**Introduction to Data Visualization**



**Data Report on**

**Exploratory Analysis on Titanic Dataset Using Tableau**

Table of Contents

[DATASET 3](#_Toc78303992)

[**INTENDED AUDIENCE OF THE REPORT** 4](#_Toc78303993)

[SOME FACTS ABOUT TITANIC 4](#_Toc78303994)

[METADATA 4](#_Toc78303995)

[OBJECTIVE OF THE PROJECT 6](#_Toc78303996)

[DESCRIPTIVE DATA ANALYSIS using SAS 6](#_Toc78303997)

[**DATA VISUALISATION** 8](#_Toc78303998)

[EXPLORATORY DATA ANALYSIS or FINDINGS 9](#_Toc78303999)

[LIMITATIONS OF THE DATASET 15](#_Toc78304000)

[**IMPORTANT INFORMATION FROM OUTSIDE RESOURCES** 15](#_Toc78304001)

[**RECOMMENDATIONS** 16](#_Toc78304002)

[DECLARATION 17](#_Toc78304003)

# DATASET

This dataset contains information about passengers who travelled on the titanic ship. The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. The goal of the

project is to **predict the survival of passengers** based off a set of datasets given. The information provided for passengers includes weather they survived, their ticket-class, place of boarding.

The objective of the Project is to determine and conclude Survival Rate based influencing variables such as age, passenger class, gender, port of embarkation and other variables with high correlation value. The idea is **to keep Survived as the main variable** and determine the Survival Rate using correlation and various statistical analysis.

Using this dataset, we need to predict whether a passenger on the titanic would have been survived or not. It provides information on the fate of passengers on the Titanic, summarized according to economic status (class), sex, age, and survival.

A snippet of the data set is shared below (The whole data set is emailed along with Table

Description automatically generatedthis file)The project of data visualization on the titanic data set aims to research and find out interesting facts on the tragic voyage that the iconic ship faced in the history. While most of the analyses revolve around survival rates as classified by pclass, gender, age etc. other facts relating to composition of passenger with reference to port of embarkment etc. are also analyzed by using visual techniques in tableau.

# INTENDED AUDIENCE OF THE REPORT

The report and project are intended for following set of users-

* **Movie Geeks**-Who are touched by the movie and wants to find out more about the voyage
* **Historians**-Doing research on ship voyages in past.
* **Cruise** **Owners**-Can use the information for formulating rules for evacuation.
* **Insurance** **Companies**-Can formulate different insurance policies according to categories like gender, pclass, age etc.
* **Anyone** **touched** **by** **/** **interested** **in the titanic story**-Can read the report to satisfy one’s inquisitiveness by reading the report.

# SOME FACTS ABOUT TITANIC

1. Titanic was a British passenger ship operated by White star line.
2. In its maiden voyage only, it sank after striking with an iceberg in north Atlantic Ocean. It was travelling from Southampton UK to New York US.
3. Of estimated 2224 passengers and crew members only 1500 survived, making it one of the deadliest ship tragedies.
4. At its time it was the largest ship afloat.
5. It was under the command of Captain Edward smith, who also went down with the ship.
6. It carried one of the richest persons of that time and large number of immigrants seeking better life in United States.
7. The safety measurements adopted at titanic were inadequate as found by later studies.

# METADATA

Data is taken from [kaggle.com](http://kaggle.com) as attached in the form of object here.



The data set contains 12 variables and 891 total observations. A brief description of all the variables is given

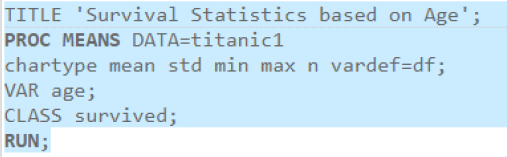
|  |  |
| --- | --- |
| **Variables** | **Description** |
| PassengerId | Number of passengers that were on the ship. |
| Survival | Number of people that survived on the ship are denoted by 1 and those who didn't survive are denoted by 0. |
| Pclass | Class of ticket denoted by 1, 2 and 3  1 = 1st (upper)  2 = 2nd (middle)  3 = 3rd (lower) |
| Name | Full name of passenger |
| Sex | Gender of passenger |
| Age | Age of the passenger in years. |
| Ticket | Ticket number bought by the passenger |
| Fare | Cost of travel. |
| Cabin | Cabin number of the passenger. |
| Embarked | Port of departure denoted by C, Q and S.  C = Cherbourg  Q = Queenstown  S = Southampton |

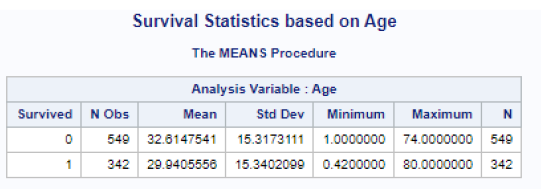
# OBJECTIVE OF THE PROJECT

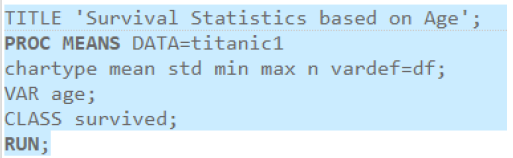
The data set is analyzed using tableau to find interesting insights about the infamous voyage using tableau for visualization, and we have also used SAS for explaining a small section of how exploratory analysis of data in SAS works using this data set. The main focus of the report is to find about the survivability of a passenger based on economic status, sex, age etc.

