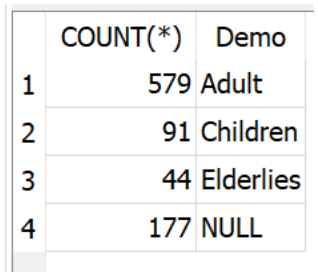
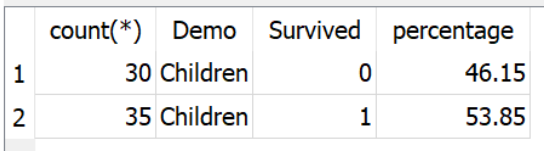
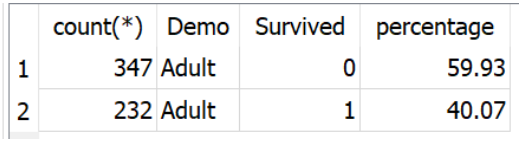
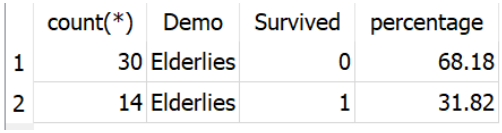
DATA EXPLORATION FOR TITANIC DATASET WITH SQL

1. **Are children and elderlies have a higher survival rate in this accident?**

   Children aged 18 years old and below have the highest survival rate at 53.85% while elderlies aged higher than 60 years old actually had the lowest survival rate at 31.82%.

Steps	SQL Code	Output
First, segregate age group from dataset into "Demographics"	WITH Demographics AS (SELECT *, CASE WHEN Age BETWEEN 0 AND 18 then "Children" WHEN Age > 59 then "Elderlies" WHEN Age > 18 AND Age < 60 then "Adult" Else "NULL" END AS Demo FROM passengers) SELECT COUNT(*), Demo FROM Demographics GROUP BY Demo	
Calculate the percentage survival rate of children	WITH Demographics AS (SELECT *, CASE WHEN Age BETWEEN 0 AND 17 then "Children" WHEN Age > 59 then "Elderlies" WHEN Age > 18 AND Age < 60 then "Adult" Else "NULL" END AS Demo FROM passengers) SELECT count(*), Demo, Survived, round(count(*) * 100.0 /(SELECT count(*) FROM (SELECT * FROM Demographics WHERE Demo = 'Children')),2) AS percentage FROM Demographics GROUP BY Demo, Survived Having Demo = 'Children'	

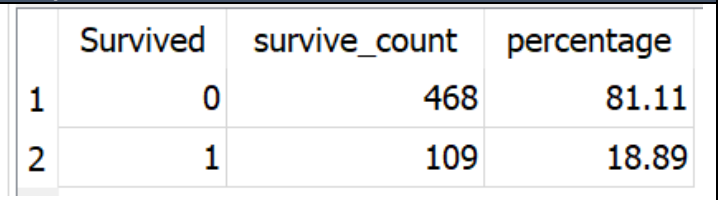
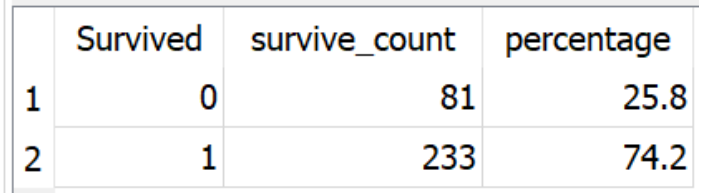
Steps	SQL Code	Output
Calculate the percentage survival rate of adults	WITH Demographics AS (SELECT *, CASE WHEN Age BETWEEN 0 AND 18 then "Children" WHEN Age > 59 then "Elderlies" WHEN Age > 18 AND Age < 60 then "Adult" Else "NULL" END AS Demo FROM passengers) SELECT count(*), Demo, Survived, round(count(*) * 100.0 /(SELECT count(*) FROM (SELECT * FROM Demographics WHERE Demo = 'Adult')),2) AS percentage FROM Demographics GROUP BY Demo, Survived Having Demo = 'Adult'	
Calculate the percentage survival rate of elderlies	WITH Demographics AS (SELECT *, CASE WHEN Age BETWEEN 0 AND 17 then "Children" WHEN Age > 59 then "Elderlies" WHEN Age > 18 AND Age < 60 then "Adult" Else "NULL" END AS Demo FROM passengers) SELECT count(*), Demo, Survived, round(count(*) * 100.0 /(SELECT count(*) FROM (SELECT * FROM Demographics WHERE Demo = 'Elderlies')),2) AS percentage FROM Demographics GROUP BY Demo, Survived Having Demo = 'Elderlies'	

	count(*)	Demo	Survived	percentage
1	347	Adult	0	59.93
2	232	Adult	1	40.07

	count(*)	Demo	Survived	percentage
1	30	Elderlies	0	68.18
2	14	Elderlies	1	31.82

2. **Are female more likely to survive in this accident?**

Yes, female passengers have a significantly higher survival rate than male passengers (74.2% vs 18.89%)

