Analysis of the Survivor Pattern of the RMS Titanic

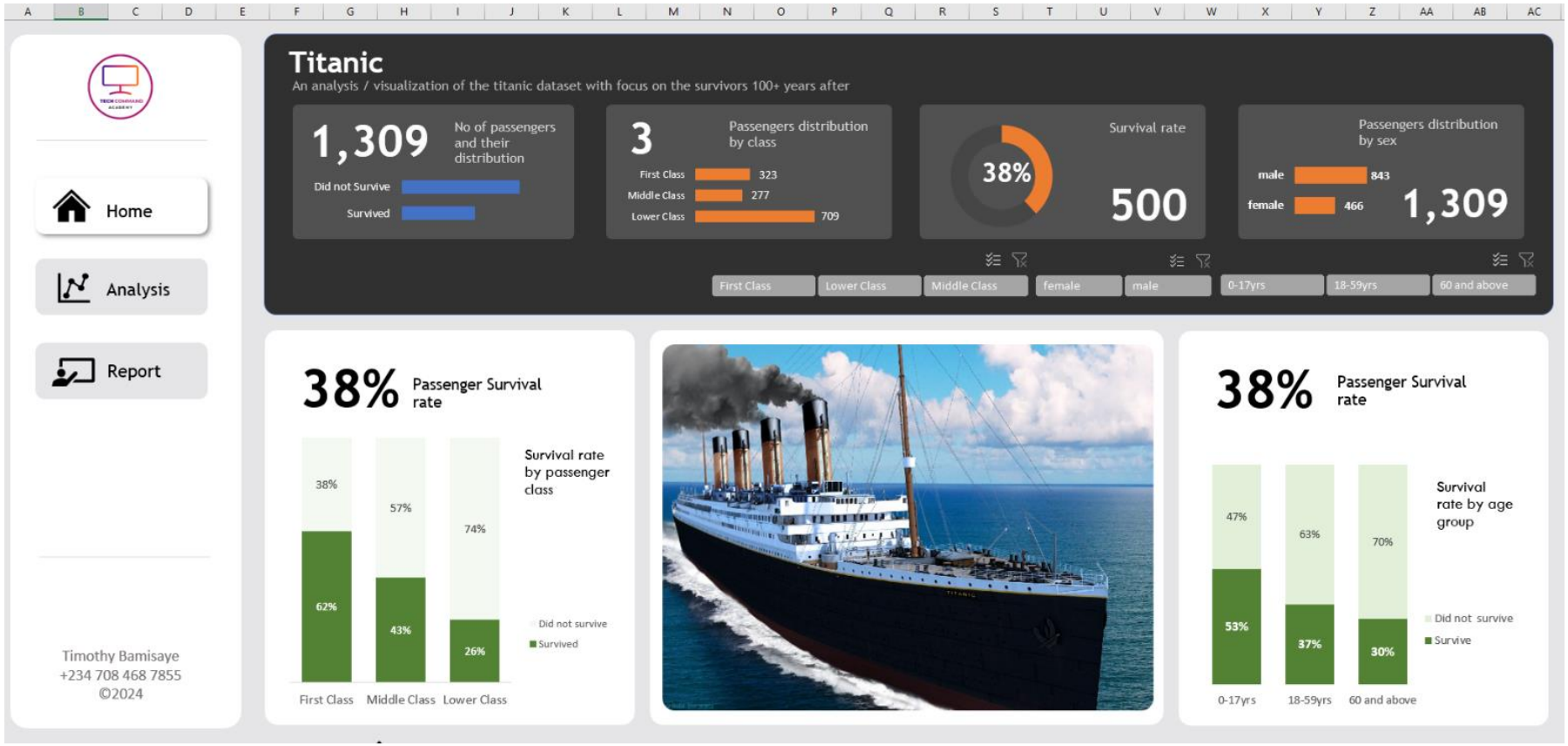
An analysis conducted by

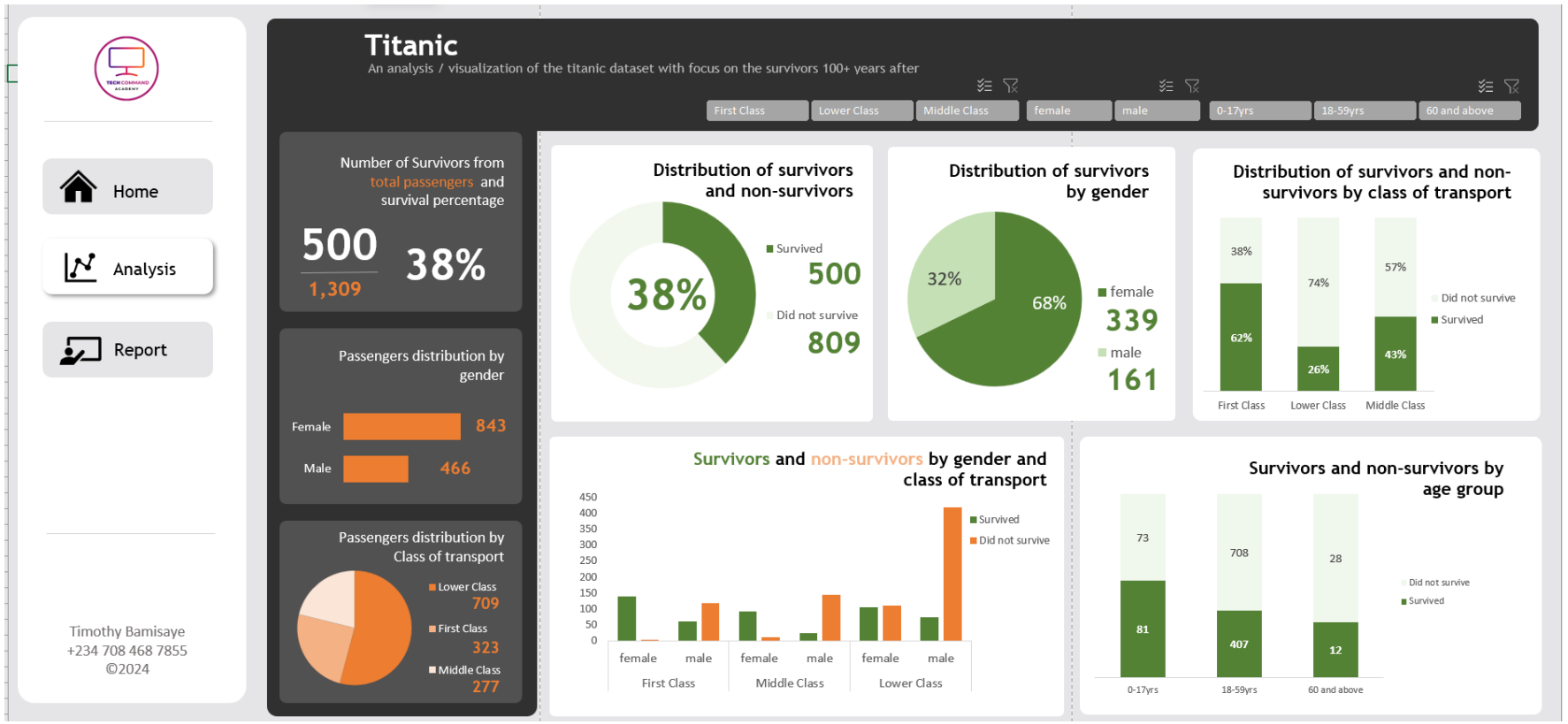
Timothy Bamisaye

+234 708 468 7855

https://bamisaye3.github.io/portfolio | bamisaye3@gmail.com | https://linkedin.com/in/bamsayetimothy

View Dashboard👇 👇




Titanic
An analysis / visualization of the titanic dataset with focus on the survivors 100+ years after
First Class
Lower Class
Middle Class
female
male
0-17yrs
18-59yrs
60 and above
Home
Analysis
Report
Timothy Bamisaye
+234 708 468 7855
©2024
Number of Survivors from total passengers and survival percentage
500
1,309
38%
Passengers distribution by gender
Female
843
Male
466
Passengers distribution by Class of transport
Lower Class
709
First Class
323
Middle Class
277
Distribution of survivors and non-survivors
38%
Survived
500
Did not survive
809
Distribution of survivors by gender
32%
68%
female
339
male
161
Distribution of survivors and non-survivors by class of transport
38%
62%
74%
26%
57%
43%
First Class
Lower Class
Middle Class
Did not survive
Survived
Survivors and non-survivors by gender and class of transport
450
400
350
300
250
200
150
100
50
0
female
male
First Class
Middle Class
Lower Class
Survived
Did not survive
Survivors and non-survivors by age group
73
81
708
407
28
12
0-17yrs
18-59yrs
60 and above
Did not survive
Survived

INTRODUCTION

The RMS Titanic popularly known as Titanic, one of the most infamous maritime disaster in history was a British passenger liner, it was the largest ship in the world at the time measuring 270m in length and 28m in width.

