

# Exploratory Data Analysis on Titanic Passenger Data

University of California, Berkeley | School of Information

DATASC W200 Introduction to Data Science Programming | WBL 004 | Neha Singh | Spring 2022

Project 2 Exploratory Data Analysis

Team #2 | Kelianne Heinz & Ivy Chan

GitHub repository | [Project2 Chan Heinz](#)

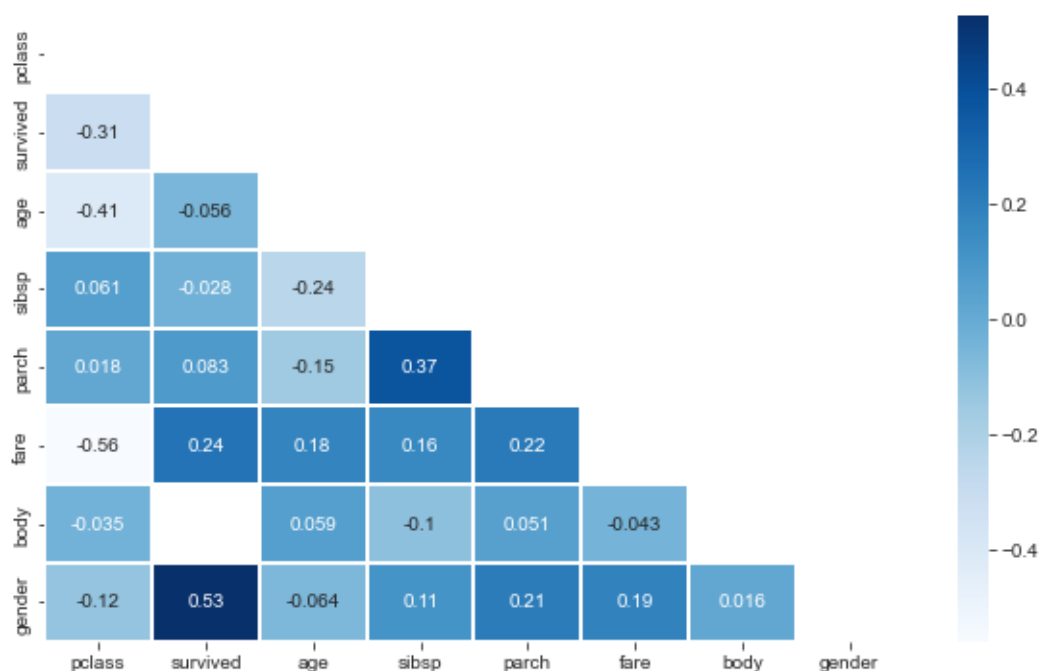
April 12, 2022

## Introduction

The Royal Mail Ship (RMS) Titanic, a luxury ocean liner, was considered unsinkable due to a series of compartment doors that could be closed to maintain buoyancy. However, during its maiden voyage on April 15, 1912, it sank after colliding with an iceberg in the North Atlantic Ocean, killing 1502 out of 2224 passengers and crew. This infamous tragedy inspired many stories, films, and musicals and gave rise to numerous myths. (History.com Editors, 2021; Piech, Chris, n.d.; Rajput, 2020; Tikkanen, 2022). This project will perform exploratory data analysis (EDA) on the Titanic passenger dataset (Frank, n.d.) to illustrate the demographic data distribution and the survival rate by various parameters such as age, passenger class, and sex. We will analyse the dataset to answer the following three questions:

- Who was on board the Titanic when it sank?
- Where were these passengers coming from, and where were they going?
- What were the demographics most likely to survive the sinking of the Titanic?

