Name: Humaira binti Shamshuddin

Project 1 – Exploring Titanic Database

Guiding questions for free exploration:

- In the movie, children, elderlies and females can get onboarded to rescue boat first, so
  - Are children and elderlies have a higher survival rate in this accident?
  - Are females more likely to survive in this incident?

- Are rich people have a higher survival rate because they can get onboard to the rescue boat sooner (like what is shown in the movie)?

Feel free to come up with other questions that you want to answer with the data!

1. The total number of passengers in Titanic group by age classification:

Query	Results

```
SELECT count(name) AS
total_passengers,
CASE
WHEN age < 18 THEN "children"
WHEN age >= 60 THEN "elderlies"
WHEN age >= 18 AND AGE < 60
THEN "adult"
END AS age_classification
FROM passengers
GROUP BY age_classification
```

total_passengers	age_classification
177	
608	adult
65	children
41	elderlies

2. Total number of passengers **survived the Titanic incident**, classified by **age**:

Query	Results

```
SELECT count(Survived) AS
total_survivors,
CASE
WHEN Age >= 60 THEN
"elderlies"
WHEN Age < 18 THEN "children"
WHEN Age >= 18 AND Age < 60
THEN "adults"
END AS survivor_age_class
FROM passengers
WHERE Survived = 1
GROUP BY survivor_age_class
```

total_survivors	survivor_age_class
52	
243	adults
35	children
12	elderlies

Question 1: Which age class has the highest survival rate?

Table 1 and Table 2 show the number of passengers on Titanic before and after the incident happened, classified by age.

Adults were those in between 18 to 59 years old, **Children** were those below 18 years old and **Elderlies** were those 60 years old and above. (blank space in age class indicates that there is no data about the passengers' age, hence no age class applicable to the group).

Both tables illustrates that:

- Less than half of adults survived. (40%)
- More than half children were survived. (53.8%)
- Only three out of ten elderlies were survived. (29.3%)

From these numbers, we can conclude that children have higher survival rate in comparison to the other age groups as rescue team will prioritize in rescuing the children in any emergency. The elderlies have the lowest rate of survival, which most probably due to their physical condition.

3. Total number of females and males on Titanic:

Query	Results
SELECT sex AS Gender, count(name) AS Total_passengers FROM passengers GROUP BY sex	Gender: female, Total_passengers: 314 Gender: male, Total_passengers: 577

4. Total number of females and males survived **after the Titanic incident**:

Query	Results
SELECT Sex AS Gender, count(Survived) AS Total_survivors FROM passengers WHERE Survived = 1 GROUP BY Sex	Gender: female, Total_survivors: 233 Gender: male, Total_survivors: 109

Question 2 : Does female has higher survival rate than male?

Table 3 and Table 4 indicate the number of passengers before and after the Titanic incident grouped by gender.

Both tables shows that:

- More than 2/3 of the total female passengers survived. (74.2%)
- Less than 1/5 of the total male passengers survived. (18.9%)

These numbers most probably because, in any emergency, women and children are often rescued first, followed by elders and adult men. Hence, we can conclude that women have higher survival rate than men in any emergency.

5. Total number of passengers based on Class:

Query	Results
SELECT Pclass AS Passengers_class, count(Pclass) AS Total_passengers FROM passengers GROUP BY Pclass	

Passengers_class	Total_passengers
1	216
2	184
3	491

6. Total number of passengers **survived the Titanic incident** based on **Class**:

Query	Results
SELECT Pclass AS Passengers_class, count(Pclass) AS Total_survivors FROM passengers WHERE Survived = 1 GROUP BY Pclass	

Passengers_class	Total_survivors
1	136
2	87
3	119

Question 3 : Do the passengers' Class status influence their survival rate?

Based on Table 5 and Table 6 above, we can see the total number of passengers classified by class before, and after the titanic incident.

Both tables depicted that:

- More than half of passengers from Class 1 were survived. (70%)
- Less than half of the passengers from Class 2 were survived. (47.3%)
- Less than a quarter of passengers from Class 3 were survived. (24.2%)

The inference that can be done from these data was that the passengers from a higher Class got the access to rescue faster than the passengers from the middle and lower Class. We can conclude that passengers from the higher class have higher survival rate in comparison to the lower class.

7. Total number of passengers in each Cabin class before the Titanic incident:

Query	Results
SELECT substring(Cabin,1,1) AS Cabin, count(Cabin) AS total_passengers FROM passengers GROUP BY substring(Cabin,1,1)	

Cabin	total_passengers
	0
A	15
B	47
C	59
D	33
E	32
F	13
G	4
T	1

8. Total number of survivors in each **Cabin** class **after the titanic incident**:

Query	Results
SELECT substring(Cabin,1,1) AS Cabin, count(Cabin) AS total_survivors FROM passengers WHERE Survived = 1 GROUP BY substring(Cabin,1,1)	

Cabin	total_survivors
	0
A	7
B	35
C	35
D	25
E	24
F	8
G	2

Question 4 : Which Cabin has the highest number of survivors?

Based on Table 7 and Table 8, we can prove that:

- Most survivors came from Cabin D, which has the highest survival rate at 75.8%
- Followed by Cabin E and Cabin B at 75% and 74.5% of survival rate respectively
- Cabin F and Cabin C have survival rate of 61.5% and 59.3% respectively
- Cabin G has 50% of survival rate, Cabin A has 46.7% and Cabin T has no survivor

(Better conclusion can be made if some of the Cabin details are not null.)