SQL

A1)

**SELECT** \* **FROM** runners **WHERE** id **NOT** **IN** (**SELECT** winner\_id **FROM** races) this query can give error as there is null values

An alternate approach can be

SELECT \* FROM runners WHERE id NOT IN (

SELECT winner\_id

FROM races

WHERE winner\_id IS NOT NULL)

Ans2)

SELECT \*

FROM table\_a

LEFT JOIN table\_a ON table\_a = table\_b

WHERE table\_b IS NULL;

We use an anti join to get all the rows from first table but excluding those common in second table

**Statistics**

**Q1)**

Six sigma is a statistical probability of failure , ie how far a process deviates from perfection. A process that operates at six sigma has a failure rate of only 0.00034%

**Q2)**

Distribution in which there is very high inequality tend to not have a normal distribution

Ex in India only 10% people earn above 25k monthly , and no. is even less for for those who earn 50k monthly this type of distributions are not normal distributuions

Another example can be rolling a die , this will be uniform distribution as all events have same probability of occurring

**Q3)**

The five-number summary is a set of descriptive statistics that provides a concise summary of the distribution of a dataset. It includes five key values: the minimum, first quartile (Q1), median (second quartile, Q2), third quartile (Q3), and the maximum. These values help convey essential information about the central tendency and spread of the data.

Here a breakdown of five-number summary:

1. Minimum : The smallest value
2. First Quartile(Q1): the value below which 25% of data falls
3. Median : median of the data ie the 50 percentile
4. Third Quartile : The value below which 75% of data falls
5. Maximum : the lagest value in dataset

We typically use boxplot to see a 5 point summary which also gives us information about outliers if any in the plot

**Q4**

Correlation depicts how much two variables are related to one another or how much they are proportional , ir change in one will also cause other to change either positively or nrgatively.it shows the linear relationship between the features

**Deep Learning**

Q1: a

There are many uses of deep learning in real world applications like

Fraud detection, spam classifiers , encoder decoder models, all the LLM’s use complex deep learning architecture , the advantage of deep learning over machine learning is deep learning can capture more complex patterns,info from a given data but also requires more better hardware to perform.

Deep learning is better than ml in real world use cases like NLP and computer vision

Q1: b

Activation Functions are used to add non linearity to the network

Without activation function our model will just be a linear model and adding layers would not improve performance of our model

We add a activation function to model because

1. It adds non linearity to data before it outputs to next layer
2. Helps model learn more complex patterns
3. Better model generalization
4. Less vanishing gradients and exploding gradient problems