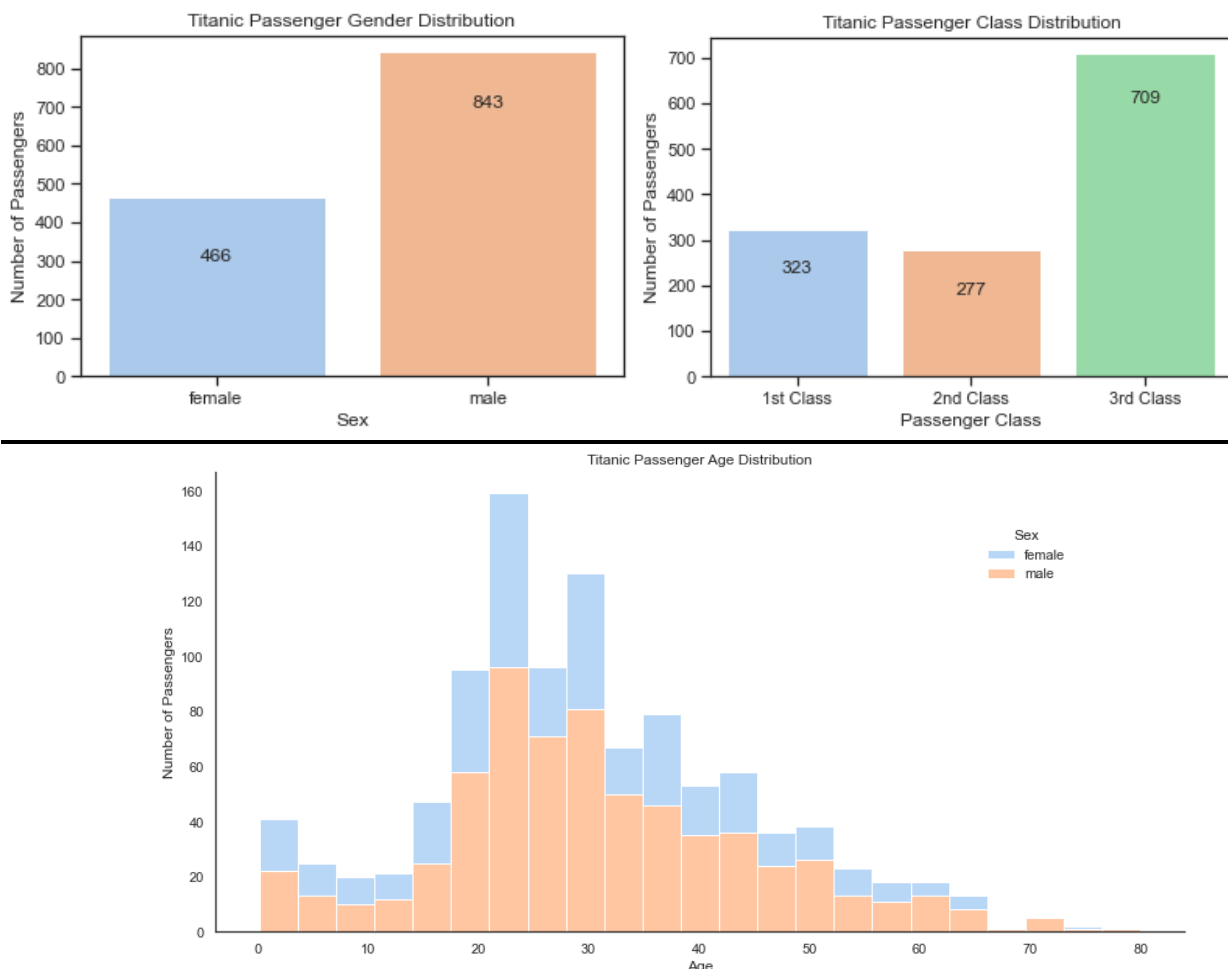
The Titanic passenger dataset contains information on 1309 passengers with 14 variables. According to the metadata, the dataset doesn't contain the crew information. Let's review the correlations between the variables. For the following correlation heat map, the sex variable has been converted to a boolean column labelled 'gender', where 'male' is represented as 0.



