SQL

Question-1

This will return an empty set because NOT IN condition contains any values that are null, then the outer query here will return an empty set.

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

Question-3

```
SELECT username, training_date
FROM users u
INNER JOIN training_details t ON(u.user_id = t.user_id)
GROUP BY username, tranning_date
HAVING COUNT(*) > 1
ORDER BY username, training_date DESC
```

Question 2-

```
Select id.* from test a

Left join test b on (testa.id = testb.id) where testb.id is null
```

Question-4

```
SELECT Manager_id, Manager, AVG(Salary) AS Average_Salary, COUNT(*) AS Under_Manager
FROM (
   SELECT Manager_id, Emp_name AS Manager, Salary
   FROM employee
   WHERE Manager_id IS NOT NULL
) AS managers
GROUP BY Manager_id;
```

STATISTICS

QUESTION 1

Six Sigma is a statistical approach used by companies to improve quality and reduce defects in their processes. It aims to achieve a level of performance where the process produces very few defects, ideally no more than 3.4 defects per million opportunities. For example, a manufacturing company may use Six Sigma to reduce the number of defective products they produce, leading to higher customer satisfaction and lower costs.

QUESTION 2

Non-Gaussian data refers to data that does not follow a normal (Gaussian) distribution or a log-normal distribution. Examples include data with exponential decay (like radioactive decay), uniform distribution (equal probability for all values in a range, like rolling a fair die), power-law distribution (where the frequency of occurrence decreases as a power of the value), and bimodal distribution (two distinct peaks).

QUESTION 3

The five-number summary in statistics gives a quick snapshot of the key characteristics of a dataset. It includes the minimum, maximum, and three key points in the middle: the first quartile (Q1), the median, and the third quartile (Q3).

For example:

- Minimum: The smallest value.
- Q1: The value below which 25% of the data falls.
- Median: The middle value.
- Q3: The value below which 75% of the data falls.
- Maximum: The largest value.

It helps us understand the spread and central tendency of the data quickly.

QUESTION 4

Correlation is a measure that tells us how two sets of data are related. If they move in the same direction, it's a positive correlation; if one goes up while the other goes down, it's a negative correlation. If there's no clear relationship, the correlation is close to zero.

For example, let's say we have data on students' study hours and their test scores. If students who study more tend to have higher scores, there's a positive correlation. We can represent this visually in a Jupyter Notebook using a scatter plot, where study hours are on the x-axis and test scores are on the y-axis. If the points form an upward trend, it indicates a positive correlation. If the points form a downward trend, it indicates a negative correlation. If there's no clear trend, it indicates no correlation.

DEEP LEARNING

QUESTION_1
a)

Below are the few steps required for implementing DL in a real-world application-

Problem Definition: Clearly define the problem you want to solve with Deep Learning (DL).
Data Preparation: Collect and preprocess high-quality data for training the DL model.
Model Selection: Choose an appropriate DL model architecture based on your problem and data.
Implementation: Implement the chosen DL model using a framework like TensorFlow or PyTorch.
Training: Train the DL model on the training data, optimizing hyperparameters as needed.
Evaluation: Evaluate the model's performance on a validation set and fine-tune as necessary.
Deployment: Deploy the trained model to a production environment for real-world use.
Monitoring and Maintenance: Continuously monitor the model's performance and update as needed.
Ethical and Legal Considerations: Consider ethical and legal implications, such as privacy and bias.
Iterate and Improve: Iterate on the model and application based on feedback and new requirements for continuous improvement.

b)

1. **Introducing Non-Linearity**: Imagine you're trying to represent data with straight lines, like plotting points on graph paper. But real-world data isn't always so simple. Sometimes, relationships between data points are curved or wavy. Activation functions help neural networks handle these curved patterns. They allow neurons to activate in a non-linear way, which lets the network learn and understand complex patterns in the data, like recognizing different shapes in images or understanding the meaning of words in text.

2. **Enabling Complex Mapping**: Activation functions help neural networks create intricate maps from inputs to outputs. Think of it like making a detailed map of a city with all its streets, buildings, and landmarks. Each activation function adds a layer of detail to the map, allowing the network to understand finer and finer distinctions in the data. Without activation functions, the network would only be able to create very simple maps, like drawing straight lines on a blank piece of paper. This would limit its ability to understand and process complex information.

In simple terms, activation functions help neural networks understand and process complex patterns in data by adding non-linearities and enabling detailed mappings from inputs to outputs. Without them, neural networks wouldn't be able to learn effectively from data with non-linear relationships, and their performance would be severely limited.