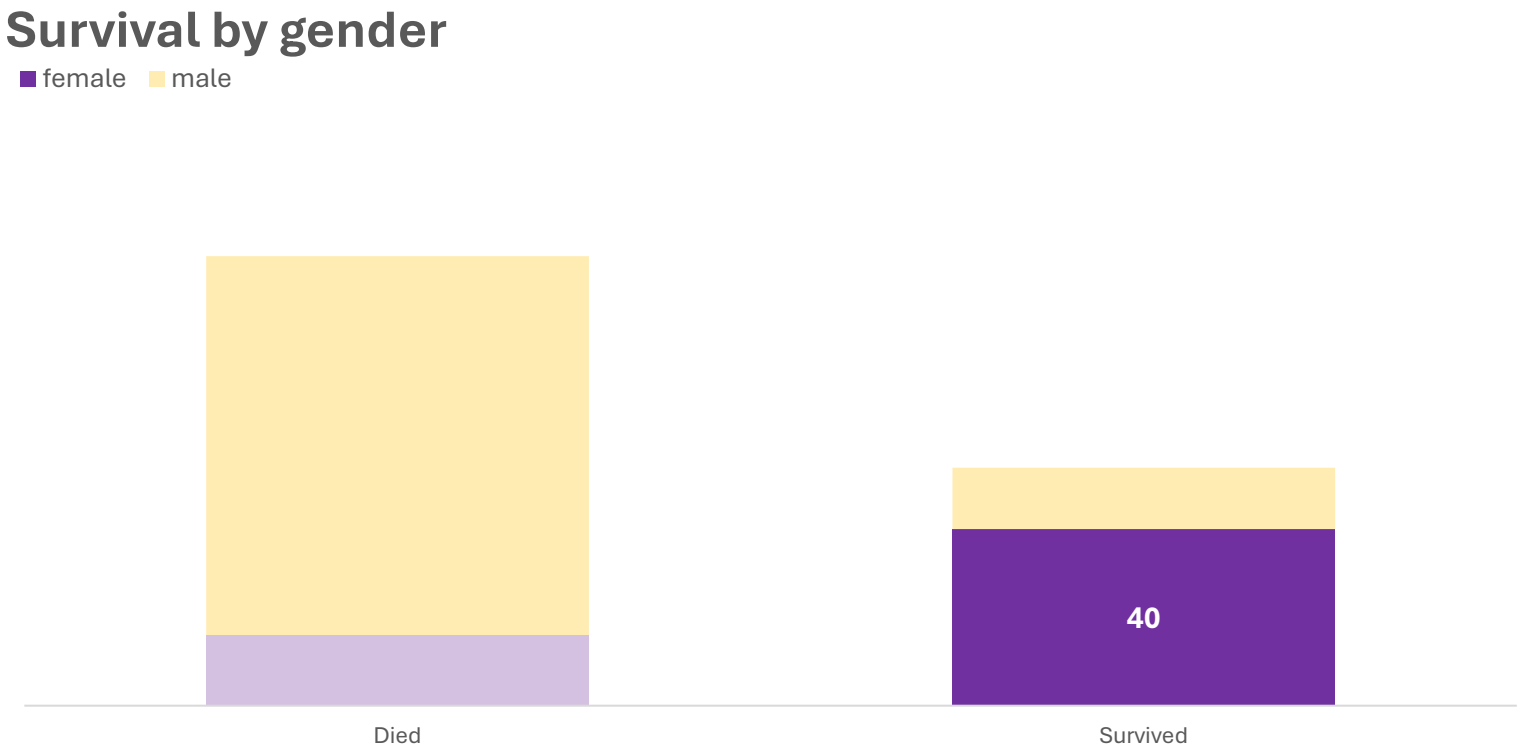
- 56 of them were women, and only 16 died.



Gender analysis.

Out of 156 passengers analyzed:

- 40 women survived.



Gender analysis

The survival rates for each gender were:

- Male: 14%.
- Female: 71.4%.

With this we can confirm that women were, in fact, prioritized while evacuating the ship.