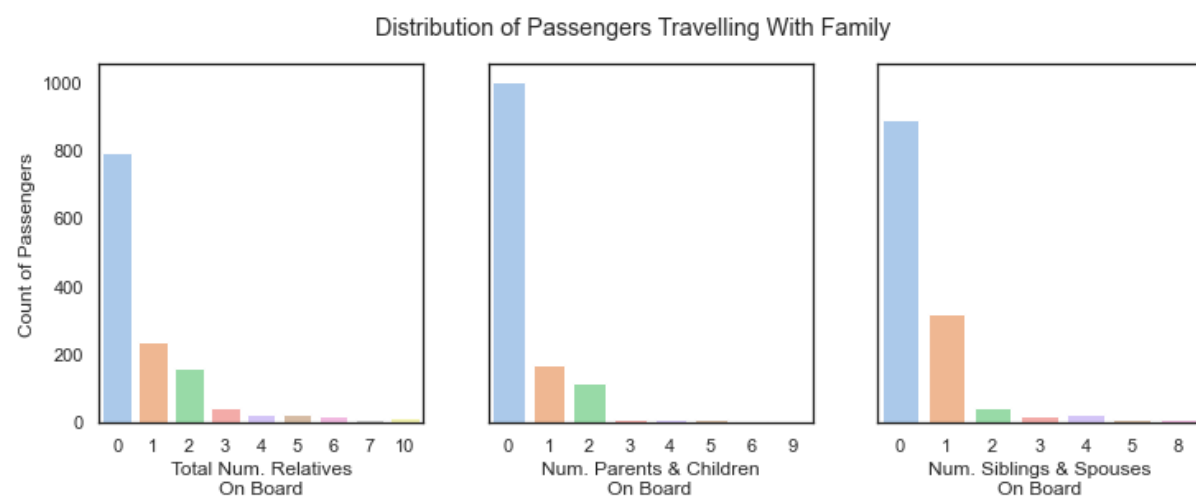
The most correlated variables are 'fare' and 'pclass', with a moderate negative correlation indicating that passengers with a higher passenger class (e.g. 3rd class passengers rather than 1st class passengers) paid less for their tickets. The only other variables that demonstrated moderate correlation are 'survived' and 'gender', with a positive correlation indicating that women were more likely to survive. The variables 'sibsp' (siblings/spouses onboard) and 'parch' (parents/ children onboard) are weakly positively correlated. The 'survived' and 'fare' (ticket purchase price) also have a very weak positive correlation. No correlation coefficient was calculated between 'survived' and 'body' (body identification number) as in every instance where 'survived' is 1, the 'body' variable is a null value. We will analyse the dataset further to find out more information.

## Who was on board the Titanic when it sank?

Titanic had over 1300 passengers on board when it began its maiden voyage. We explored the gender, passenger class, and age distributions of these Titanic passengers. We found that the majority of passengers were male, with 377 more male passengers than female passengers on board. The 3rd passenger class had the most passengers of the three passenger classes, about 54% of the total passengers. The passengers on the Titanic ranged in age from newborns to 80 year olds, though the majority of the Titanic passengers were in the age range of 20-30, especially the early 20s. Please see the details in the three graphs below.



790 passengers out of the 1309 on board travelled alone on the Titanic. About 40% of passengers travelled with one or more direct relatives. These direct relatives were parents and children ('parch' column), and siblings and spouses ('sibsp' column). Most travelled with a single relative, and that relative was more frequently a sibling or spouse than a parent or child. Please see the three graphs below.

