

Uncovering the Titanic: A Data-Driven Analysis Using SQL



Overview:



Overview:

This project presents a data-driven analysis of the Titanic disaster using SQL. By exploring passenger data such as age, gender, class, fare, and survival status, we aim to uncover key insights and patterns that influenced survival rates. The analysis is conducted using SQL queries on the Titanic dataset, demonstrating how structured data can be used to extract meaningful information.

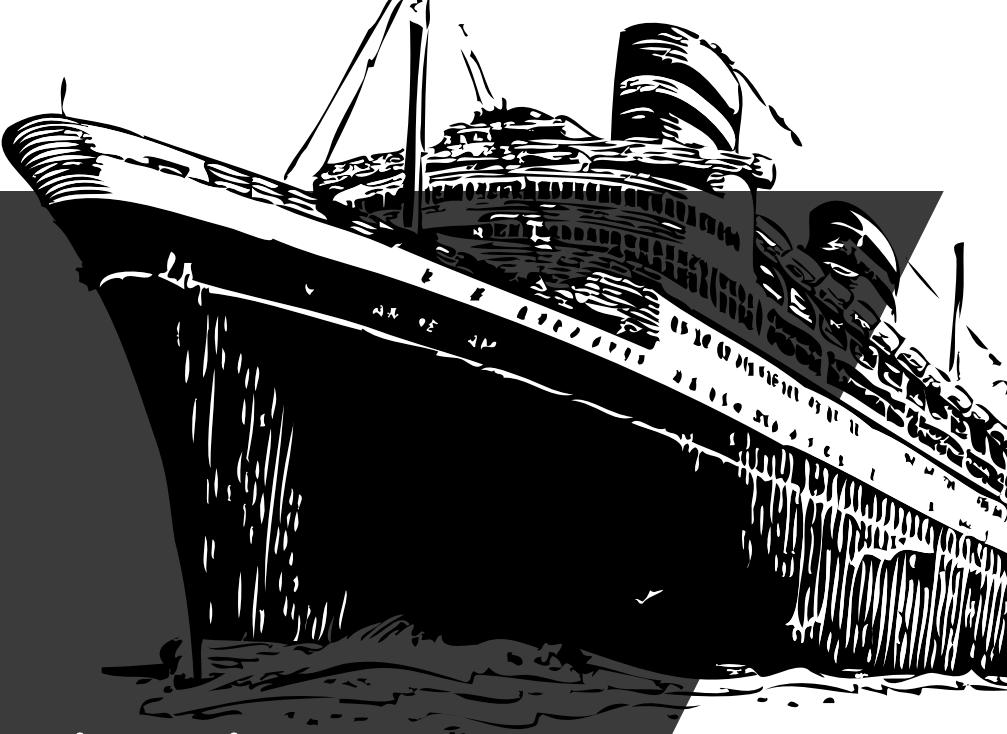


Objective:



Objective:

The primary objective of this project is to explore the Titanic dataset using SQL to uncover meaningful insights about the passengers and the factors that influenced their survival. This analysis aims to understand the distribution of passengers based on age, gender, class, and embarkation point, and to identify key patterns related to survival outcomes. By writing and executing various SQL queries, the project enhances data analysis skills such as filtering, grouping, aggregating, and joining tables. Ultimately, the goal is to present a data-driven perspective on the Titanic tragedy through clear interpretations and visualizations.



Questions:



1. How many passengers were onboard the Titanic?

Query:

```
SELECT COUNT(*) AS total_passengers FROM titanic_data;
```

Output:

	total_passengers	bigint
1		891

2. How many passengers survived?

Query:

```
SELECT COUNT(*) AS survivors FROM titanic_data WHERE Survived = 1;  
SELECT Sex,  
       COUNT(*) AS total,  
       SUM(CASE WHEN Survived = 1 THEN 1 ELSE 0 END) AS survived,  
       ROUND(100.0 * SUM(CASE WHEN Survived = 1 THEN 1 ELSE 0 END)/COUNT(*), 2) AS survival_rate  
FROM titanic_data  
GROUP BY Sex;
```

Output:

	survivors	bigint
1		342

3. What is the survival rate by gender?

Query:

```
SELECT Embarked, COUNT(*) AS total_passengers  
FROM titanic_data  
GROUP BY Embarked;
```

Output:

	sex	total	survived	survival_rate
	text	bigint	bigint	numeric
1	female	314	233	74.20
2	male	577	109	18.89

4. How many passengers embarked from each port?

Query:

```
SELECT Name, Age  
FROM titanic_data  
WHERE Survived = 1 AND Age IS NOT NULL  
ORDER BY Age ASC  
LIMIT 1;
```

Output:

	embarked	total_passengers
	text	bigint
1	[null]	2
2	Q	77
3	C	168
4	S	644



5. Who are the youngest and oldest passengers who survived?

Query:

```
SELECT Name, Age
FROM titanic_data
WHERE Survived = 1 AND Age IS NOT NULL
ORDER BY Age DESC
LIMIT 1;
```

Output:

	name text	age real
1	Barkworth, Mr. Algernon Henry Wilson	80

6. Which ticket had the highest fare, and who was the passenger?

Query:

```
SELECT Name, Ticket, Fare
FROM titanic_data
ORDER BY Fare DESC
LIMIT 1;
```

Output:

	name text	ticket text	fare real
1	Ward, Miss. Anna	PC 17755	512.3292



7. Find the survival rate of passengers who were alone vs. with family.

Query:

```
SELECT
    CASE
        WHEN SibSp = 0 AND Parch = 0 THEN 'Alone'
        ELSE 'With Family'
    END AS travel_type,
    COUNT(*) AS total,
    SUM(CASE WHEN Survived = 1 THEN 1 ELSE 0 END) AS survived,
    ROUND(100.0 * SUM(CASE WHEN Survived = 1 THEN 1 ELSE 0 END)/COUNT(*), 2) AS survival_rate
FROM titanic_data
GROUP BY travel_type;
```

Output:

	travel_type	total	survived	survival_rate
	text	bigint	bigint	numeric
1	With Family	354	179	50.56
2	Alone	537	163	30.35

8. Which cabin prefix (e.g., 'C', 'D', etc.) had the highest survival rate?

Query:

```
SELECT LEFT(Cabin, 1) AS cabin_prefix,
       COUNT(*) AS total,
       SUM(CASE WHEN Survived = 1 THEN 1 ELSE 0 END) AS survived,
       ROUND(100.0 * SUM(CASE WHEN Survived = 1 THEN 1 ELSE 0 END)/COUNT(*), 2) AS survival_rate
FROM titanic_data
WHERE Cabin IS NOT NULL
GROUP BY LEFT(Cabin, 1)
ORDER BY survival_rate DESC;
```

Output:

	cabin_prefix	total	survived	survival_rate
	text	bigint	bigint	numeric
1	D	33	25	75.76
2	E	32	24	75.00
3	B	47	35	74.47
4	F	13	8	61.54
5	C	59	35	59.32
6	G	4	2	50.00
7	A	15	7	46.67
8	T	1	0	0.00

Conclusions:



Conclusion:

This analysis using SQL provided key insights into the Titanic tragedy, revealing how factors like gender, class, and age impacted survival. It demonstrated the power of SQL in exploring real-world data and uncovering meaningful patterns.





Thank You