# **SQL Queries for Titanic Dataset Analysis**

## **Task 1: Handling Missing Values**

### 1. Checking Missing Values -- This query checks for missing values in Age, Fare, and Embarked columns.

SELECT COUNTIF(Age IS NULL) AS missing\_age,  
 COUNTIF(Fare IS NULL) AS missing\_fare,  
 COUNTIF(Embarked IS NULL) AS missing\_embarked   
FROM `titanic-data-analysis.titanic\_analysis.passengers`;

### 2. Median Age -- This query calculates the median age of passengers, excluding null values.

WITH age\_data AS (  
 SELECT Age  
 FROM `titanic-data-analysis.titanic\_analysis.passengers`  
 WHERE Age IS NOT NULL  
 ORDER BY Age  
)  
SELECT   
 APPROX\_COUNT\_DISTINCT(Age) AS total\_ages,   
 APPROX\_QUANTILES(Age, 3)[OFFSET(1)] AS median\_age  
FROM age\_data;

### 3. Median Fare -- This query calculates the median fare of passengers, excluding null values.

WITH fare\_data AS (  
 SELECT Fare  
 FROM `titanic-data-analysis.titanic\_analysis.passengers`  
 WHERE Fare IS NOT NULL  
 ORDER BY Fare  
)  
SELECT   
 APPROX\_COUNT\_DISTINCT(Fare) AS total\_fares,   
 APPROX\_QUANTILES(Fare, 3)[OFFSET(1)] AS median\_fare  
FROM fare\_data;

### 4. Create New Table & Update Null Values -- This query creates a new table 'cleaned\_passengers' and updates null values in Age, Fare, and Embarked columns.

CREATE OR REPLACE TABLE `titanic-data-analysis.titanic\_analysis.cleaned\_passengers` AS  
SELECT   
 PassengerId,  
 Survived,  
 Pclass,  
 Name,  
 Sex,  
 COALESCE(Age, 23.0) AS Age,  
 SibSp,  
 Parch,  
 Ticket,  
 COALESCE(ROUND(Fare, 1), 8.7) AS Fare,   
 COALESCE(Embarked, 'S') AS Embarked   
FROM   
 `titanic-data-analysis.titanic\_analysis.passengers`;

## **Task 2: Analyzing Survival Rates**

### 1. Overall Survival Rate -- This query calculates the overall survival rate percentage.

SELECT   
 ROUND(COUNTIF(Survived = 1) / COUNT(\*) \* 100, 2) AS survival\_rate\_percentage  
FROM   
 `titanic-data-analysis.titanic\_analysis.cleaned\_passengers`;

### 2. Survival by Passenger Class (Pclass) -- This query calculates survival rate percentage based on passenger class.

SELECT   
 Pclass,  
 ROUND(COUNTIF(Survived = 1) / COUNT(\*) \* 100, 2) AS survival\_rate\_percentage  
FROM   
 `titanic-data-analysis.titanic\_analysis.cleaned\_passengers`  
GROUP BY   
 Pclass  
ORDER BY   
 Pclass;

### 3. Survival by Gender (Sex) -- This query calculates survival rate percentage based on gender.

SELECT   
 Sex,  
 ROUND(COUNTIF(Survived = 1) / COUNT(\*) \* 100, 2) AS survival\_rate\_percentage  
FROM   
 `titanic-data-analysis.titanic\_analysis.cleaned\_passengers`  
GROUP BY   
 Sex  
ORDER BY   
 Sex;

### 4. Fare vs. Survival -- This query calculates the average fare paid by survivors and non-survivors.

SELECT   
 Survived,  
 ROUND(AVG(Fare), 2) AS avg\_fare  
FROM   
 `titanic-data-analysis.titanic\_analysis.cleaned\_passengers`  
GROUP BY   
 Survived  
ORDER BY   
 Survived;

### 5. Age vs. Survival -- This query calculates the average age of survivors and non-survivors.

SELECT   
 Survived,  
 ROUND(AVG(Age), 2) AS avg\_age  
FROM   
 `titanic-data-analysis.titanic\_analysis.cleaned\_passengers`  
GROUP BY   
 Survived  
ORDER BY   
 Survived;

### 6. Survival by Embarkation Port (Embarked) -- This query calculates survival rate percentage based on embarkation port.

SELECT   
 Embarked,  
 ROUND(COUNTIF(Survived = 1) / COUNT(\*) \* 100, 2) AS survival\_rate\_percentage  
FROM   
 `titanic-data-analysis.titanic\_analysis.cleaned\_passengers`  
GROUP BY   
 Embarked  
ORDER BY   
 Embarked;

### 7. Family Size vs. Survival (SibSp and Parch) -- This query calculates survival rate based on family size, combining SibSp and Parch.

SELECT   
 (SibSp + Parch) AS family\_size,  
 ROUND(AVG(Survived), 2) AS survival\_rate  
FROM   
 `titanic-data-analysis.titanic\_analysis.cleaned\_passengers`  
GROUP BY   
 family\_size  
ORDER BY   
 family\_size;

### 8. Top 10 Survivors by Fare Paid -- This query retrieves the top 10 survivors based on fare paid.

SELECT   
 Name,  
 Fare,  
 Pclass,  
 Age,  
 Sex  
FROM   
 `titanic-data-analysis.titanic\_analysis.cleaned\_passengers`  
WHERE   
 Survived = 1   
ORDER BY   
 Fare DESC   
LIMIT 10;