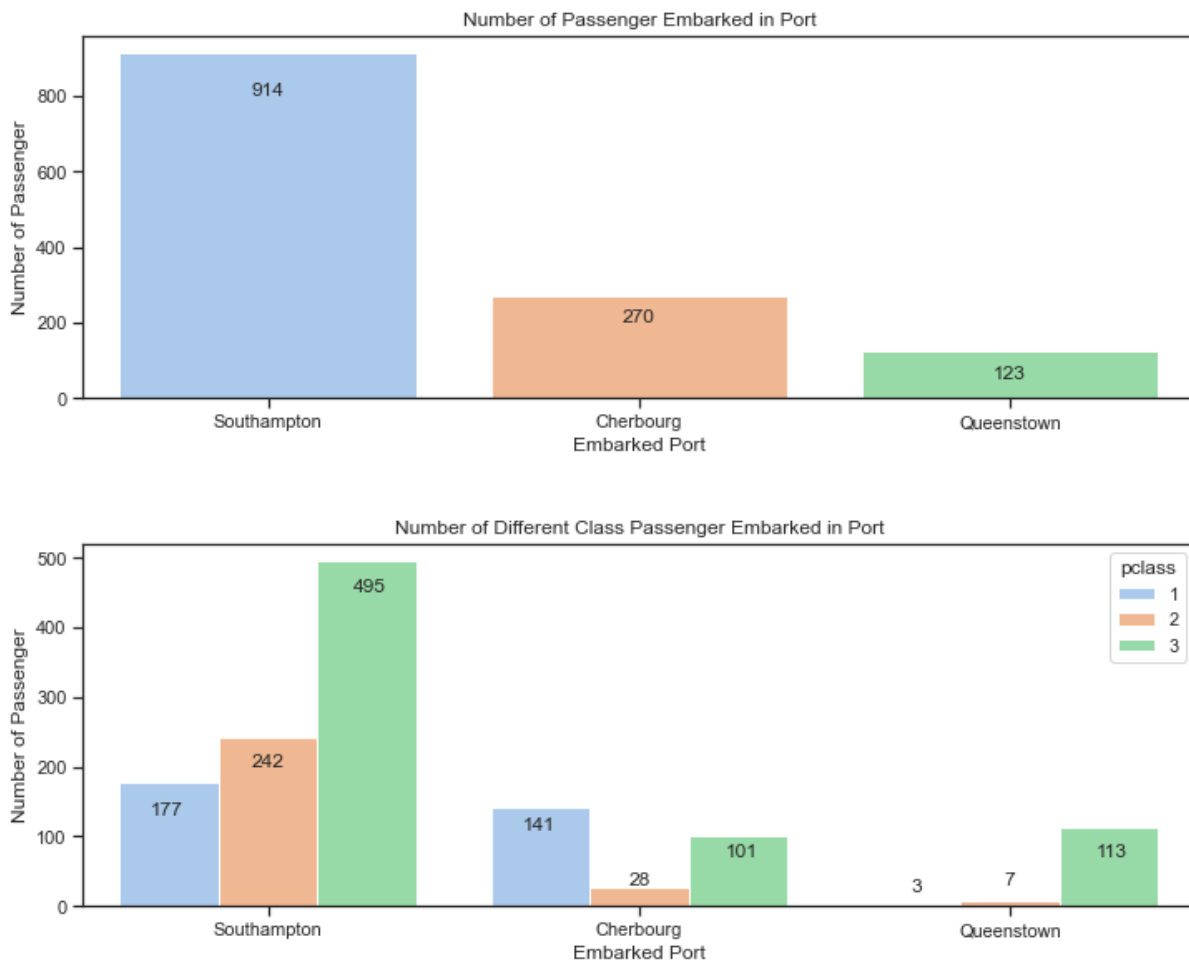


### Where were these passengers coming from, where were they going, and how much did they pay to get there?

914 passengers, almost 70% of the total passengers, boarded the ship in Southampton, England, as the embarkation port. In addition, 270 passengers boarded in Cherbourg, France, and 123 passengers boarded in Queenstown, Ireland.

Among those who boarded in Southampton, the majority (~54%) were in the 3rd passenger class, followed by 2nd passenger class (~26%) and 1st passenger class (~19%). Among those who boarded in Cherbourg, the majority (~52%) were 1st class passengers, followed by 3rd class passengers (~37%), and finally by 2nd class passengers (~10%). In Queenstown, the vast majority of passengers were in the 3rd class (~92%), followed by 2nd class (~6%) and 1st class (~2%).

Interestingly, only 10 passengers out of 123 boarded in Queenstown as 1st and 2nd classes, about 8% of its total passengers. Compared to the percentages of 1st and 2nd class passengers boarded in Southampton, which was about 46%, and Cherbourg, which was 63%, this was low. Please see the below two graphs.



When we began investigating where passengers were travelling to, we found that 564 records (~43%) didn't have the home/destination specified. For the passengers who did have a specified home/destination, we found some irregularities in the data. Most locations in the home/destination were cities in the United States and Canada, but oddly, some European cities such as London, Paris, and Belfast were found. This did not make sense with the transatlantic route the Titanic was travelling. Additionally, many entries included two different locations in the 'home.dest' column, while many others included only a single location. We postulate that the passengers might have been unsure if they should enter their home town or their destination town. Another explanation may be that this column was meant to capture the home town and destination of each passenger, but that this was done inconsistently. Therefore, we couldn't confirm the accuracy of the destination where most Titanic passengers were heading. Of records with a home.dest value, we found that most were going to New York, NY. Please see the table below.

