APRIL 2024

TITANIC DATA ANALYSIS REPORT

Exploring the Titanic passenger dataset to uncover insights and trends using Power BI's intuitive data visualization capabilities

PREPARED BY

DATA ELITES POWER BI TEAM A

Table of Contents

Project Overview

The objective of this project is to conduct a comprehensive analysis of the Titanic dataset using Power BI. We aim to extract valuable insights regarding passenger demographics, survival rates, and other relevant factors to gain a deeper understanding of the factors influencing survival on the Titanic

Dataset Overview

Data collection method

The Titanic dataset was provided by the Data Elites team.

The Dataset comprises the following columns

- PassengerId: A unique identifier assigned to each passenger in the dataset
- Survived: Indicates whether the passenger survived (1) or not (0)
- Pclass: Represents the passenger's class (1st, 2nd, or 3rd).
- Name: The name of the passenger.
- Sex: Gender of the passenger, categorized as male or female.
- Age: The age of the passenger at the time of boarding the Titanic.
- SibSp: The number of siblings or spouses aboard the Titanic for the
- Parch: The number of parents or children aboard the Titanic for the
- Ticket: Ticket number associated with the passenger

- Fare: The fare paid by the passenger for their ticket.
- Cabin: Cabin number.
- Embarked: Port of embarkation (Cherbourg, Queenstown, or Southampton)

Data Summary

This dataset provides valuable information about passengers aboard the Titanic, including their demographics, family relationships, and survival outcomes. Through comprehensive analysis, we aim to uncover insights into the factors influencing survival rates and passenger demographics during this historic event

Data Cleaning And Modeling

Data Cleaning

Handling Missing Values

Missing values in each column were replaced with the word "Unspecified

Removing Unnecessary Column

The column named "Cabin" was removed from the dataset as it was not utilized in the analysis

Standardizing Values

Values "C", "Q", and "S" in the column named "Embarked" were replaced with "Cherbourg", "Queenstown", and "Southampton", respectively.

Renaming and Recategorizing Column

The column named "Passenger" was renamed to "Passenger Class".

The values in the "Passenger Class" column were replaced as follows:

- 1 replaced with "First Class".
- 2 replaced with "Middle Class".
- 3 replaced with "Third Class".

The Column named “Survived was renamed to “Survival Status

The values in the Survival Status Column were replaced as follows:

- 0 replaced with “Deceased”
- 1 replaced with “Survived”

Creation of New Columns

Age Category: Created from the "Age" column to categorize the age distribution.
Fare Distribution: Created from the "Fare" column to categorize the fare distribution of each passenger.

Data Modeling

Flat Data Schema

Utilized a flat data schema since the dataset consists of a single table with no dimension tables.
This approach simplifies the data structure and allows for easier analysis and visualization directly from the single table.

DAX Measure Created

1. **Total Passenger:**
   Calculates the overall count of passengers onboard the Titanic.

2. **Total Deceased:**
   Calculates the count of passengers who did not survive (deceased).

3. **Total Survived:**
   Calculates the count of passengers who survived.

4. **Family Size:**
   Calculates the family size onboard for each passenger based on the 'SibSp' and 'Parch' columns.

