

# 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

April 12, 2022

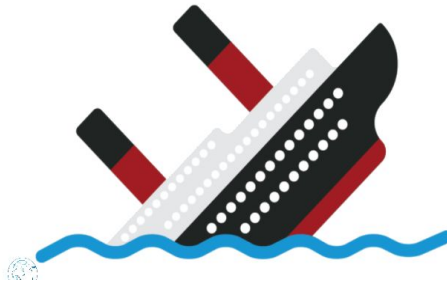


Image source:  
[https://www.kindpng.com/imgv/TToowoJo\\_transparent-titanic-png-transparent-titanic-clip-art-png/](https://www.kindpng.com/imgv/TToowoJo_transparent-titanic-png-transparent-titanic-clip-art-png/)

**Berkeley**  
UNIVERSITY OF CALIFORNIA

- Hello everyone! Welcome to our presentation for the DATASCI W200 final.
- My name is Kelianne and I am presenting with my teammate, Ivy.
- Today, we will be talking about our exploratory data analysis of the Titanic passenger dataset.
- Next slide, please

# Overview

Exactly 110 years ago this week...

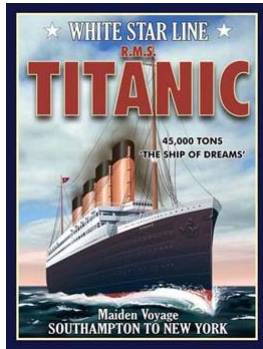


Image source:  
<https://i.stack.imgur.com/8807w>  
he-titanic-ever-publicised-as-unsinkable



This Titanic voyage map outlines the route Titanic intended to take between Southampton and New York.

Image source: <https://ultimatetitanic.com/titanic-voyage/>

**Berkeley**  
UNIVERSITY OF CALIFORNIA

- The RMS Titanic was a luxury ocean liner
- It was considered unsinkable due to a series of compartment doors that could be closed to maintain buoyancy
- It took three years to build and was promoted as “The Ship of Dreams” on the poster to the left.
- It set sail on its maiden voyage from Southampton, England on April 10, 1912, exactly 110 years ago this week, destined for a transatlantic trip to New York, NY.
- On April 15, 1912, only 4 days after leaving Southampton, it collided with an iceberg and sank in the North Atlantic Ocean.
- This tragedy killed 1502 out of 2224 passengers and crew.
- Next slide, please

## Three Questions

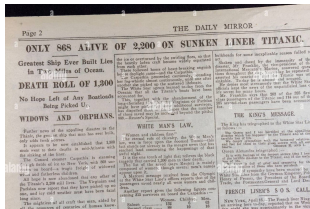


Image source:  
<https://www.alamy.com/headlines-inside-announcing-titanic-death-toll-the-daily-mirror-replica-newspaper-from-17th-april-1912-following-the-sinking-of-the-rms-titanic-image424946457.html>

1. Who was on board the Titanic when it sank?
2. Where were these passengers coming from, where were they going, and how much did they pay to get there?
3. What were the demographics most likely to survive the sinking of the Titanic?

**Berkeley**  
UNIVERSITY OF CALIFORNIA

- For this project, we performed exploratory data analysis on the Titanic Passenger Data set, which included data from 1309 passengers. No crew data was included in this data set.
- We wanted to investigate the data to help us answer three main questions:
  - Who was on board the Titanic when it sank?
  - Where were these passengers coming from, where were they going, and how much did they pay to get there?
  - And what were the demographics most likely to survive the sinking of the Titanic?
- Next slide, please

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

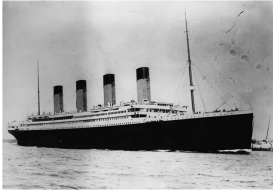
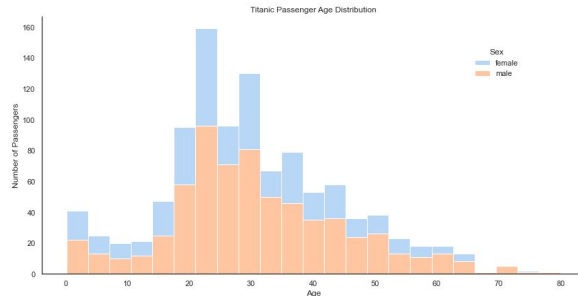
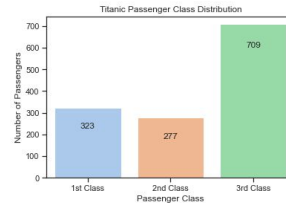
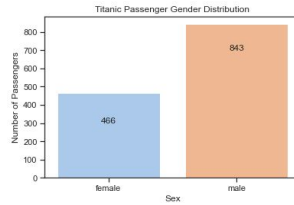