#### Top 10 Destinations of Titanic passengers:

Home/ Destination	Number of Passengers
New York, NY	64
London	14
Montreal, PQ	10

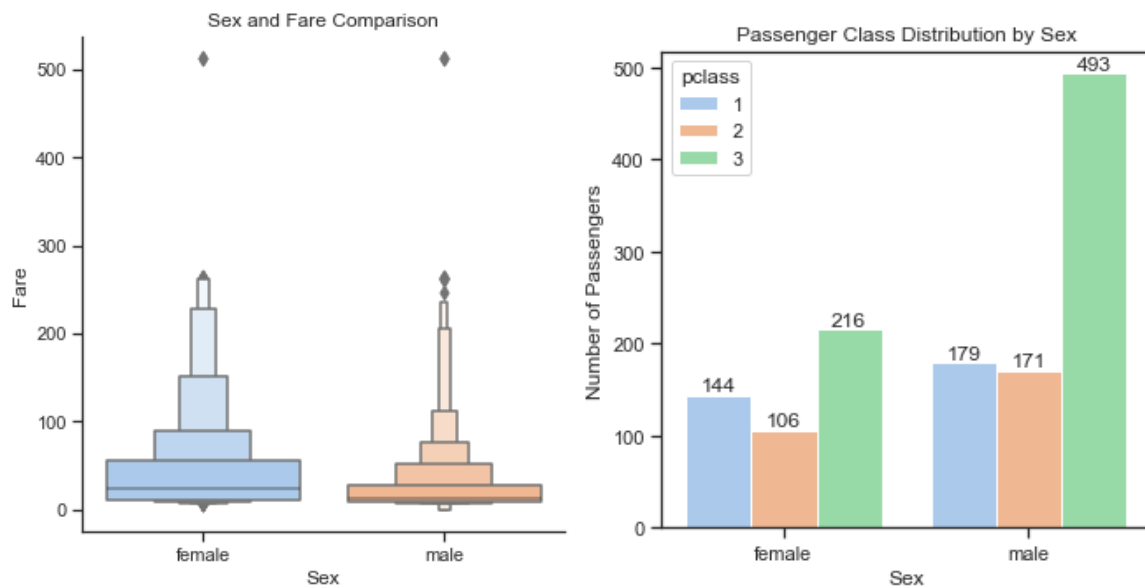
Home/ Destination	Number of Passengers
Paris, France	9
Cornwall / Akron, OH	9
Wiltshire, England Niagara Falls, NY	8
Winnipeg, MB	8
Philadelphia, PA	8
Belfast	7
Sweden Winnipeg, MN	7

When we inspected the Titanic fare for each port, we discovered that the price range was enormous. For example, the most expensive fare was the 1st class ticket with the embarkation port of Cherbourg, which cost £512.3292, while the cheapest fare was any class ticket with the embarkation port of Southampton, which cost £0.0000. We were suspicious about the no-cost ticket; therefore, we further investigated the passengers who boarded in Southampton and found 17 passengers distributed in all three classes with the £0.0000 fare. As a result, we suspected that the dataset didn't have complete fare records for these individuals. However, we couldn't conclude. Please see the two tables below.

embarked_port	pclass	fare		
		mean	min	max
Cherbourg	1	106.469871	25.7417	512.3292
	2	23.300593	12.0000	41.5792
	3	11.021624	4.0125	22.3583
Queenstown	1	90.000000	90.0000	90.0000
	2	11.735114	9.6875	12.3500
	3	10.390820	6.7500	29.1250
Southampton	1	72.148094	0.0000	263.0000
	2	21.206921	0.0000	73.5000
	3	14.435422	0.0000	69.5500