In 1912 it began it maiden voyage with over 2200 passenger and crew, on April 15th the titanic struck an iceberg on it starboard (right) which led to significant damage to the ship and eventual sinking that led to the loss of over 1500 lives. This data attempt to analyze factors that influence survival if any, and the chances of survival of the different passenger aboard the ship.

This data was sourced from Kaggle and a big credit to vinicius150987 who made it dataset available. You can find a copy of the dataset here.

https://www.kaggle.com/datasets/vinicius150987/titanic3

OBJECTIVES

The objective of this analysis is to attempt to analyze factors that influence survival if any, and the chances of survival of the different passenger aboard the ship.

DATA CEANING & TRANSFORMATION

As usual every data must be cleaned and transformed before being analyzed, to do that, I used power Query for my data cleaning.

1. Verified and ensure all relevant columns to be used in the analysis contain the right data type and those with the wrong type were duly converted to the appropriate type
2. Created Age groups: For the analysis, I will love to see survival rate in relation to the age of the passengers aboard the ship. This is to ascertain if there is a correlation between passengers age and chances of survival, I therefore created 3 age groups those between 0 and 17years were grouped as 0-17years, then young adults as 18-59years and elderly as 60 and above.
3. Replaced values and removed columns: It was observed that the age column contained a number of blank cells – to handle this, we could either leave as blank and analyze it or use the median or mean to replace all the missing values. I chose to replace the missing values with the median which turned out to be 29. Also note that the missing values was about 20% of the total rows available which is within the acceptable range of less than 30%
4. Removed Columns: Columns with more than 30% missing values were automatically removed as they are not fit for analysis, also other columns not used within our analysis were removed - these columns are Ticket, Fare, Home Destination, Port of Embarkment

MODELLING, ANALYSIS (FINDINGS) AND VISUALIZATION

The data was analyzed using Pivot Tables and visualized using pivot charts

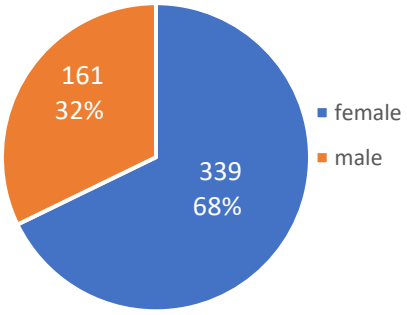
1. **Passenger:** of the 1309 passengers recorded, 843 were male and 466 were female
2. **Survivors:** of the 1309 passengers 500 of them survived. Giving a survival rate of 38%.
3. **Distribution of Survivors by gender:** of the 466 females aboard the ship a whooping 339 survived, this amounts to 73% of the female population surviving. For the male gender only 161 of the 843 survived giving a survival rate of 19%

Row Labels	Survived	Did not survive	Grand Total	Survival %
female	339	127	466	73%
male	161	682	843	19%
Grand Total	**500**	**809**	**1,309**	38%

FINDINGS: Despite having less female aboard the ship, more of them survived as compared with the male

4. **Comparing survivors by gender:**
Trying to dig deeper into the survivor data by gender, we found that of the total 500 survivors, 339 female amounts to 68% of the total survivors and the male amounts to 32%

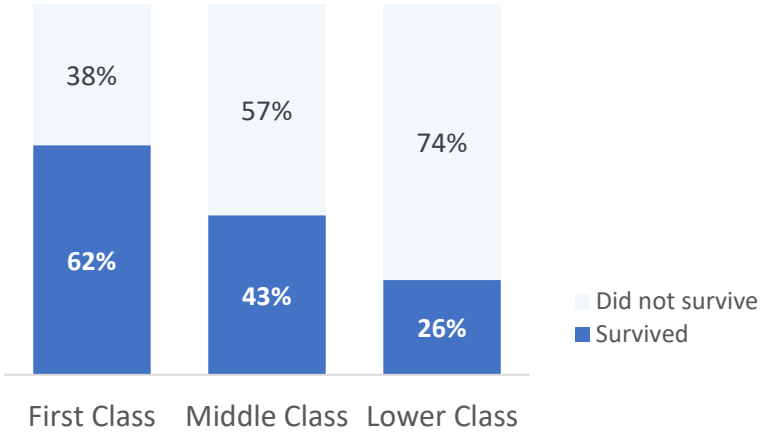
Row Labels	Survivors	%
Female	339	68
Male	161	22
Grand Total	**500**	**100**



FINDINGS: we found that of every 100 survivors, about 68 of them were female, this obviously is not just a random occurrence considering there were more male aboard and the male are stronger, this only point to a more structured approached to evacuation of the passengers and this approach tilted slightly towards favouring female survivors.

5. **Analysis by passenger class:** The ship has 3 different classes denoted by 1,2 and 3 in the dataset each of them implying First class, Middle Class and Lower Class respectively.