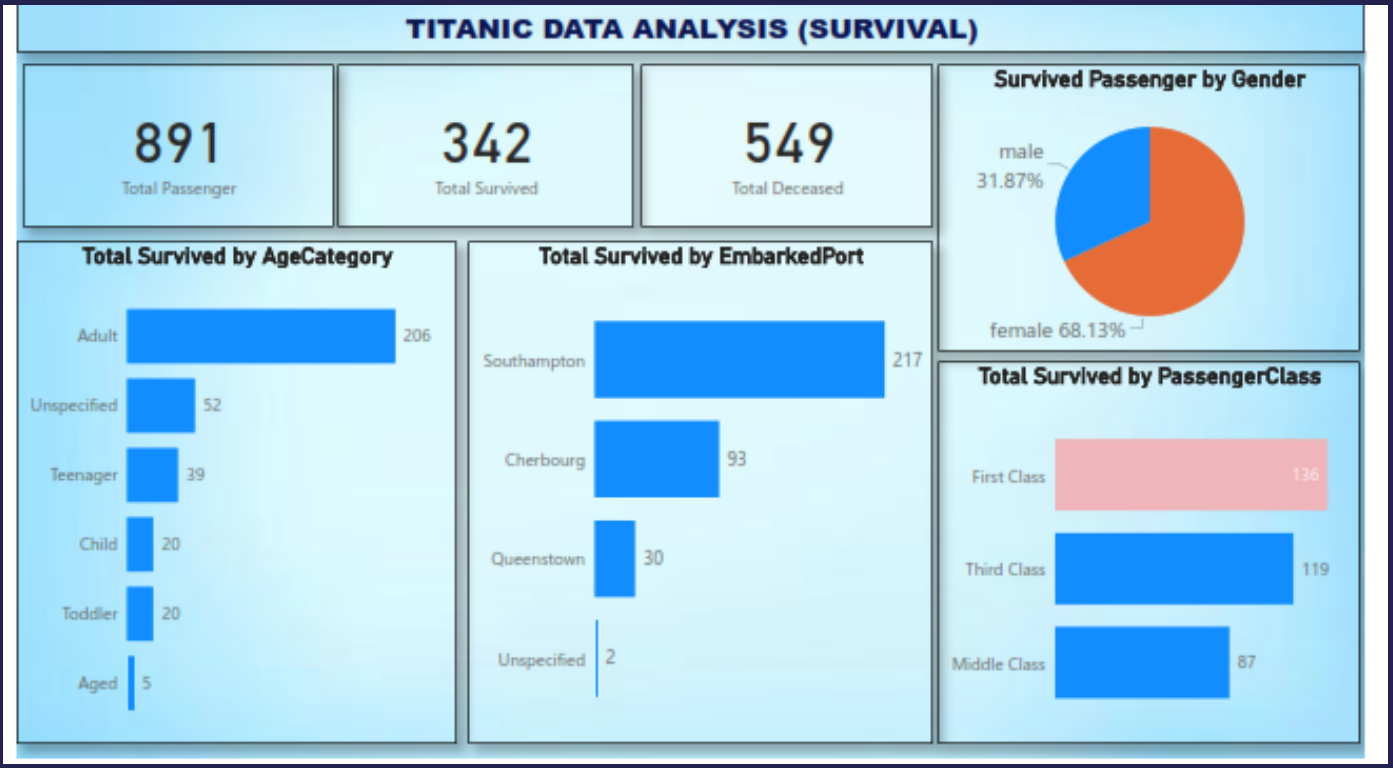
5. **Family Size Category:**
   Categorizes passengers into those who traveled alone and those who traveled with family based on the calculated family size.

6. **Survived by Passenger Based on Sex:**
   Shows the distribution of passenger survival status (died or survived) based on their sex (gender)

Result:
Measures have been created to facilitate easier analysis and visualization of the Titanic dataset.
These measures provide insights into passenger demographics, survival rates, and family dynamics onboard the Titanic, aiding in comprehensive data exploration and interpretation

Data Visualization

Survival Analysis



Key Insights:

1. **Total Passengers Onboard:** 891 passengers boarded the ship.

2. **Survival Statistics:**
   Survived: 342 passengers survived.
   Deceased: 549 passengers did not survive.

3. **Gender Analysis:**
   Female passengers had the highest survival rate.