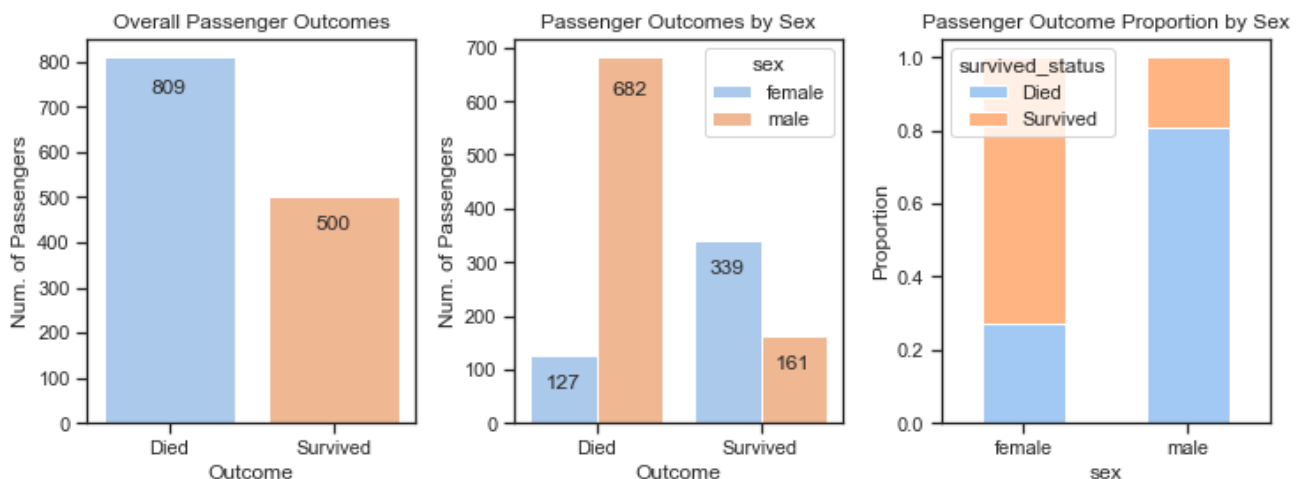
pclass	embarked	fare
7	1	S 0.0
70	1	S 0.0
125	1	S 0.0
150	1	S 0.0
170	1	S 0.0
223	1	S 0.0
234	1	S 0.0
363	2	S 0.0
384	2	S 0.0
410	2	S 0.0
473	2	S 0.0
528	2	S 0.0
581	2	S 0.0
896	3	S 0.0
898	3	S 0.0
963	3	S 0.0
1254	3	S 0.0

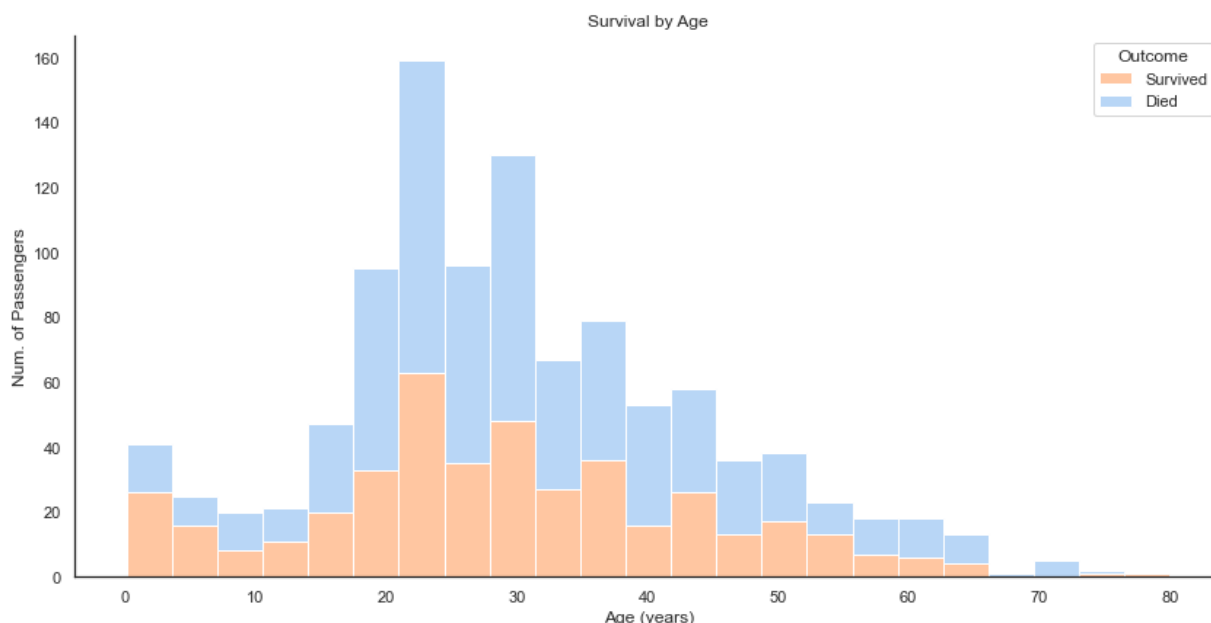
In addition, we compared the fare between male and female passengers, and we found that the average female passengers' fare was higher than the male passengers' fares. This may be explained by looking at the distribution of passenger class and sex; far more men were travelling in 3rd class (the class with the lowest fares) than women. Please see the graphs below.