Row Labels	Survivors	Passengers	%
First Class	200	323	62%
Middle Class	119	277	43%
Lower Class	181	709	26%
Grand Total	**500**	**1309**	38%



From the analysis, 62% of those with First Class tickets survived, the rate dropped to 43% for those in middle class and only 26% of those in the lower class survived
VERDICT: The class of ticket played a huge role in the survival of passengers, seems those with higher status, were able to secure necessary requirements that ensured their survival.

6. **Analysis by passenger class and gender:**

	Survived	Did not survive	Grand Total	
First Class	**200**	**123**	**323**	**62**
female	139	5	144	97%
male	61	118	179	34%
Middle Class	**119**	**158**	**277**	**43%**
female	94	12	106	89%
male	25	146	171	15%
Lower Class	**181**	**528**	**709**	**26%**
female	106	110	216	49%
male	75	418	493	15%
Grand Total	**500**	**809**	**1309**	**38%**

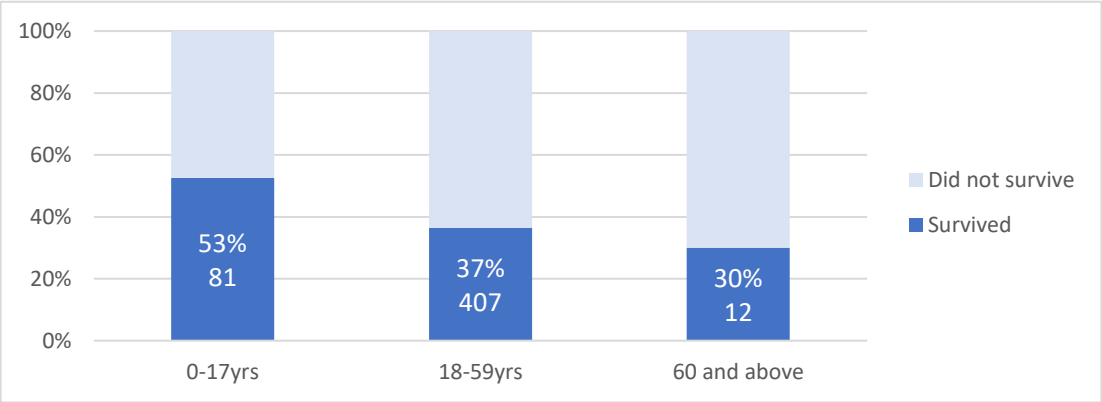
With it already being established that passenger class played a role in the survival of the passengers, a further check showed even among each class, the gender also played a role. For those with first class ticket, only 5 of the 144 females onboard did not survive implying 97% survival rate for female in the first class cabin. The same cannot be said for the male gender, as a drastic drop was recorded with only 34% surviving (61 out of 179). The same trend holds true for those in the middle class, 89% of women survived with only 15% of the men survived and finally in the lower class 49% of the female survived while just 15% survived.

VERDICT: with the consistency across board, this cannot be a random occurrence, it can be said that even amongst the different passenger classes, conscious effort was made to ensure more female survivors

7. Survival by Age Groups: For this analysis, the passengers were grouped into 3 age categories. Age 0-17 were grouped as 0-17years, age 18-59, as 18-59years and those above 60 as 60 and above.

Row Labels	Survived	Did not survive	Grand Total	%
0-17yrs	81	73	154	53%
18-59yrs	407	708	1115	37%
60 and above	12	28	40	30%
Grand Total	**500**	**809**	**1309**	38%

A total of 154 passengers were aged between 0 and 17years, 53% of this population survived, 407 of the 1115 people were between 18 and 59years which amounts to 37% survival rate for this age group, and for the elderly (60 and above), only 30% of them survived.
VERDICT: there seem to have been conscious effort to ensure more children survived, followed by the working population (18-59years).



CHALLENGES

Quite a number of challenges were encountered during while analyzing this data, while in reality there were over 2200 passengers aboard the titanic, this dataset only records details of 1309 passengers

Missing values, some of the columns were completely removed as too many rows were missing from those columns, some others were also removed as they played no significant role in the analysis.

The age column used in the analysis contained some blank cells, we used the median age of 28 to replace all the missing values. Also note that the missing values was about 20% of the total rows available which is within the acceptable range of less than 30%

SUMMARY

From our analysis, through the pandemonium and crisis occurring, there were conscious effort by the captain and crew of the ship to ensure priority were given to children and young adults, we also found out that class of ticket played a major role in survival of passengers, finally, the female were also prioritized over men during the evacuation process.