PROJECT 1 – EXPLORING TITANIC DATABASE

Objective: To identify the survival rate based on age group, gender, socioeconomic status and etc..

Exploration Question:

1. What is the total number of passengers in the Titanic and their overall survival rate?
2. Do children and elderlies have a higher survival rate in this accident?
3. What is the survival rate of adult from each location (Assuming they are born and live in the location where they embarked Titanic)
4. Are females more likely to survive in this incident?
5. 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)?
6. Based on Question 4&5, is there a higher survival rate if the passenger is of specific gender and economic status?

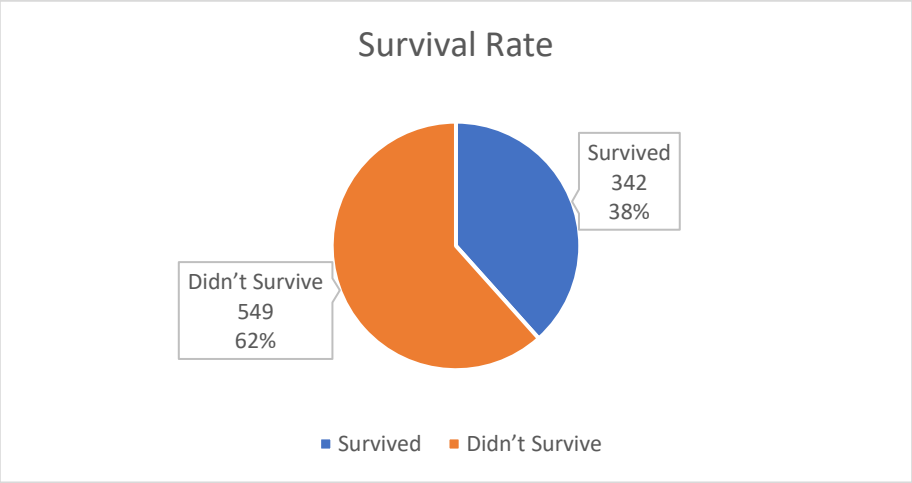
Discussion

1. What is the total number of passengers in the Titanic and their overall survival rate?

SQL Step:

```
SELECT Survived, COUNT(*) FROM passengers GROUP BY Survived
```

Result:



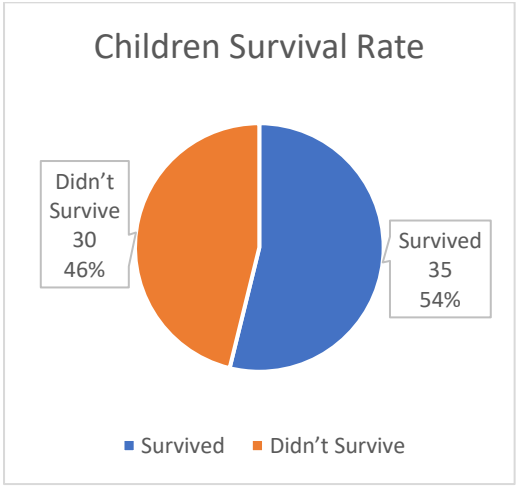
Out of 891 passengers in the Titanic, 38% of the passengers survived.

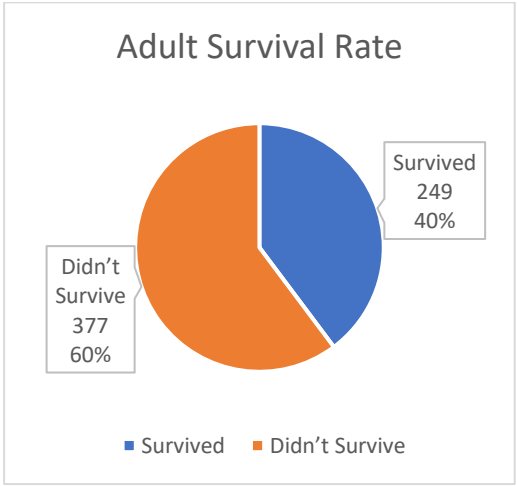
2. Do children and elderlies have a higher survival rate in this accident?