Age group analysis

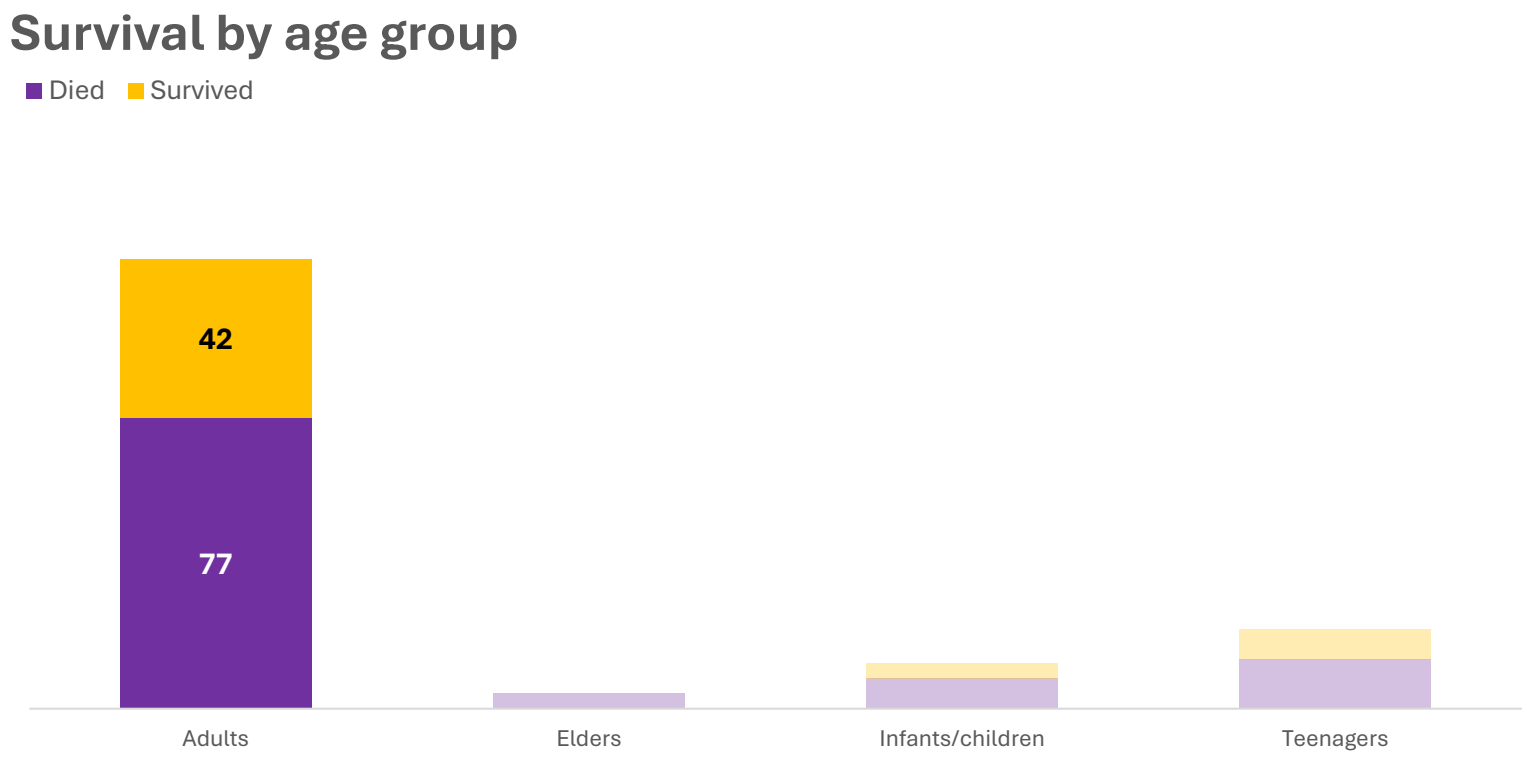
Age groups were created based on different age ranges. It's important to consider that the ranges used here may not align with the ones used around a century ago. The age groups were defined based on the ages spanning the stages of human development and are the following:

- Infants/children, for those under 12.
- Teenagers, for those 12 and over and under 20.
- Adults, for those 20 and over and under 65.
- Elders, for those 65 and over.

Age group analysis.