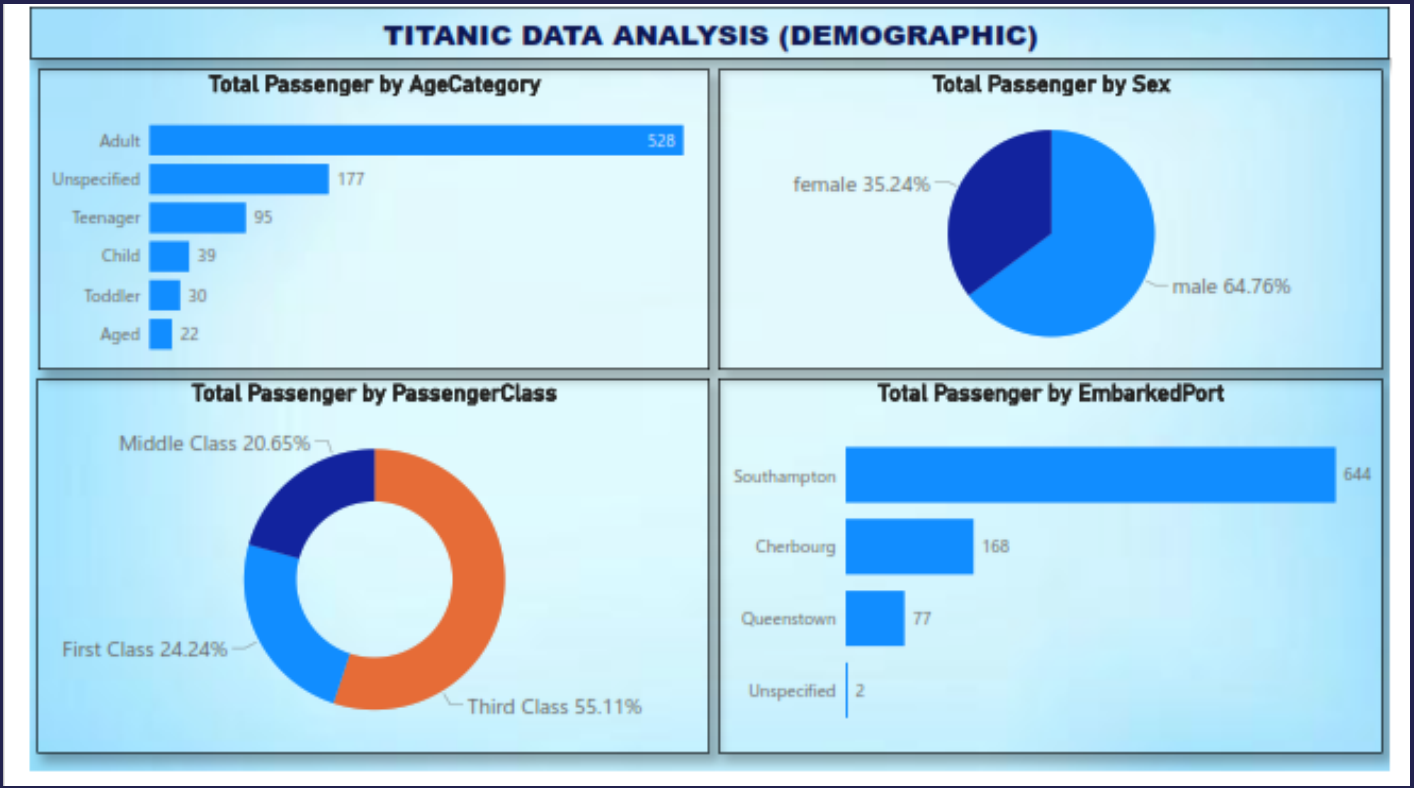
4. **Age Group Analysis:**
   Passengers between the ages of 20 and 60 had the highest survival rate.

5. **Passenger Class Analysis:**
   Firstclass passengers had the highest survival rate.

6. **Embarkation Port Analysis:**
   Passengers from Southampton had the highest survival rate.

Data Visualization

Demographic Analysis



Key Insights:

Passenger Class Distribution:

Third-class passengers boarded the ship the most.

Age Group Distribution:

Adults boarded the ship the most.