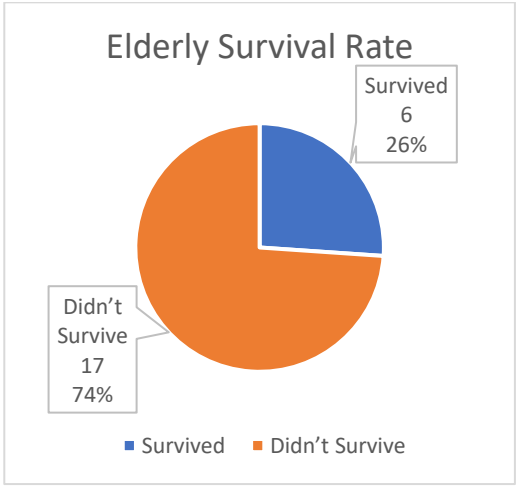
SQL Step:

```
WITH passengers_age_group AS (
        SELECT *,
                CASE
                        WHEN age < 18 THEN "Children"
                        WHEN age > 65 THEN "Elderly"
                        ELSE "Adult"
                END AS age_group
        FROM passengers WHERE age is not null
)
SELECT age_group, Survived, COUNT(*) FROM passengers_age_group GROUP BY age_group, Survived
```

Result:







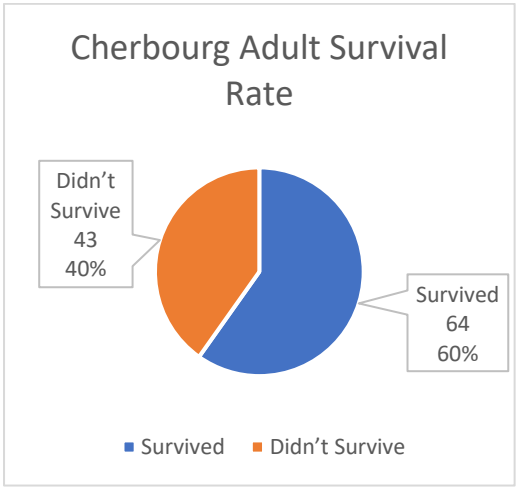
In overall, children have the highest survival rate, followed by adult and elderly

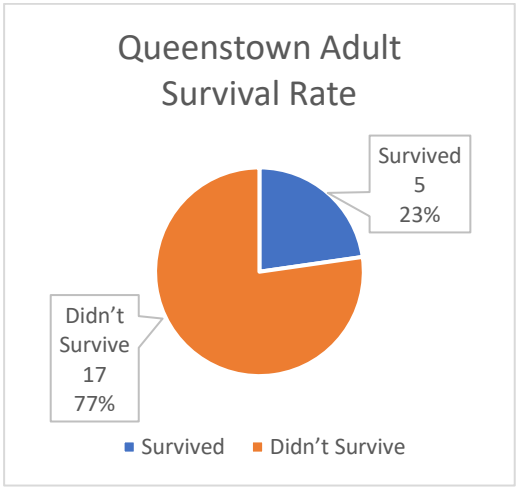
3. What is the survival rate of adult from each location (Assuming they are born and live in the location where they embarked Titanic)

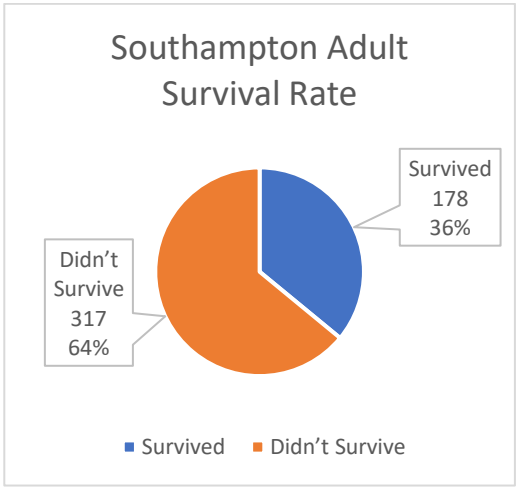
SQL Step:

```
WITH passengers_age_group AS (
        SELECT *,
                CASE
                        WHEN age < 18 THEN "Children"
                        WHEN age > 65 THEN "Elderly"
                        ELSE "Adult"
                END AS age_group
        FROM passengers WHERe age is not null
)
SELECT Embarked, Survived, COUNT(*) FROM passengers_age_group WHERE age_group = 'Adult' AND Embarked is not NULL GROUP BY Embarked, Survived
```

Result:







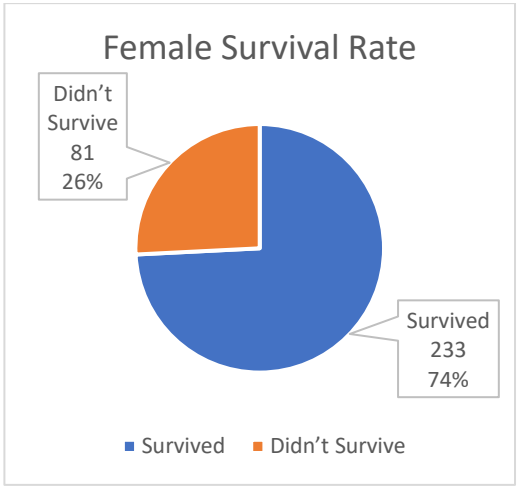
Adult from Cherbourg have the highest survival rate

