PROJECT 1:

DATA EXPLORATION

FOR TITANIC DATABASE

TABLE OF CONTENT

01

TITANIC DATABASE
BUSINESS CONTEXT

The background and context of the database

BUSINESS CONTEXT

- What is the cause of the dataset existed?
  - A tragic accident happens on 15 April 1912, where the RMS Titanic sunk in the North Atlantic Ocean, after it collided with an iceberg during its journey to New York City from Southampton.

- What is the used of this dataset?
  - The dataset provides the necessary information to predict the fate of the passengers on the Titanic.

- What is titanic data analysis?
  - The total samples are 891 or 40% of the actual number of passengers on board the Titanic.

02 TITANIC DATABASE
TECHNICAL CONTEXT

Technical aspects of the database

TECHNICAL CONTEXT

- Where are the source of the data?
- – The principal source for the data about the Titanic passengers is the Encyclopedia Titanica.

- What are the errors of this dataset?
- – The data contains empty data fields.

- How are the data collected?
- – The datasets are collected from a variety of investigators and researchers.

03 TITANIC DATABASE TABLES & FIELDS

Understanding the database using SQLite

THE TABLES AND AND FIELDS

The Titanic database have one table, called passengers. The columns in this table are:

PassengerId

Definition: Id of each passngers
Type: Integer

Survived

Definition: Survival
Type: Integer

Pclass

Definition: Ticket Class
Type: Integer

Name

Definition: Passenger's name
Type: Text

Sex

Definition: Gender
Type: Text

Age

Definition: Age in years
Type: Text

THE TABLES AND AND FIELDS

The Titanic database have one table, called passengers. The columns in this table are:

SibSp

Definition: Number of siblings/spouse aboard
Type: Integer

Parch

Definition: Number of parents aboard
Type: Integer

Ticket

Definition: Ticket number
Type: Text

Fare

Definition: Passenger fare
Type: Integer

Cabin

Definition: Cabin number
Type: Text

Embarked

Definition: Port of embarkation
Type: Text

04

TITANIC DATABASE
FREE EXPLORATION

The analysis and the conclusions made to the database

RESEARCH QUESTIONS

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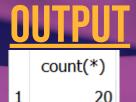
  Are children and elderlies have a high survival rate in this accident?
  Are females more likely to survive in this accident than males?

- Are rich people have a survival rate because they can get onboard to the rescue boat sooner?

- In the movie, the children, elderlies and females can get onboarded to the rescue boat first, so

  Are children and elderlies have a high survival rate in this accident?

SQL QUERY

```
SELECT count(*) FROM passengers
WHERE
    Survived = 1
    AND (Age <= 12 or Age >= 65)
```

OUTPUT

	count(*)
1	20

This SQL query count the number of passengers that are children and elderlies that survived the incident.

SQL QUERY

```
SELECT count(*) FROM passengers
WHERE
    Survived = 0
    AND (Age <= 12 or Age >= 65)
```

OUTPUT

	count(*)
1	27

This SQL query count the number of passengers that are children and elderlies who did not survived.

DATA EXPLORATION

- In the movie, the children, elderlies and females can get onboarded to the rescue boat first, so

  Are children and elderlies have a high survival rate in this accident?
  In this analysis, passengers with the age less than 13 years old is considered as a child, while passengers with the age of 65 or higher is considered as an elder.
  From the analysis, it can be concluded that, there is only 20 passengers in the group of children and elderlies that survived out of 47 passengers which makes the group have a survivability rate of 42.55%.

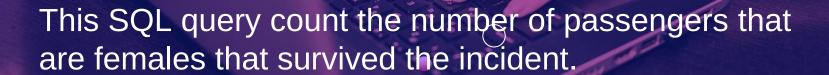Are females more likely to survive in this accident than males?