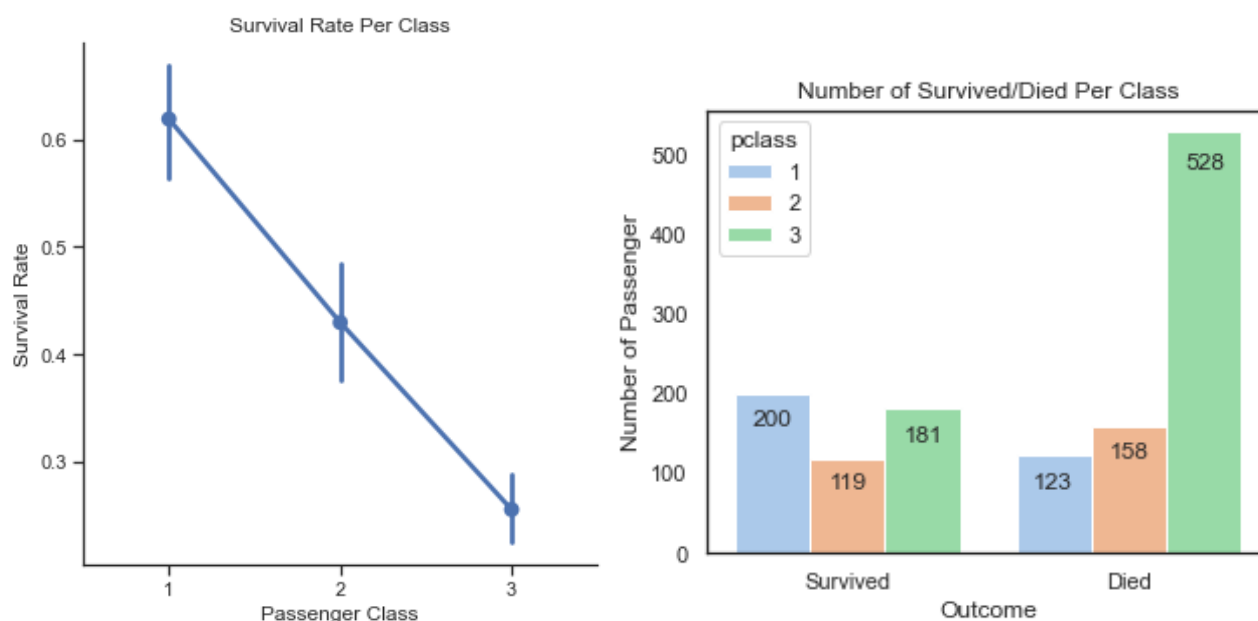
## What were the demographics most likely to survive the sinking of the Titanic?

The Titanic's collision with the iceberg and sinking in the North Atlantic Ocean was a tragedy. More than half of the passengers died during this accident; 809 passengers died, while only 500 passengers survived. The number of female survivors was more than double that of male survivors, despite female passengers accounting for only ~36% of the total passengers. Female deaths were also only a fifth of the male deaths. We can see by looking at the proportion of female and male passengers who died that male passengers were likely to die while female passengers were likely to survive. In terms of passenger age distribution, it is relatively normal, with all age groups experiencing deaths. Please see the four graphs below.