Gender Distribution:

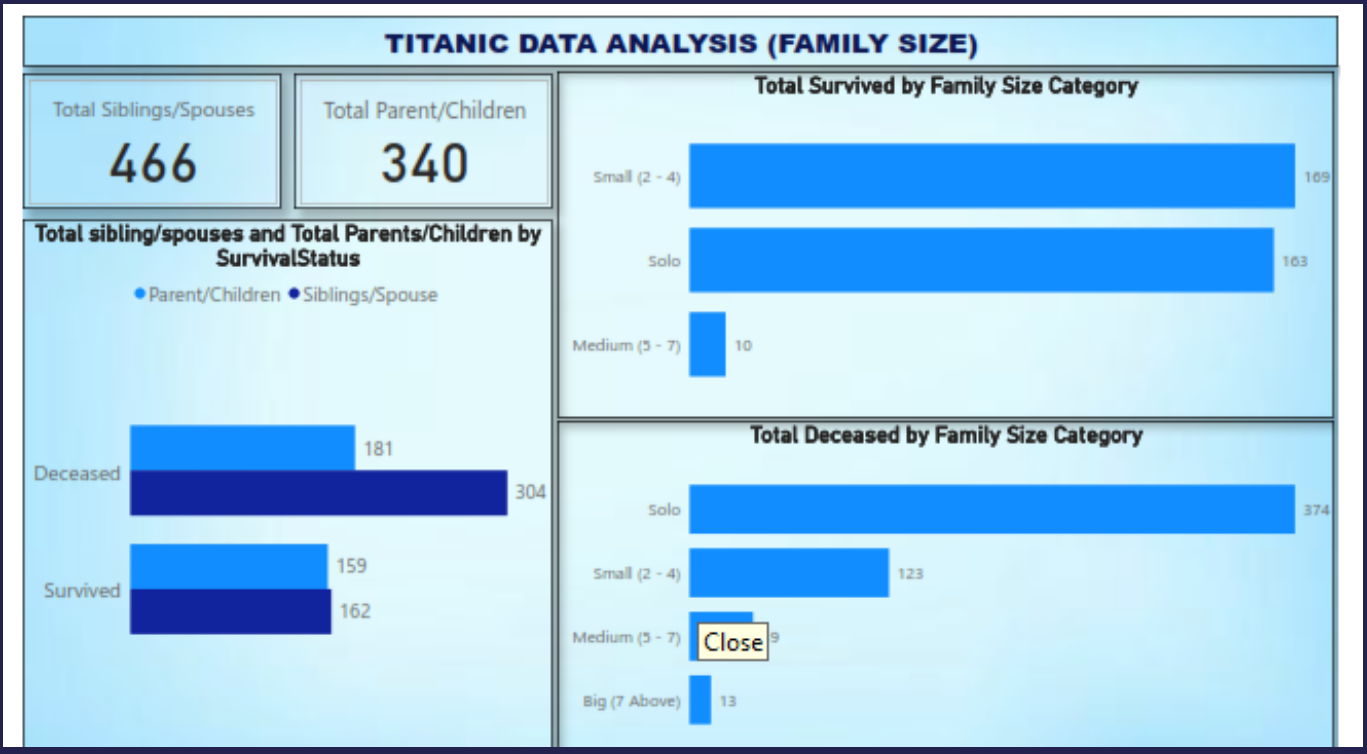
Male passengers boarded the ship the most.

Embarkation Port Distribution:

People embarked from Southampton the most.

Data Visualization

Family Size Analysis



Key Insights:

Survival by Family Size:

Passengers traveling alone died the most, followed by those with small family sizes.

Passengers traveling with 2-3 siblings survived the most.

None of the big families survived.

