# PROJECT 1: EXPLORING TITANIC DATABASE *by* *Luqman Idlan*

# Step 1: Understanding the Business Context

## *What are these data for?*

Titanic database provides the information on the passengers’ survival, identity, sex, age, and economic status.

1. *Why do we need this database?*

This database will help answering some questions related to the Titanic shipwreck incident.

1. *Where are these data collected?*

The data is collected from Encyclopedia Titanica which is an online reference work that contains extensive and constantly updated information on RMS Titanic.

# Step 2: Understanding the Technical Context

1. *How are these data collected?*

The data were collected through manual inspection from the tragedy and recorded.

1. *Where are the sources of these data?*

The data is stored in Encyclopedia Titanica website.

1. *What are the systems that touch or use/modify these data*?

Machine learning systems or manual data change may modify these data.

1. *What are some of the error sources of this data?*

There is an error on the age dataset where the oldest person age (80) is not the age recorded during the disaster. It is the age of post survival of the person’s death.

1. *Is the data complete? Would there be missing pieces of data?*

The data is completed. However, there is constantly updated information on the website.

# Step 3: Understanding the Tables and Fields

1. *How many tables do we have?*

One table

1. *What are the tables? What are these tables representing?*

The table is the passengers table. The table represents the information such as name, survival, age, sex, tickets, and economic status.

1. *Is the data messy? And how?*

The data is well sorted according to their category.

1. *Should I clean the data first? Or ignore*

There are no null values on the data of interest. Therefore, no cleaning will be done.

# Step 4: Free Exploration

Research Questions and Answers with SQL

1. *In the movie, women are prioritised to get into the safe boats more than men. Is that true?*

First, we need to count for number of men who survived.



Text

Description automatically generated with low confidence

Figure : Number of Men who survived and SQL commands

Based on figure 1, we can see the number of men who survived the tragedy is 109. Then, we need to compare with the number of women who survived the tragedy.





Figure : Number of Women who survived and SQL commands

Based on figure 2, we can see the number of women who survived the tragedy is 233 which is more than men. Thus, we can conclude that during the actual incident, women are prioritised for the seats in the safety boats because more women survived than men.

1. *In the movie, the number of people with high economic status and survived are more than people with lower economic status and survived. Is that true?*

First, we need to count the number of people with economic status with value 1 and survived.



Chart

Description automatically generated

Figure : Number of economic status value 1 and SQL commands

Next, we need to count the number of people with economic status with value 2 and survived.



A picture containing text

Description automatically generated

Figure : Number of economic status value 2 and SQL commands

Next, we need to count the number of people with economic status with value 3 and survived



A picture containing diagram

Description automatically generated

Figure : Number of economic status value 3 and SQL commands

Based on figure 3, figure 4 and figure 5, we can conclude that people with class 1 of economic status survived the most. Therefore, it is true that people with higher economic status survived more than people with lower economic status in the titanic tragedy.

1. *In the movie, children are prioritised more for the safety boat than adults. Is it true?*

First, we get the number of children with the assumption age is below than 12.



A picture containing shape

Description automatically generated

Figure : Number of children on board and SQL commands

Next, we get the number of children on board and survive

A picture containing text

Description automatically generated

A picture containing shape

Description automatically generated

Figure : Number of children on board and survived and SQL commands

Based on figure 6 and figure 7, we can conclude that out of 20 children on board, 13 children survived. This means it is true that children are prioritised more than adults for the safety board as most of the children survived the tragedy.