When we reviewed the survival rates for each passenger class, we found that the 1st class had a high survival rate of more than 60%, but the survival rate for the 3rd class was lower than 30%. When we checked the breakdown for the survival outcome for each class, we confirmed that the breakdown aligned with the survival rates. The 3rd class had more deaths than the 1st and 2nd classes. The 1st and 2nd classes had 123 and 158 deaths, while the 3rd class had 528 deaths. When we checked the surviving numbers, we found that there wasn't a notable difference among them. The 1st, 2nd, and 3rd classes had 200, 119, and 181 passengers survived. Please see the below two graphs.



We also looked for the youngest and oldest survivors and dead. The youngest survivor was only nine weeks old, and the oldest survivor was eighty years old. We searched the Internet and found that the youngest Titanic survivor, Millvina Dean, died on May 31, 2009, at 97. Please see the below tables for details.

#### Youngest & Oldest Survivors:

Name	Sex	Age	Number of Siblings/Spouses Aboard	Number of Parents/Children Aboard	Passenger Class	Embarkation Port
Dean, Miss. Elizabeth Gladys "Millvina"	Female	0.17	1	2	3rd Class	Southampton
Barkworth, Mr. Algernon Henry Wilson	Male	80.0	0	0	1st Class	Southampton

#### Youngest & Oldest Dead:

Name	Sex	Age	Number of Siblings/Spouses Aboard	Number of Parents/Children Aboard	Passenger Class	Embarkation Port
Danbom, Master. Gilbert Sigvard Emanuel	Male	0.33	0	2	3rd Class	Southampton
Svensson, Mr. Johan	Male	74.0	0	0	3rd Class	Southampton

## Conclusion

The Titanic, a British luxury passenger liner, set sail on its maiden voyage on April 10, 1912. It claimed that it was unsinkable. However, on April 15, 1912, four days after leaving England, the ship collided with an iceberg and sank in the North Atlantic Ocean. This tragedy killed 1502 out of 2224 passengers and crew. (History.com Editors, 2021; Piech, Chris, n.d.; Rajput, 2020; Tikkanen, 2022). We performed the exploratory data analysis (EDA) on the Titanic passenger dataset and discovered that it had more than 1300 passengers on board, including 843 males and 466 females. Their age range was from nine weeks old to eighty years old, and most of the passengers were in the 3rd passenger class. More than half travelled alone, and others travelled with one or more relatives. Most passengers embarked in Southampton, England, and headed to New York, NY. The fare cost varied widely depending on the embarkation port and the passenger class. After the ship sank, more passengers died than survived in the wreck and more female passengers were saved than male passengers, and many male deaths were in their early 20. We also confirmed that the 1st passenger class survival rate was much higher than the 3rd passenger class. Finally, the youngest Titanic survivor, Millvina Dean, died about thirteen years ago; she died on May 31, 2009, at 97. (Biography.com Editors, 2021; Burns, 2009).

## References

Biography.com Editors. (2021). "Millvina Dean Biography." Retrieved from <https://www.biography.com/historical-figure/millvina-dean#:~:text=Millvina%20Dean%2C%20the%20last%20survivor,hull%20on%20May%2031%2C%201911>

Burns, John F. (2009). "Millvina Dean, Titanic's Last Survivor, Dies at 97." Retrieved from <https://www.nytimes.com/2009/06/01/world/europe/01dean.html>



Frank, M.L. (n.d.). Titanic Datasets.

<http://campus.lakeforest.edu/frank/FILES/MLFfiles/Bio150/Titanic/TitanicMETA.pdf>

History.com Editors. (2021). "Titanic." Retrieved from <https://www.history.com/topics/early-20th-century-us/titanic>

Piech, Chris. (n.d.). A Titanic Probability. Retrieved from

<https://web.stanford.edu/class/archive/cs/cs109/cs109.1166/problem12.html#:~:text=On%20April%2015%2C%201912%2C%20the,better%20safety%20regulations%20for%20ships.>

Rajput, Faizah. (2020). "What The Titanic Looks Like Now Vs The Day It Sank." Retrieved from <https://www.thetravel.com/titanic-now-vs-sinking/>

Tikkanen, Amy. (2022). "Titanic." Retrieved from

<https://www.britannica.com/topic/Titanic>

**Slide Deck with Speaker Notes:**

<https://drive.google.com/file/d/1QXSvB1s0SrhVM4AwqVpR3VsLIHsj1Un4/view?usp=sharing>