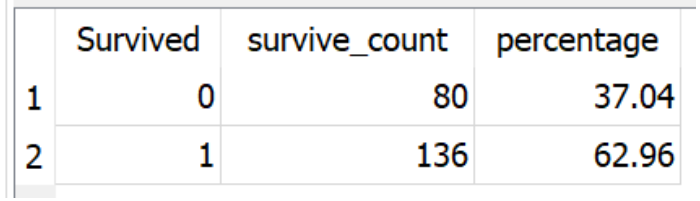
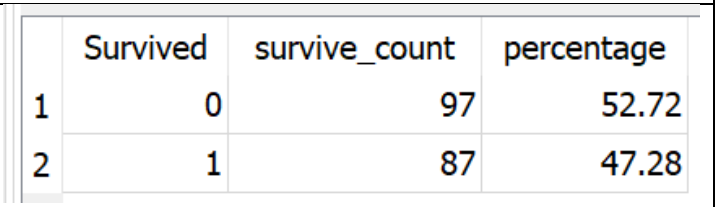
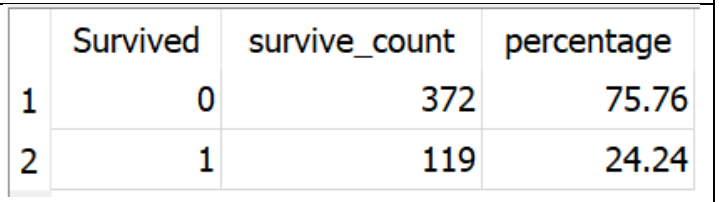
Steps	SQL Code	Output
Calculate the percentage survival rate of male passengers	SELECT Survived, count(*) AS survive_count, round(count(*) * 100.0 / (SELECT count(*) FROM passengers WHERE Sex = 'male'),2) AS percentage FROM passengers WHERE Sex = 'male' GROUP BY Survived	
Calculate the percentage survival rate of female passengers	SELECT Survived, count(*) AS survive_count, round(count(*) * 100.0 / (SELECT count(*) FROM passengers WHERE Sex = 'female'),2) AS percentage FROM passengers WHERE Sex = 'female' GROUP BY Survived	

	Survived	survive_count	percentage
1	0	468	81.11
2	1	109	18.89

	Survived	survive_count	percentage
1	0	81	25.8
2	1	233	74.2

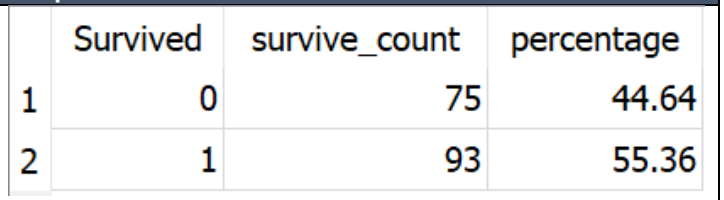
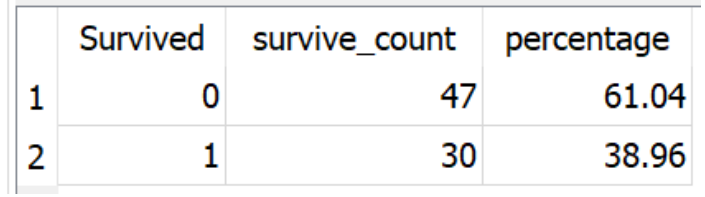
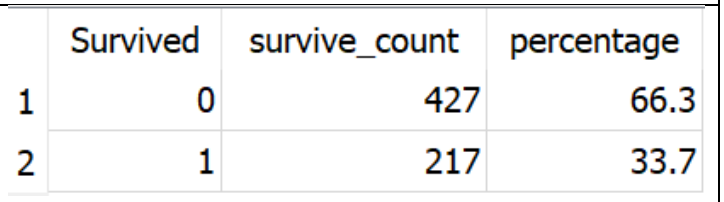
3. **Are rich people have a higher survival rate because they can get onboard to the rescue boat sooner (like what is shown in the movie)?**

   Yes, upper class passengers did have the highest survival rate at 62.96% followed by middle class passengers at 47.28% and lower class passengers had the lowest survival rate at 24.24%

Steps	SQL Code	Output
Calculate the percentage survival rate of upper class passengers	SELECT Survived, count(*) AS survive_count, round(count(*) * 100.0 / (SELECT count(*) FROM passengers WHERE Pclass = '1'),2) AS percentage FROM passengers WHERE Pclass = '1' GROUP BY Survived	
Calculate the percentage survival rate of middle class passengers	SELECT Survived, count(*) AS survive_count, round(count(*) * 100.0 / (SELECT count(*) FROM passengers WHERE Pclass = '2'),2) AS percentage FROM passengers WHERE Pclass = '2' GROUP BY Survived	
Calculate the percentage survival rate of lower class passengers	SELECT Survived, count(*) AS survive_count, round(count(*) * 100.0 / (SELECT count(*) FROM passengers WHERE Pclass = '3'),2) AS percentage FROM passengers WHERE Pclass = '3' GROUP BY Survived	

4. **Which embarkation port has the lowest survival rate?**

Passengers who embarked from Southampton had the lowest survival rate at 33.7% while passengers who embarked from Cherbourg had the highest survival rate at 55.36%

Steps	SQL Code	Output
Calculate the percentage survival rate of passengers from Cherbourg	SELECT Survived, count(*) AS survive_count, round(count(*) * 100.0 / (SELECT count(*) FROM passengers WHERE Embarked = 'C'),2) AS percentage FROM passengers WHERE Embarked = 'C' GROUP BY Survived	
Calculate the percentage survival rate of passengers from Queenstown	SELECT Survived, count(*) AS survive_count, round(count(*) * 100.0 / (SELECT count(*) FROM passengers WHERE Embarked = 'Q'),2) AS percentage FROM passengers WHERE Embarked = 'Q' GROUP BY Survived	
Calculate the percentage survival rate of passengers from Southampton	SELECT Survived, count(*) AS survive_count, round(count(*) * 100.0 / (SELECT count(*) FROM passengers WHERE Embarked = 'S'),2) AS percentage FROM passengers WHERE Embarked = 'S' GROUP BY Survived	

	Survived	survive_count	percentage
1	0	75	44.64
2	1	93	55.36

	Survived	survive_count	percentage
1	0	47	61.04
2	1	30	38.96

	Survived	survive_count	percentage
1	0	427	66.3
2	1	217	33.7