Out of 156 passengers analyzed:

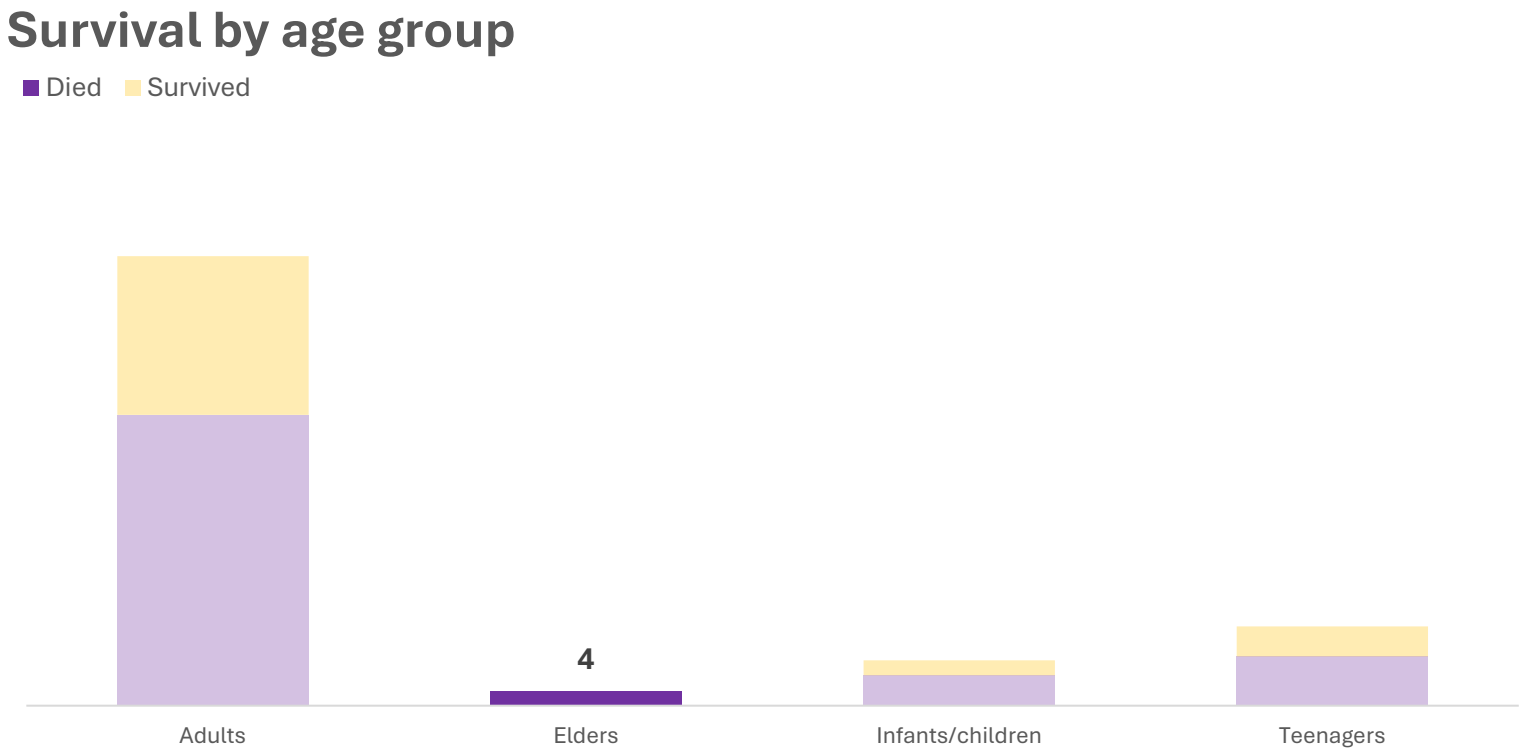
- 119 were adults, of which 42 survived and 77 died.



Age group analysis.

Out of 156 passengers analyzed:

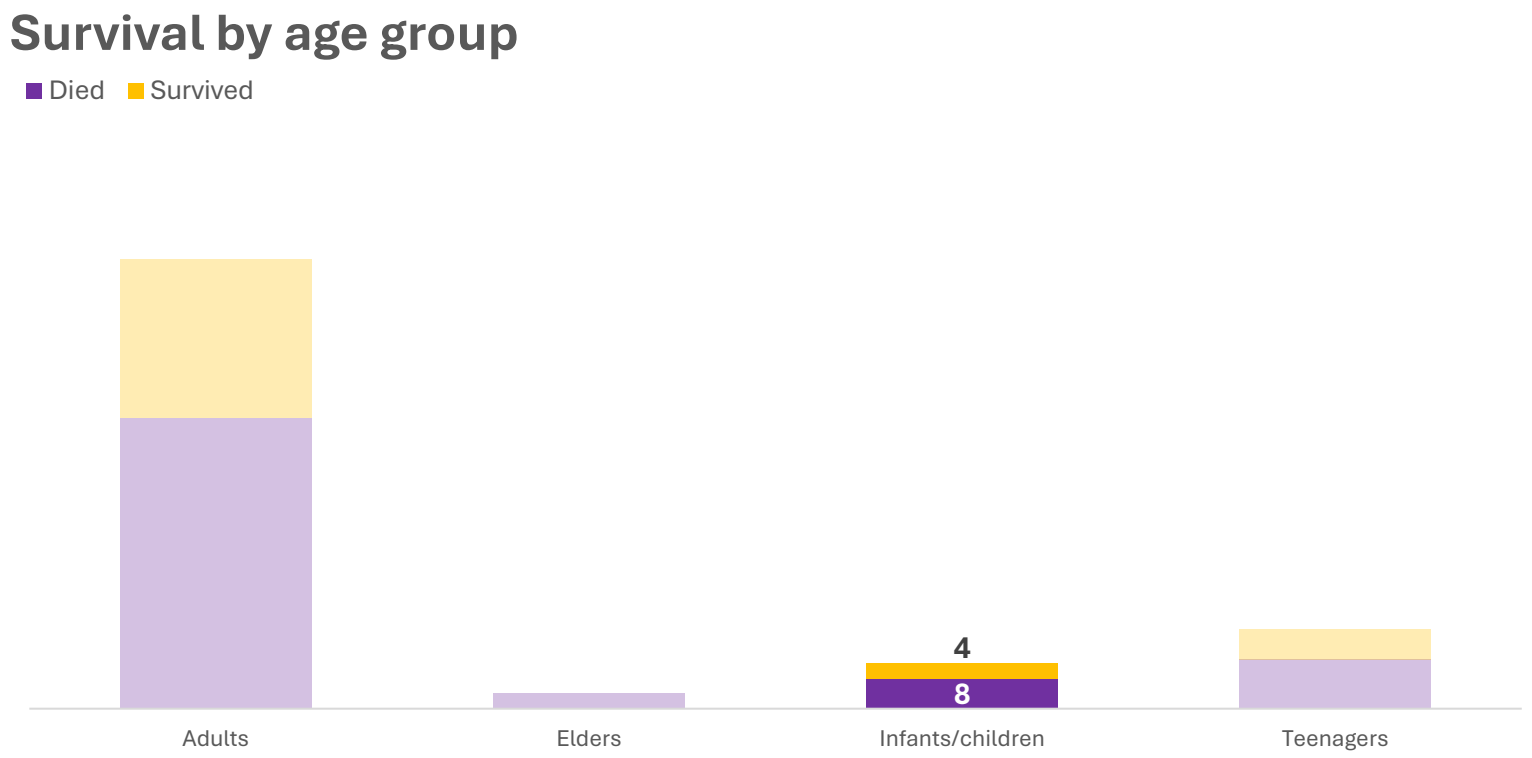
- All 4 elders died.



Age group analysis.

Out of 156 passengers analyzed:

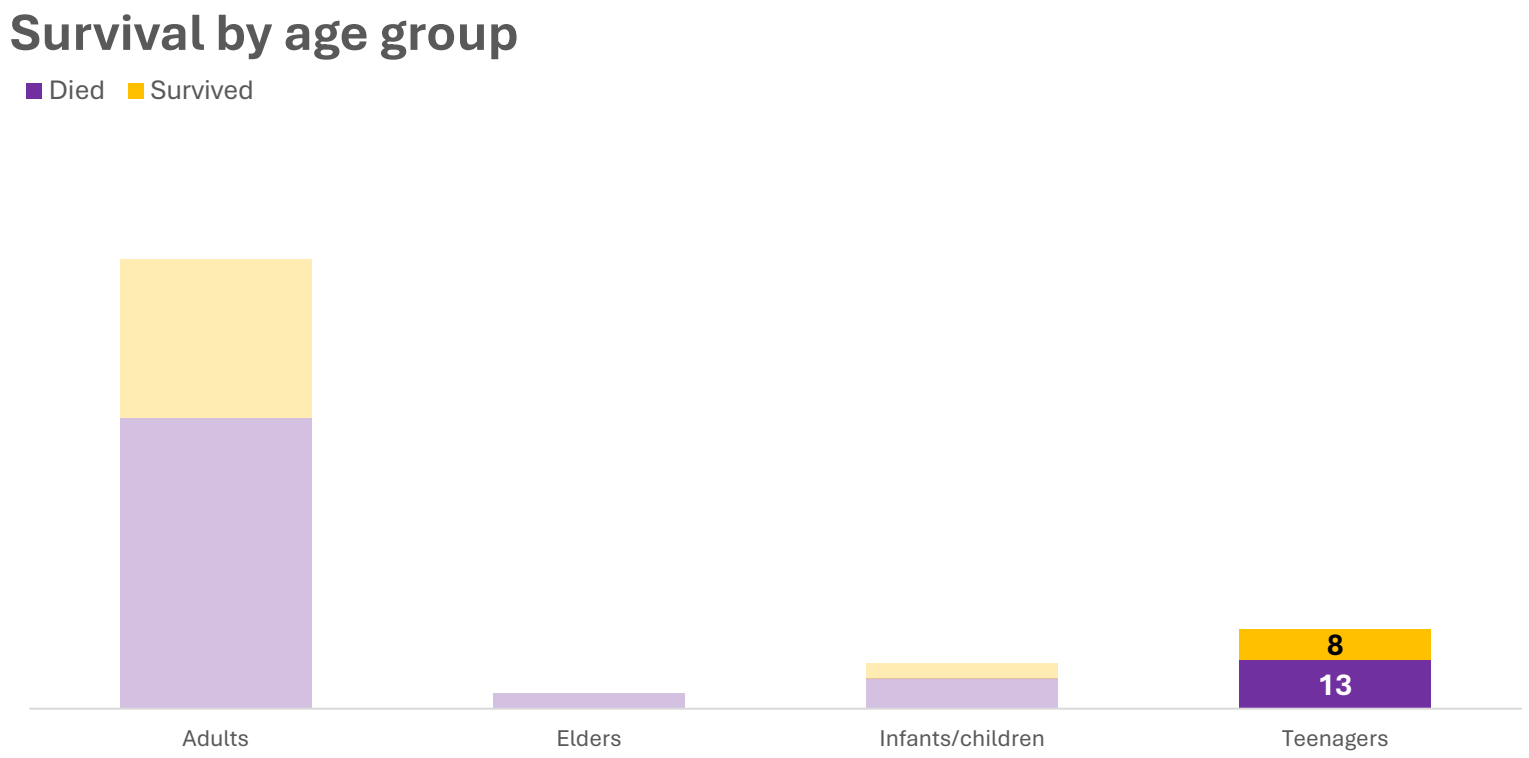
- 12 of them were infants/children, of which only 4 of them survived, while 8 died.



Age group analysis.

Out of 156 passengers analyzed:

- 21 were teenagers, out of which 13 died and 8 survived.



Age group analysis.

The survival rates for each group were:

- Infants/children: 33.3%
- Teenagers: 38.1%
- Adults: 35.3%
- Elders: 0%

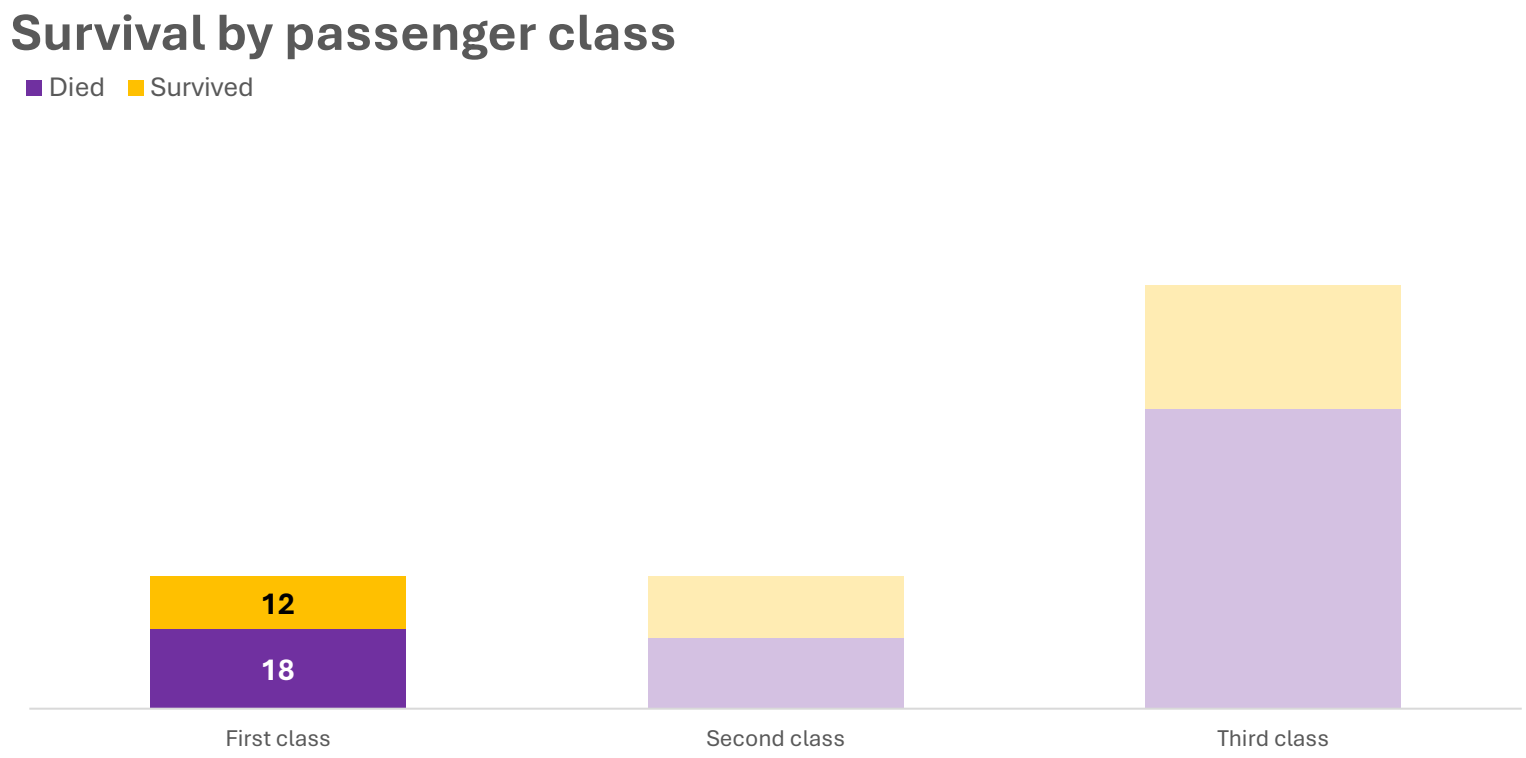
So, it doesn’t seem that children had a higher priority after all, as they had the second lowest survival rate. Perhaps the children on board got scared and started running around in panic when the incident happened, becoming hard for other people to find them and evacuate them.

Teenagers had the highest survival rate.

Passenger class analysis.

Out of 156 passengers analyzed:

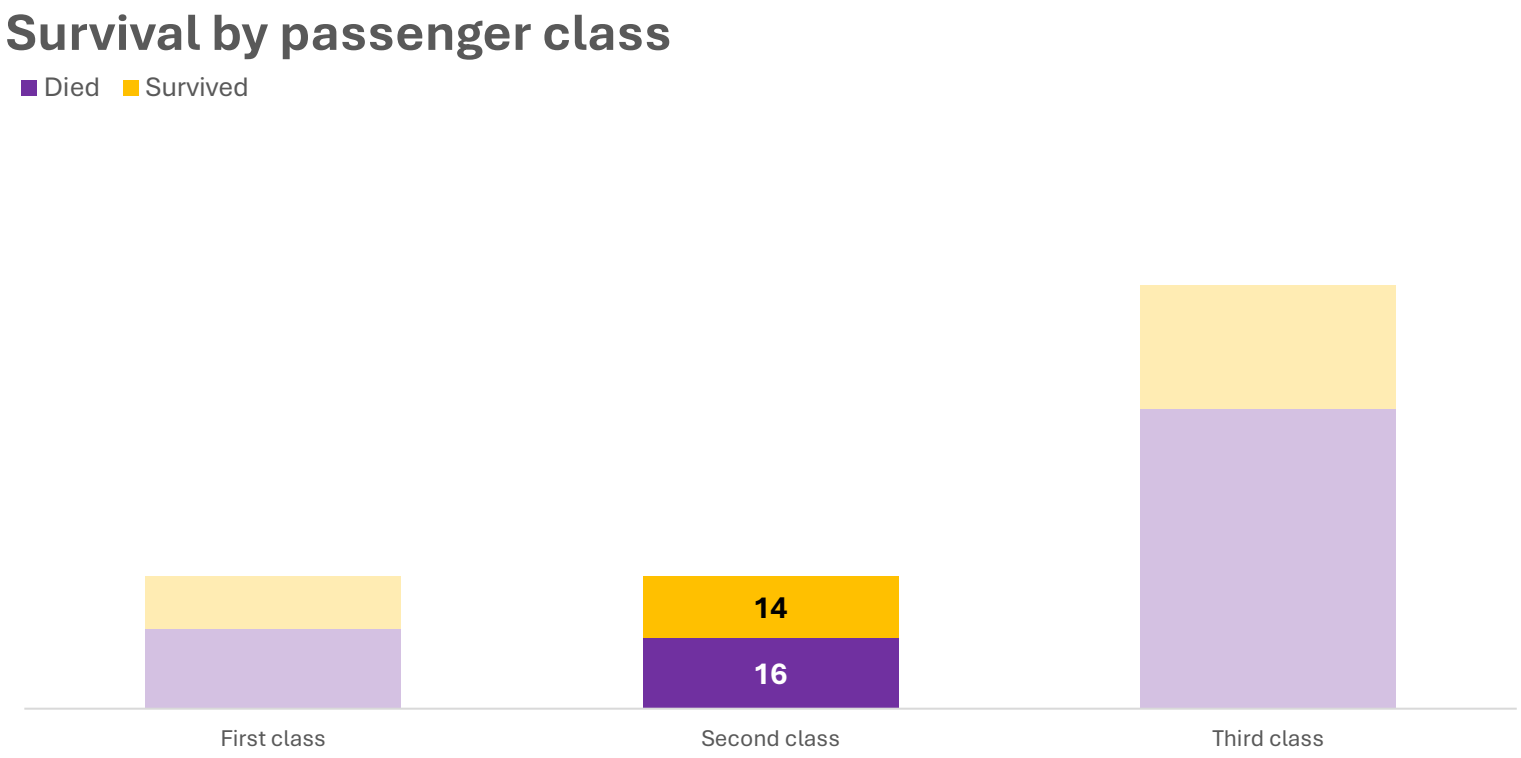
- 30 were first-class, out of which 12 survived and 18 died.



Passenger class analysis.

Out of 156 passengers analyzed:

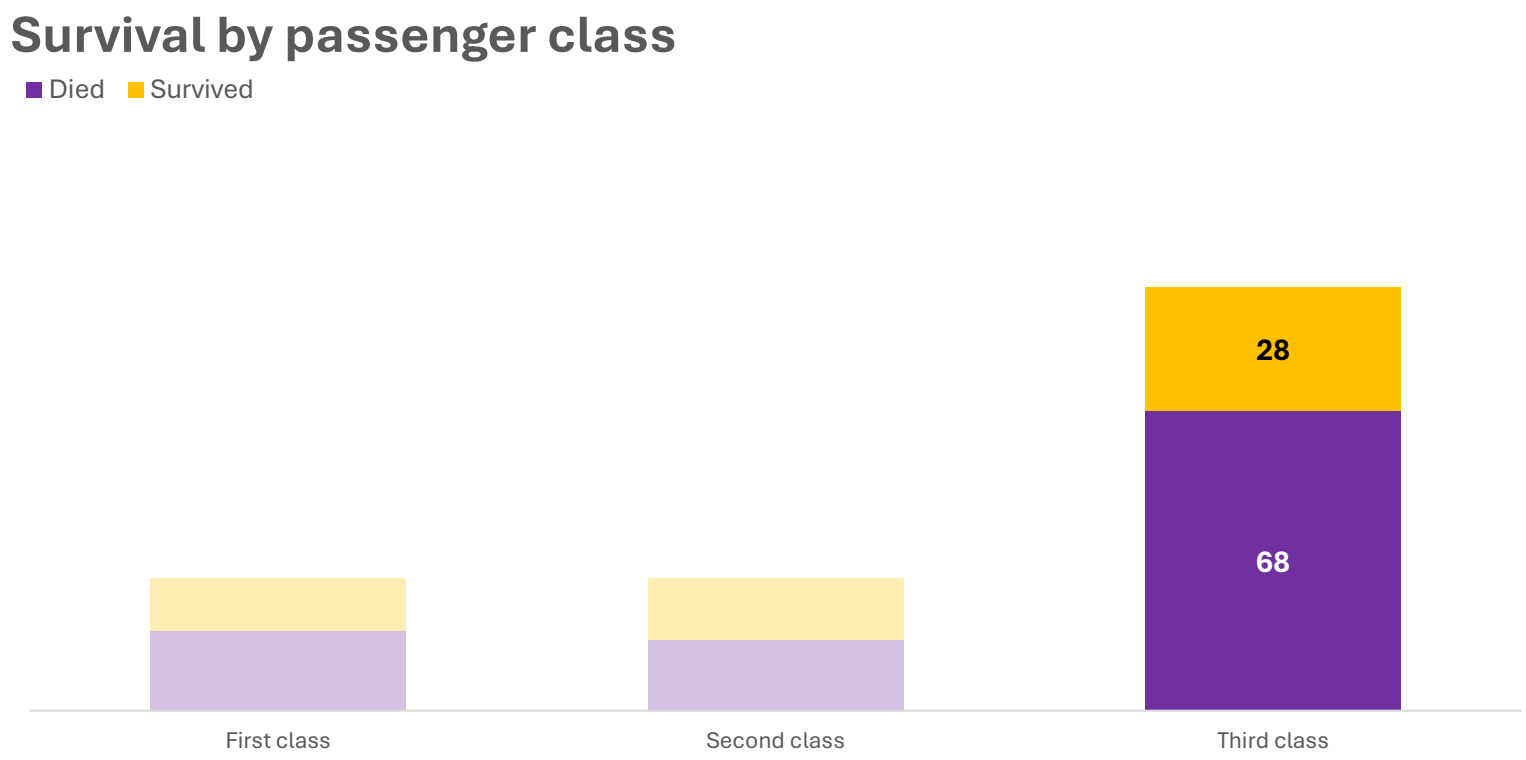
- 30 were second-class and only 14 of them survived.