# DESCRIPTIVE DATA ANALYSIS using SAS

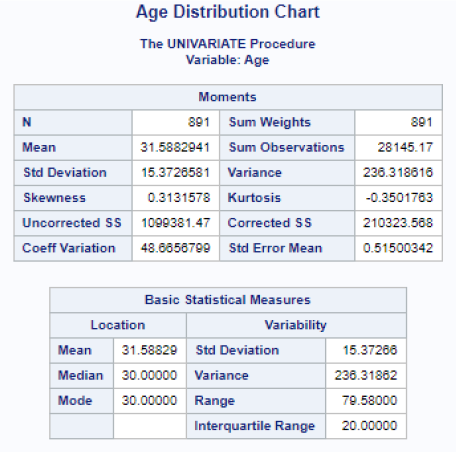
This analysis is done using SAS, which involves descriptive statistics





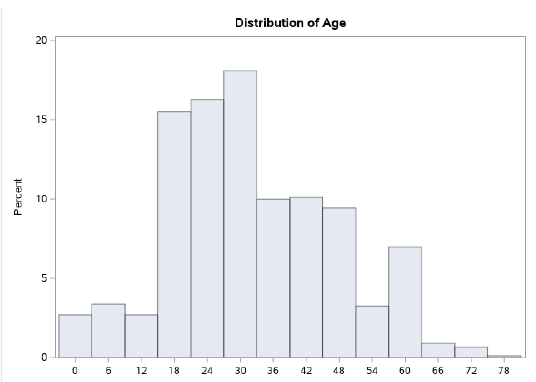


Table

Description automatically generated

A picture containing table

Description automatically generated



# Icon Description automatically generatedChart, line chart Description automatically generated**DATA VISUALISATION**

**Building a Titanic Story**: The data set in the excel file is connected to the tableau and the dependent variable that is the survived is put under row and independent variables like (age, pclass) are put under column. The resultant figure is used for visualization and understanding data.

ROUTE OF TITANIC

Titanic started its journey from Southampton UK, then to Cherbourg France and Queenstown Ireland. It got hit with giant iceberg near Newfound Island (Canada).

**Chart, line chart

Description automatically generated**

# EXPLORATORY DATA ANALYSIS or FINDINGS

Some of the interesting findings of the report are-

1. **Passenger survival** -Only 38.38% of the passengers survived. This figure signifies the tragedy related to the voyage. This means a huge number of people onboard lost their life, but how and Chart, pie chart

   Description automatically generatedwhy?
2. **Composition of passengers boarded on Titanic-**

* Most of the passengers were males. Females were only 35% whereas males were 65%
* Majority of the passengers were in the age group from 20-40 years. Only 7% of the passengers were above 60 years of age.
* More than 70% of the passengers boarded the voyage from Southampton
* 55% of the passengers were from class-3.24% of the passengers wereChart, bar chart

  Description automatically generated from class 1

Logo, company name

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

1. **Money increases the survival rate-**

* Of the 216 people travelling in the class -1, 136 survived and 80 perished.
* Similarly for class -2, 87 people survived and 97 perished.
* But for class-3 from 491 people travelling only 119 survived and 372 people died.
* The difference in survival rates for three p-class are whopping. It points out to the policy of saving lives of rich in priority to those not economically sound.

1. Chart, bar chart, box and whisker chart

   Description automatically generated**Survivors by Passenger category-**

Men had least survival rate whereas women and children had higher survival rates. This is in line with a code of conduct made in 1852 where lives of men and women were to be saved first in life threatening situations when survival resources such as lifeboats were limited.

The low survival rate of men can also be attributed to the fact that they were higher in percentage as compared to women and children.

Chart, bar chart

Description automatically generated

1. **Survival by age-**

Children aged between 1-15 had the maximum number survival rate and then survival rate for age group 15-45 can be seen low. For other ages, the survivability was low as seen by black Chart, bar chart

Description automatically generatedcolor.

1. **Passengers accompanied by family-Their survival rate**

Most of the passengers were accompanied by their children, parents, siblings, and spouse. Approximately 20% of the passengers had their children or parents accompanying them and about 30% of the passengers had their spouses or siblings on board with them

Survival rates for accompanied and unaccompanied passengers approximately corresponds to their percentage in total number of passengers. Therefore, being accompanied did not increase or decrease the survival rates.

