**Titanic Project SQL Code**

**Data Cleaning**

#UPDATE `your\_project.your\_dataset.train`

SET Age = (SELECT AVG(Age) FROM `your\_project.your\_dataset.train` WHERE Age IS NOT NULL)

WHERE Age IS NULL;

UPDATE `your\_project.your\_dataset.train`

SET Fare = (SELECT AVG(Fare) FROM `your\_project.your\_dataset.train` WHERE Fare IS NOT NULL)

WHERE Fare IS NULL;

UPDATE `your\_project.your\_dataset.train`

SET Embarked = 'S'

WHERE Embarked IS NULL;

**Overall Survival Rate**:

SELECT

ROUND(AVG(Survived) \* 100, 2) AS survival\_rate

FROM

`your\_project.your\_dataset.train`;

**Survival by Passenger Class**:

SELECT

Pclass,

ROUND(AVG(Survived) \* 100, 2) AS survival\_rate

FROM

`your\_project.your\_dataset.train`

GROUP BY

Pclass

ORDER BY

Pclass;

**Survival by Gender**:

SELECT

Sex,

ROUND(AVG(Survived) \* 100, 2) AS survival\_rate

FROM

`your\_project.your\_dataset.train`

GROUP BY

Sex;

**Fare vs. Survival**:

SELECT

Survived,

ROUND(AVG(Fare), 2) AS average\_fare

FROM

`your\_project.your\_dataset.train`

GROUP BY

Survived;

**Age vs. Survival**:

SELECT

Survived,

ROUND(AVG(Age), 2) AS average\_age

FROM

`your\_project.your\_dataset.train`

GROUP BY

Survived;

**Survival by Embarkation Port**:

SELECT

Embarked,

ROUND(AVG(Survived) \* 100, 2) AS survival\_rate

FROM

`your\_project.your\_dataset.train`

WHERE

Embarked IS NOT NULL

GROUP BY

Embarked;

**Family Size vs. Survival**:

SELECT

(SibSp + Parch) AS family\_size,

ROUND(AVG(Survived) \* 100, 2) AS survival\_rate

FROM

`your\_project.your\_dataset.train`

GROUP BY

family\_size

ORDER BY

family\_size;

**Top 10 Survivors by Fare Paid**:

SELECT

Name, Fare

FROM

`your\_project.your\_dataset.train`

WHERE

Survived = 1

ORDER BY

Fare DESC

LIMIT 10;