Passenger class analysis.

Out of 156 passengers analyzed:

- 96 were third-class, out of which 28 survived and 68 died.



Passenger class analysis.

The survival rates for each segment were:

- First-class passengers: 40%
- Second-class passengers: 46.67%
- Third-class passengers: 29.17%

So, it doesn’t seem like the “higher class were also prioritized when evacuating” hypothesis is completely true, at least for this data set. It is true in the sense that both first- and second-class passengers had higher priority than third-class. However, one would expect the highest survival rate to belong to first class passengers, but it doesn’t, so the hypothesis would be false in that sense, as survival rate wouldn’t increase with class.

Gender, age and class analysis.

This analysis resulted in 22 different groups. Out of those 22 groups:

- 6 of them had no survivors.
- 5 of them had 100% survival rate (male, infants/children, second-class (1/1); female, adults, first-class (8/8); female, teenagers, first-class (1/1); female, infants/children, second-class (2/2); female, teenagers, second-class (2/2)). The male group mentioned previously was the only male group with a survival rate higher than 20%.
- The lowest survival rate for a group of women was 25%, which was the group of female, infants/children, third-class, in which only 1 out of 4 girls survived.
- The most numerous group (male, adults, third-class) had a survival rate of 14%, with only 7 survivors out of 50 people.
- The second-most numerous group (female, adults, third-class) had a survival rate of 71.43%, with 15 out of 21 women surviving.
- For male teenager groups, the only class that had a survivor was third class.
- For the remaining groups, a higher class indicated higher chances of survival.

Significant relationships.

The dataset is composed of different variables, and some of these variables are significantly related. This relationship means that a change of values in one will affect the values of the other. Here are some of the most significant relationships within the variables:

- People with more siblings or spouses on board tended to be of younger ages.
- People with more siblings or parents on board had also more spouses or siblings on board.
- Females tended to survive more than males.
- People of higher ages tended to belong to a higher class.
- Higher class people paid more for their tickets.

Conclusion.

After analyzing the data, it seems clear that women's survival was, in fact, prioritized, as women had the higher survival rate (of 71.4%). Surprisingly, though, children were the second-least likely to survive, so their survival wasn't as prioritized as previously hypothesized.

Also, when analyzed by class, second-class people had the highest survival rate, however, when analyzed by gender, age group and class, most of the small groups with the highest survival rates were from first class. So, people of higher-classes (first- and second-class) did have a higher survival rate, but the survival rate didn't increase with the class. This small discrepancy may be due to the size of the data set.

Some other things discovered through the data analysis is that younger people tended to travel accompanied by their families, and that older people tended to be of higher classes.