Chart, bubble chart

Description automatically generated

1. **Survival and port of boarding-**

**Southampton** passengers had most survival rate as seen by the heat map high intensity. This can be because maximum people boarded from Southampton, and it had a larger percentage of affluent passengers (have better survival) than Cherbourg and Queenstown

Chart, treemap chart

Description automatically generated

1. **Port of embarkment and average prices**

Cherbourg (France) had the highest average fare whereas Queenstown (Ireland) had the lowest fare. Fare difference might be due to distance or class of passengers.

**Chart

Description automatically generated**

# LIMITATIONS OF THE DATASET

* The data set used for project is although helpful in finding many interesting aspects about the voyage, leave some questions unanswered such as the safety measures in the ship as it was the biggest and most expensive ship of that time.
* Data about lifeboats is unavailable, capacity of lifeboats and were all the lifeboats used is not present.
* Titanic is most remembered for the tragedy that stuck it into the ocean and took many lives and hence information about lifeboats and its usage is of utmost importance and should have been included.
* The data about crew members is missing. The information about crew members would have provided an insight on whether they had better survival chances or was there rescue prioritized.
* Without these vitals a complete picture of what transpired through Titanic’s voyage cannot be visualized completely.

# IMPORTANT INFORMATION FROM OUTSIDE RESOURCES

1. **Flaws in Titanic’s safety measures-**

* It had watertight compartments but the walls separating it were few feet apart, that means water could pop from one apartment to another if the ship had to pitch forward.
* The most fatal security lapse was that it carried only 20 lifeboats that could accommodate only 1100 people whereas there were 3300 people on board (Including the crew)
* Wrong assumptions of Captain Edward-He thought even after repeated warnings, he thought travelling through ice field at speed of 22.5 knots (high speed) would be safe. But his faulty assumption proved to be fatal for himself and the passengers too.
* No drills with lifeboats- Crew has insufficient training for lowering and using lifeboat in case of emergency. No safety drill for the passengers was conducted.
* Lifeboats were lowered after an hour when iceberg struck ship
* 472 -is the number of unused lifeboats seats.
* Only 18 lifeboats could be lowered.2 were not launched, and they simply floated away.

1. **Rescue-**

RMS Carpathia, which was 3 hours away came to the rescue of stranded passengers. Of 3000 people on board only 700 survived.30,000 people gathered in New York city to welcome the survivors.

Carpathia could lift only 13 lifeboats. Rest of the lifeboats could not be rescued.

1. **Crew of Titanic-**

* Titanic had about 600 crew members.
* About 24% of the crew members survived. Female crew had a survival rate of 87% whereas male crew had survival rate of about 22%.
* The code of conduct for evacuation was followed by crew also as females were prioritized.
* Crew members did not have an advantage in surviving as their survival rate is less than passenger survival rate.

1. **Aftermath of titanic-**

* Role of SS Californian-It was only few miles away from Titanic but did not responded to her distress calls or responded to her signal rockets. Later inquires found that captain of the ship acted improperly in failing to do so.
* Britain and USA governments constituted inquiry into the tragedy and found that inadequacy of lifeboats and life drills increased the intensity of tragedy.
* Changes into the safety code were made. And life drills were made a mandatory requirement for ship’s crew training.
* The tragic titanic voyage became a popular saga and a blockbuster movie depicting the unfortunate incident was made in 1997.

**Links**:

<https://www.anesi.com/titanic.htm>

<https://titanicfacts.net/titanic-survivors/>

<https://www.shiftcomm.com/insights/never-let-go-titanic-survival-101/>

<https://www.history.com/topics/early-20th-century-us/titanic>

<https://www.insider.com/photos-titanic-passengers-getting-rescued-by-carpathia-2020-4>

<https://titanicfacts.net/titanic-survivors/>

# RECOMMENDATIONS

* The data set used in this project is usual for analyzing various historic aspects of the infamous voyage, but for more curious audience information about lifeboats and crew members can be added.
* As can be inferred from the above analysis, in titanic money did increased the survival rate. This seems unethical and present cruise ships should make a policy of rescue without regard to one’s economic status.
* **Historians**- can infer from the report about social circumstances of that time. For instance, during society of that time wealth had considerable effect on treatment a person received. Also, the number of females travelling were considerably lower than man. This might be due to gender inequality prevailing at that time.
* **Insurance** **Companies**-For providing insurance for voyages, the company should look at safety preparedness of the ship. For lower safety measures more premium can be charged.
* **Movie geeks** and generally inquisitive people can visualize one of their favorite movies from a different perceptive and corroborate the movie with findings of this report.