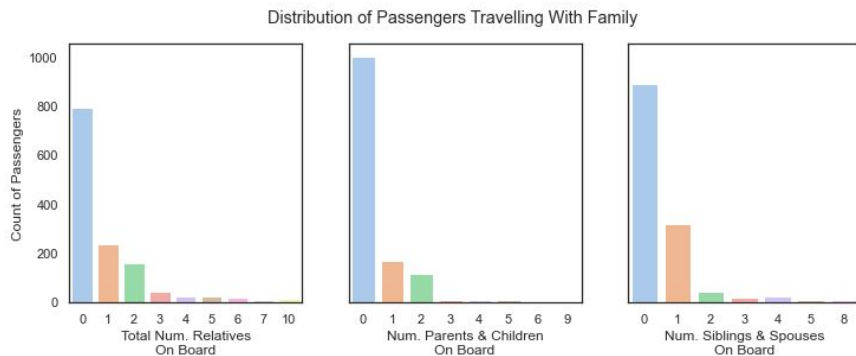
Image source:  
<https://www.si.com/tennis/2015/04/20/si-vault-richard-williams-karl-behr-titanic-tennis>



**Berkeley**  
 UNIVERSITY OF CALIFORNIA

- For the question, “Who was on board the Titanic when it sank?”, we found the following:
- The plot on the top left shows the number of female and male passengers on board the Titanic.
- From that we can see that there were a total of 1309 passengers onboard and 466, or about  $\frac{1}{3}$ , were female & 843, or about  $\frac{2}{3}$ , were male.
- The plot on the top right shows the number of passengers in each passenger class - from first class (which would have been the luxury class) to third class.
- About 54% of total passengers were in the 3rd passenger class
- The plot on the bottom of the slide shows the distribution of passenger ages, with male passengers in peach and female passengers in blue.
- Here we can see that the majority of the Titanic passengers were in the age range of 20-30, especially in the early 20 years old
- The passengers ranged in age from newborns to 80 year olds.
- Next slide, please

## Who was on board the Titanic when it sank? (cont)

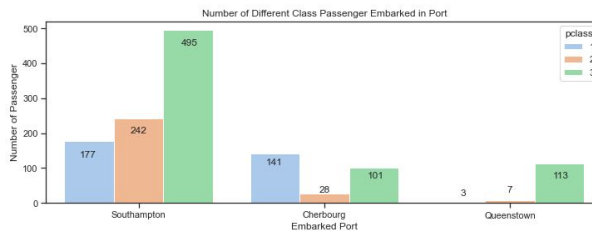
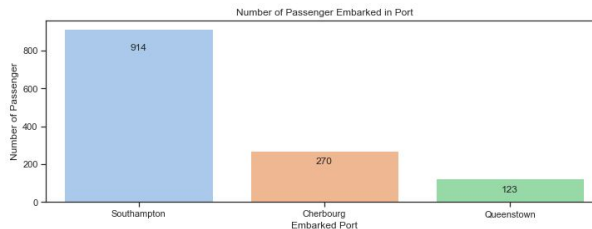


- We were interested in knowing more about the passengers on the Titanic, including who they were travelling with.
- The dataset we examined had 2 columns describing what direct relatives passengers were travelling with; one column was the number of parents and children a passenger was with, one column was the number of siblings and spouses a passenger was with.
- The first plot, on the left, shows the total number of direct relatives (including parents / children and siblings / spouses) passengers travelled with along the x-axis, and the number of passengers in each category on the y-axis. The second plot, in the center, shows the number of parents and children passengers travelled with, and the third plot on the right shows the number of siblings and spouses passengers travelled with.
- There were 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 & children and siblings & spouses
- Most travelled with a single relative
- The relative was more frequently a sibling or spouse than a parent or child
- Next slide, please

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



Image source:  
<https://www.thetravel.com/titanic-now-vs-sinking/>



**Berkeley**  
UNIVERSITY OF CALIFORNIA

- Next, we investigated where these passengers were coming from, and where they were going.
- The plot on the top of this slide shows the number of passengers that embarked from each port.
- 914 passengers, almost 70% of the total passengers on the Titanic, boarded the ship in Southampton England.
- An additional 270 boarded in Cherbourg, France, and a further 123 boarded in Queenstown, Ireland.
- The plot on the bottom of the slide shows the passenger class for passengers who boarded at each port.
- Among those who boarded in Southampton, the majority (~54%) were in the 3rd passenger class, followed by 2nd and 1st class passengers.
- Interestingly, only 10 passengers out of 123 who boarded in Queenstown were in 1st or 2nd class, about 8% of Queenstown's total passengers.
- This was notably lower when compared to Southampton, where 46% of passengers were in the 1st or 2nd class, and Cherbourg, where 63% of passengers were in the 1st or 2nd class.
- Next slide, please

**Where were these passengers coming from, and where were they going, and how much did they pay to get there? (cont)**

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

fare				
		mean	min	max
embarked_port	pclass			
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