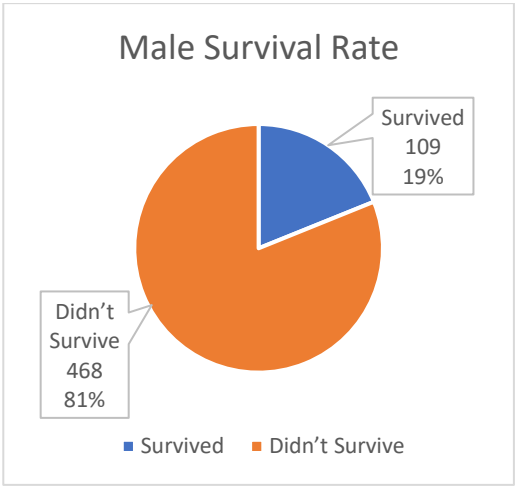
4. Are females more likely to survive in this incident?

SQL Step:

```
SELECT Sex, Survived, COUNT(*) FROM passengers GROUP BY Sex, Survived
```

Result:





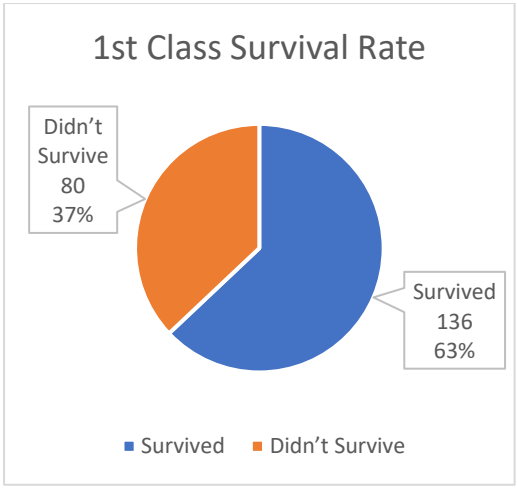
Females have a very high survival rate as compared to male

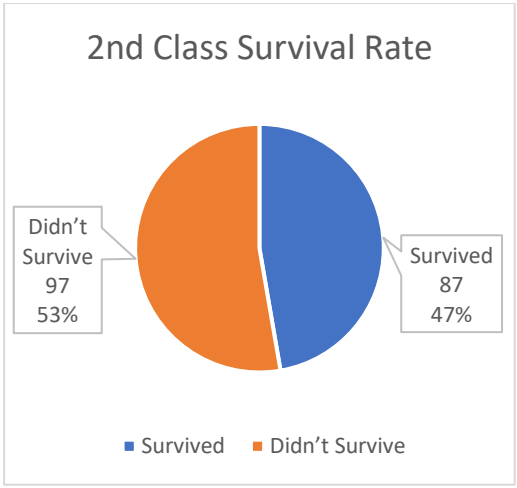
5. Are rich people have a higher survival rate because they can get onboard to the rescue boat sooner (like what is shown in the movie)?

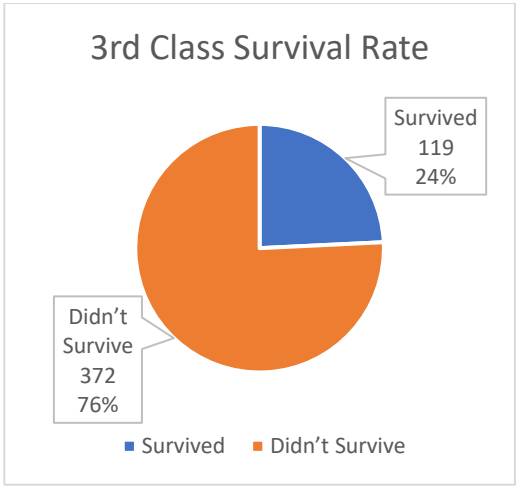
SQL Step:

```
SELECT Pclass, Survived, COUNT(*) FROM passengers GROUP BY Pclass, Survived
```

Result:





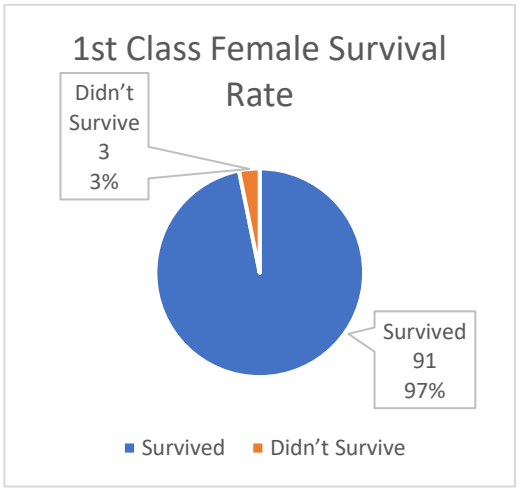


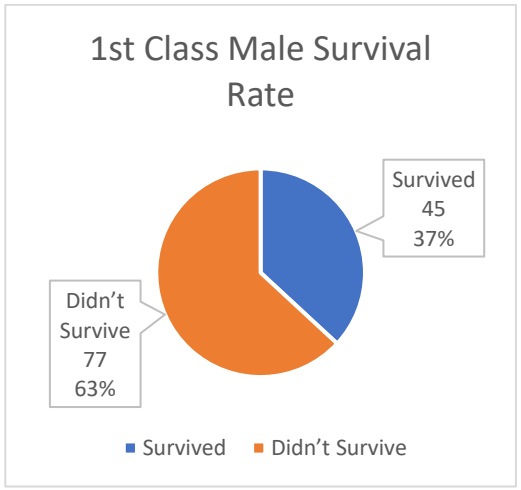
The rich have a higher survival chance of more than 50% compared to the poor.

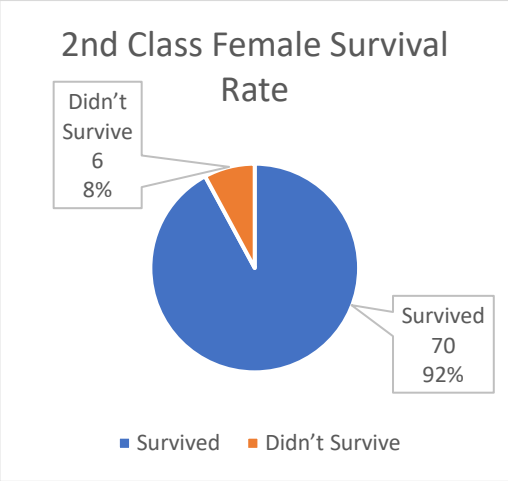
6. Based on Question 4&5, is there a higher survival rate if the passenger is of specific gender and economic status?

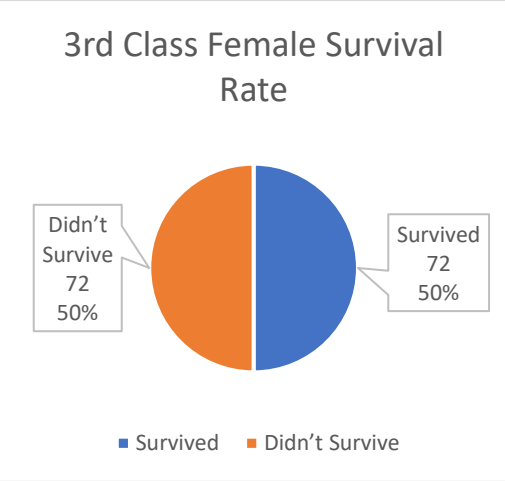
SQL Step:

```
SELECT Sex, Survived, COUNT(*) FROM passengers WHERE Pclass = 1 GROUP BY Sex, Survived
SELECT Pclass, Survived, COUNT(*) FROM passengers WHERE Sex = "female" GROUP BY Pclass, Survived
```









As shown from the data, if the passenger is a female and rich, there is a very high chance that she will survive from the crash as comparing to rich male and the poor female passenger.

Conclusion

There is a total of 891 passengers in the titanic, but there is only 342 (38%) passengers survived the crash. As compared to adult and elderly, children have the highest survival chance. Also, female also have a higher survival chance as compared to male. This might be due to children and female are given priority to board the rescue board first before adult and elderly. Similarly to what is shown in the movie, as the rich can get onboard to the rescue boat sooner, there have a higher survival rate compared to the poor.

A follow up question is raised to identify on the relationship between survival rate, gender and economic status. Seeing from the data, rich female have a higher survival rate compared to rich male and poor female. However, 2nd class female passengers have a higher survival rate compared 1st class male. In conclusion, female was given highest priority to get onboard to the rescue board as they have the highest survival rate.