SQL:

QUESTION 1:

What will be the result of the query below?

SELECT * FROM runners WHERE id NOT IN (SELECT winner_id FROM races)

Explain your answer and also provide an alternative version of this query that will avoid the issue that it exposes.

Ans: The given query will not show any result because here winner_id from races is comparing to the id column from runners table. And winner id doesn't exist which matches the id column from runners table.

To avoid the issue, we should match id's from both the tables, then we will get the answer on which runner did not participate in race.

"**SELECT * FROM runners WHERE id NOT IN (SELECT id FROM races)**"

This give answer ID=5, NAME= Lisa romero.

QUESTION 2:

Given two tables, Write a query to fetch values in table test_a that are and not in test_b without using the NOT keyword.

Ans: Using left join because it will take all the records from left table i.e., test_a and only matched from test_b. But when we apply where condition, then result will only fetch the id's from test_a which are not present in test_b.

"SELECT a.id
FROM test_a a
LEFT JOIN test_b b ON a.id = b.id
WHERE b.id IS NULL;
"

Question: 3

Write a query to to get the list of users who took the a training lesson more than once in the same day, grouped by user and training lesson, each ordered from the most recent lesson date to oldest date.

Ans: I have selected user id and name, training lesson and date from training details matching id with users table and performed count() to get how many times the particular user has been attended the training on a same day. And implemented DESC to order by recent accessed training day.

“SELECT u.id,u.name, t.tid, tdate, COUNT(*) AS lesson_count
FROM training_details t
left join users u on u.id=t.id
GROUP BY u.id, tid, tdate
HAVING COUNT(*) > 1
ORDER BY tdate DESC;
”

Question: 4:

Ans: The manager_id and emp_name are selected from the employee table (represented by 'e'), and the average salary is calculated using the AVG() function and then join condition is used on the employee table which joins with itself to retrieve the manager details and where condition is used to filter the records where the manager_id is not NULL and used group by condition on manager_id and manager_name.

“SELECT e.manager_id, m.emp_name AS Manager, AVG(e.salary) AS Average_Salary_Under_Manager
FROM employee e
JOIN employee m ON e.manager_id = m.emp_id
where e.manager_id is not null
GROUP BY m.manager_id, m.emp_name;
”

STATISTICS

Question: 1

What is the meaning of six sigma in statistics? Give proper example

Ans: Six sigma is statistical measure standard deviation (σ) from the mean.
Example:

In the graph of a distribution, It says 68% of the data will fall under 1 sigma(standard deviation), 95% falls under 2 sigma, and 99.7% falls under 3 sigma.99.993% falls under 4 sigma.99.99999% falls under 5 sigma.
6th sigma cover the area of upper fence to the lower fence.
And whatever the data points in tails which doesn't cover, can be considered as outliers.

Question: 2

What type of data does not have a log-normal distribution or a Gaussian distribution? Give proper example

1. Skewed distribution
2. Uniform distribution
3. Discrete distribution
4. Categorical distribution
5. Exponential distribution
6. Poisson distribution

Example 1. distribution of numbers that show up on the top of a fair die after a large number of throws.
Example 2: data with outliers, or heavy tail
Example 3: Unlikely events occurring which is not fair

Question: 3

What is the meaning of the five-number summary in Statistics? Give proper example

Ans: A five-number summary is a set of five values that can be used to describe a large data set. The five values are:

- The maximum value
- The minimum value
- The lower quartile
- The upper quartile
- The median

Question: 4

What is correlation? Give an example with a dataset & graphical representation on jupyter Notebook

Ans: Correlation: How much data features are co related to each other. It can be between -1 to 1. -1 indicating strong negative correlation, 1 indicating strong positive correlation.
I have performed example on iris dataset, shared in repo.