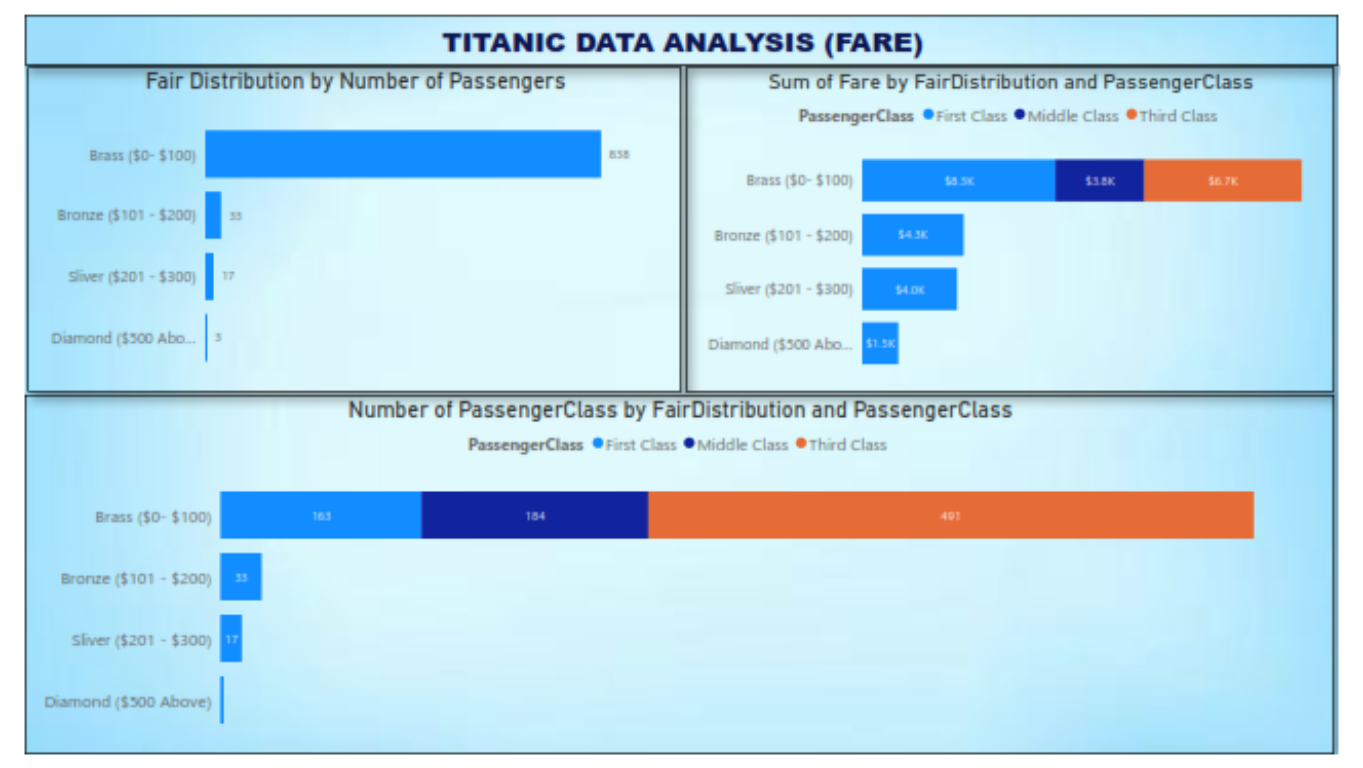
Number of Passengers with Family Members:

340 passengers had parents or children onboard.

466 passengers had siblings or spouses onboard.

Data Visualization

Fare Analysis



Key Insights:

Fare Distribution by Passenger Class:

First-class passengers paid the highest fares.

Third-class passengers paid the lowest fares.

Most Common Fare Range:

The majority of passengers paid below $100 for their fare

Conclusion

Fare Analysis Conclusion:

Fare analysis revealed significant disparities in fare amounts across passenger classes, highlighting the importance of equitable pricing strategies in the maritime industry.

Family Size Analysis Conclusion:

Family size analysis underscored the impact of social support networks on survival outcomes, emphasizing the need for inclusive emergency preparedness measures.

Demographic Analysis Conclusion:

Demographic analysis provided insights into passenger composition, informing targeted marketing efforts and onboard service enhancements to cater to diverse passenger needs.

Survival Analysis Conclusion:

Survival analysis identified demographic factors influencing survival rates, guiding the development of safety protocols and onboard facilities to prioritize passenger safety.

Future Directions:

Further research and analysis are recommended to explore additional factors influencing passenger experience and safety onboard maritime vessels.

Continued collaboration with industry stakeholders and regulatory bodies is essential to address emerging challenges and enhance passenger safety standards

Thank you!

Data Elites Power BI Team A

LinkedIn:

Data Analytics Elites Global Community