SQL QUERY

```sql
SELECT count(*) as survived FROM passengers
WHERE
    Survived = 1
    AND Sex = 'female'
```

OUTPUT

	survived
1	233

This SQL query count the number of passengers that are females that survived the incident.

```
1    SELECT count(*) as not_survived FROM passengers
2    WHERE
3        Survived = 0
4        AND Sex = 'female'
```

	not_survived
1	81

This SQL query count the number of passengers that are females who did not survived.

Are females more likely to survive in this accident than males?

SQL QUERY

```
SELECT count(*)as survived FROM passengers
WHERE
    Survived = 1
    AND Sex = 'male'
```

OUTPUT

	survived
1	109

This SQL query count the number of passengers that are males that survived the incident.

```sql
SELECT count(*)as not_survived FROM passengers
WHERE
    Survived = 0
    AND Sex = 'male'
```

not_survived
468

This SQL query count the number of passengers that are males who did not survived.

RESEARCH QUESTIONS

- In the movie, the children, elderlies and females can get onboarded to the rescue boat first, so

Are females more likely to survive in this accident than males?
From the analysis, 233 out of 314 female passengers have survived the incident, which makes the survival rate of 74.2%. As for the males, 109 out of 577 passengers have survived the incident. From that, the survival rate is 18.9%. It can be concluded that the female passengers will be more likely to survive than the males.

- Are rich people have a survival rate because they can get onboard to the rescue boat sooner?

SQL QUERY

```sql
SELECT Pclass as 'ticket class',count(*)as 'survived' FROM passengers
WHERE
    Survived = 1
    AND Pclass = 1
UNION
SELECT Pclass as 'ticket class',count(*)as 'survived' FROM passengers
WHERE
    Survived = 1
    AND Pclass = 2
UNION
SELECT Pclass as 'ticket class',count(*)as 'survived' FROM passengers
WHERE
    Survived = 1
    AND Pclass = 3
```

OUTPUT

	ticket class	survived
1	1	136
2	2	87
3	3	119

This SQL query count the number of passengers from each ticket class that survived the incident.

- Are rich people have a survival rate because they can get onboard to the rescue boat sooner?

SQL QUERY

```
SELECT Pclass as 'ticket class',count(*)as 'not survived' FROM passengers
WHERE
    Survived = 0
    AND Pclass = 1
UNION
SELECT Pclass as 'ticket class',count(*)as 'not survived' FROM passengers
WHERE
    Survived = 0
    AND Pclass = 2
UNION
SELECT Pclass as 'ticket class',count(*)as 'not survived' FROM passengers
WHERE
    Survived = 0
    AND Pclass = 3
```

OUTPUT

	ticket class	not survived
1	1	80
2	2	97
3	3	372

This SQL query count the number of passengers from each ticket class that did not survive the incident.

RESEARCH QUESTIONS

- Are rich people have a survival rate because they can get onboard to the rescue boat sooner?
Based on the analysis, the first-class passengers that survived the incident is 136 out of 216. As for the second-class, there are 87 out of 184 passengers that survived. As for the third-class, there are 119 out of 491 passengers that survived the incident. From that, the survival for the 1st class, 2nd class, and 3rd class are 62.96%, 47.28%, and 24.24% respectively.

CONCLUSION: ANALYSIS RESULTS

VARIABLES	TOTAL PASSENGERS		SURVIVAL RATE (%)
Are children and elderlies have a high survival rate in this accident?	47		42.55
Are females more likely to survive in this accident than males? (F-female/m-male)	F: 314	M: 577	F: 74.2 M: 18.9
Are rich people have a survival rate because they can get onboard to the rescue boat sooner?	1st: 216 2nd: 184 3rd: 491		1st: 62.96 2nd: 47.28 3rd: 24.24

CONCLUSION

In conclusion, the females are more likely to survive the incident due to the high survival rate and has the highest compared to the other variables. Therefore, the movie might be true that the females get onboard the rescue boat first and same goes for the 1st class passengers since they have higher survival rate compared to the other classes passengers. However, the survival rate for the children and elderlies are quiet low which is below 50%. This can be concluded, that they were not prioritized this group to be rescued first.