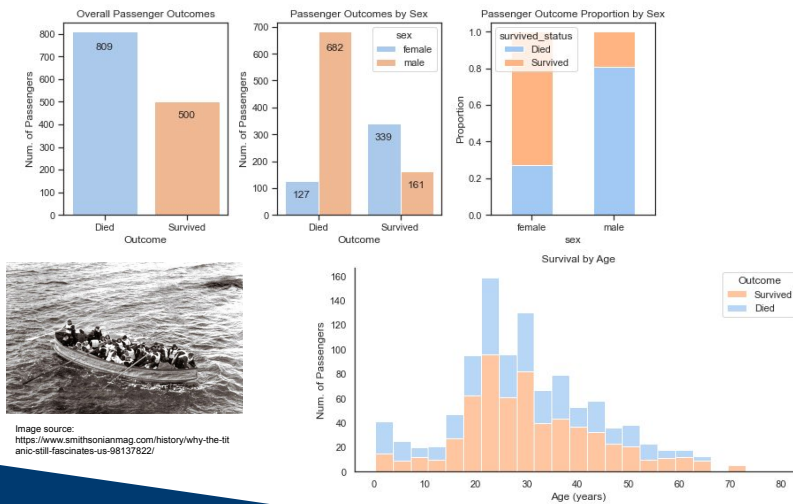
	pclass	embarked	fare
7	1	S	0.0
70	1	S	0.0
125	1	S	0.0
150	1	S	0.0
170	1	S	0.0
223	1	S	0.0
234	1	S	0.0
363	2	S	0.0
384	2	S	0.0
410	2	S	0.0
473	2	S	0.0
528	2	S	0.0
581	2	S	0.0
896	3	S	0.0
898	3	S	0.0
963	3	S	0.0
1254	3	S	0.0

- We then began investigating the destination of these passengers, but soon found data irregularities.
- 564 records (~43% of the dataset) didn't have a value in the home/destination column
- 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 as destinations with the transatlantic route the Titanic was travelling.
- We postulate that the passengers might have been unsure if they should provide 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. Some passengers had two locations listed in the columns, while others only had a single location listed, adding to our confusion.
- Therefore, we couldn't confirm the accuracy of the destination where most Titanic passengers were heading.
- We also looked in to how much people paid for their Titanic tickets, and found the fare range was enormous.
- For example, for passengers embarking from Cherbourg, fares ranges from ~4

- pounds to ~£512, the most expensive fare in the dataset.
- The cheapest fares were tickets with the embarkation port of Southampton, which cost £0
- We further investigated these passengers who boarded in Southampton and found 17 passengers distributed in all three classes with the £0 fare, which we suspect are not accurate data points.
- Next slide, please



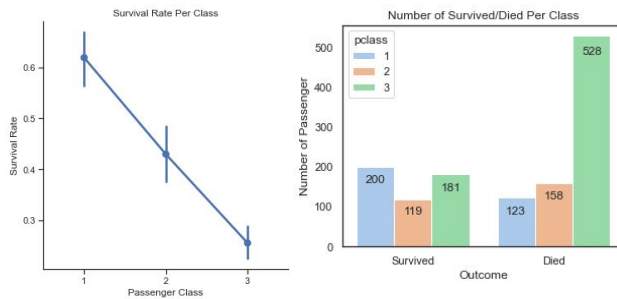
## Where were the demographics most likely to survive the sinking of the Titanic?



**Berkeley**  
UNIVERSITY OF CALIFORNIA

- For the question, “What were the demographics most likely to survive the sinking of the Titanic?”, we found the following:
- 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.
- Female deaths were only a fifth of the male deaths.
- In terms of passenger age distribution, it is relatively normal.
- Next slide, please

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



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

- 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%.
- We checked the class breakdown for the survival outcome for each class, we confirmed the breakdown aligned with the survival rate
- 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.
- We also discovered the youngest and oldest survivors.
- The youngest survivor was Miss Millvina Dean who was nine weeks old
- The oldest survivor was Mr. Wilson Barkworth who was 80 years old
- Coincidentally, they both were the youngest and oldest passengers on the Titanic.
- Next slide, please

## Conclusion



Image source:  
<https://www.businessinsider.com/titanic-sinking-animation-game-ship-2016-4>

**Berkeley**  
UNIVERSITY OF CALIFORNIA

- The sinking of the Titanic was an unexpected tragedy
- The ship took three years to build, and only fifteen minutes to sink
- More than half of the passenger lives went down with the ship in the North Atlantic Ocean.
- After answering our initial questions through our exploratory data analysis, we were interested in one last piece of research.
- We were searched for any news about the youngest survivor, and we found that Millvina Dean, who was only nine weeks old when the tragedy happened, died on May 31, 2009, at the age of 97.
- Thank you for listening to our presentation.
- We look forward to any questions or feedback you may have.

# 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/MLFiles/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>

- End of presentation
- Thank you