Python

Question: 1

You have an input dictionary given,

input\_dict = {"abc":{"def":{"ghi":{"jkl":{"mno":{"pqr":{"stu":{"vwx":{"yz":"you are finally here !!!"}}}}}}}}}

Answer. 1 Python - > <https://github.com/prachikabra121/Inueron_Assessment/blob/main/Question1.ipynb>

Question: 2

Given an array of length ‘N’, where each element denotes the position of a stall. Now you have ‘N’ stalls and an integer ‘K’ which denotes the number of horses that are mad. To prevent the horses from hurting each other, you need to assign the horses to the stalls, such that the minimum distance between any two of them is as large as possible. Return the largest minimum distance.

array: 1,2,4,8,9 & k=3

O/P: 3

Answer. 2 Python - > <https://github.com/prachikabra121/Inueron_Assessment/blob/main/Question2.ipynb>

Question: 3

Mr. Karthiken works in a door mat manufacturing company. One day, he designed a new door mat with the following specifications:

a) Mat size must be N X M. (N is an odd natural number, and M is 3 times N.)

b) The design should have ‘WELCOME’ written in the center.

c) The design pattern should only use |, . and – characters.

Sample Design is given above image, Write a python code for this.

Answer. 3 Python - > <https://github.com/prachikabra121/Inueron_Assessment/blob/main/Question3.ipynb>

Question: 4

Given an array nums of n integers, return an array of all the unique quadruplets [nums[a], nums[b], nums[c], nums[d]] such that:

a) 0 <= a, b, c, d < n

b) a, b, c, and d are distinct.

c) nums[a] + nums[b] + nums[c] + nums[d] == target

Answer. 4 Python - > <https://github.com/prachikabra121/Inueron_Assessment/blob/main/Question4.ipynb>

SQL

**Question 1**

**Answer:**

The given SQL query is attempting to retrieve all rows from the "runners" table where the "id" is not present in the result of the subquery (SELECT winner\_id FROM races).

However, this query might not work as expected in cases where there are NULL values in the "winner\_id" column of the "races" table. If there are NULL values in the subquery result, the NOT IN clause will not match those NULL values, potentially leading to unexpected results.

To avoid this issue, you can use the NOT EXISTS .So the alternative solution for this would be :

SELECT \* FROM runners r

WHERE NOT EXISTS (

SELECT 1

FROM races

WHERE races.winner\_id = r.id

);

**Question 2**

**Answer:**

SELECT a.\*

FROM test\_a a

LEFT JOIN test\_b b ON a.id = b.id

WHERE b.id IS NULL;

**Question 3**

**Answer:**

SELECT u.username, td.training\_id, td.training\_date

FROM users u

JOIN training\_details td ON u.user\_id = td.user\_id

GROUP BY u.user\_id, td.training\_id, td.training\_date

HAVING COUNT(\*) > 1

ORDER BY td.training\_date DESC;

**Question 4**

**Answer:**

SELECT Manager\_Id,

Emp\_name AS Manager,

AVG(Salary) AS Average\_Salary\_Under\_Manager

FROM Employee

WHERE Emp\_Id IN (SELECT DISTINCT Manager\_Id FROM Employee)

GROUP BY Manager\_Id, Emp\_name

ORDER BY Manager\_Id;

**Statististics**

**Question: 1**

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

**Answer:**

Six Sigma is a statistical concept that originated from the manufacturing industry and has been widely adopted across various sectors. It is a methodology used to improve business processes and reduce defects or variations in products or services. The term "Six Sigma" refers to a statistical measure that quantifies how far a given process deviates from perfection.

In a Six Sigma process, the goal is to achieve fewer than 3.4 defects per million opportunities, which is an extremely high level of quality. The term "Sigma" represents the standard deviation, a measure of how much variation exists in a set of data.

Here's a breakdown of the sigma levels and their corresponding defect rates:

* Sigma Level 6:3.4 defects per million opportunities
* Sigma Level 5:233 defects per million opportunities
* Sigma Level 4:6,210 defects per million opportunities
* Sigma Level 3:66,807 defects per million opportunities
* Sigma Level 2:308,537 defects per million opportunities
* Sigma Level 1:691,462 defects per million opportunities

The goal of Six Sigma is to reduce defects and improve processes by identifying and eliminating the causes of variation and waste. Organizations implementing Six Sigma often follow the DMAIC (Define, Measure, Analyze, Improve, Control) methodology.

Example:

Let's say you work for a company that produces widgets. The length of these widgets is a critical quality parameter. A Six Sigma project is initiated to reduce the variation in widget lengths.

* Define: Clearly outline the problem, goals, and scope of the project.
* Measure: Collect data on the length of widgets produced and calculate the current sigma level.
* Analyze: Identify sources of variation and potential causes for defects in widget length.
* Improve: Implement changes to the production process to reduce variation.
* Control: Monitor and control the process to ensure continued improvement and adherence to the new standards.

By implementing Six Sigma principles, the company aims to produce widgets with consistent lengths, ultimately leading to higher customer satisfaction and reduced waste.

**Question: 2**

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

**Answer:**

Data that does not have a log-normal distribution or a Gaussian (normal) distribution is often referred to as non-normally distributed data. Non-normally distributed data may exhibit different patterns and shapes in its probability distribution. There are various types of distributions for non-normally distributed data, and examples include:

1. Skewed Distributions:

* Positively Skewed (Right-skewed): In a positively skewed distribution, the tail on the right-hand side is longer or fatter than the left-hand side. Examples include income distribution, where a small number of people earn significantly higher incomes than the majority.

* Negatively Skewed (Left-skewed): In a negatively skewed distribution, the tail on the left-hand side is longer or fatter than the right-hand side. An example could be the distribution of grades on a very easy test, where most students score high, but a few score very low.

2. Uniform Distribution:

A uniform distribution occurs when all possible values of a variable are equally likely. For example, the outcome of rolling a fair six-sided die follows a uniform distribution.

3. Exponential Distribution:

The exponential distribution is often associated with the time between events in a process where events occur continuously and independently at a constant rate. An example is the time between arrivals of customers at a service point.

4. Weibull Distribution:

- The Weibull distribution is used to model reliability and life data. It can describe a variety of shapes, including exponential and normal distributions. It is commonly used in survival analysis.

5. Poisson Distribution:

- The Poisson distribution represents the number of events that occur in a fixed interval of time or space. Examples include the number of phone calls received at a call center in an hour or the number of emails received in a day.

6. Discrete Distributions:

- Some data may follow discrete distributions, such as the binomial or hypergeometric distributions, which are used when dealing with discrete events or counts.

**Question: 3**

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

**Answer:**

The five-number summary is a descriptive statistics technique used to summarize the distribution of a dataset. It consists of five key values that divide the data into four intervals, providing a quick snapshot of the dataset's central tendency and spread. The five numbers include the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum.

Here's a breakdown of the five-number summary:

1. Minimum (Min): The smallest value in the dataset.

2. First Quartile (Q1): The median of the lower half of the dataset, representing the 25th percentile.

3. Median (Q2): The middle value of the dataset when it is sorted in ascending order, representing the 50th percentile.

4. Third Quartile (Q3): The median of the upper half of the dataset, representing the 75th percentile.

5. Maximum (Max): The largest value in the dataset.

Example:

Consider the following dataset:

12, 18, 20, 24, 30, 32, 36, 42, 48, 52

1. Minimum (Min): The smallest value is 12.

2. First Quartile (Q1): The median of the lower half is 20, average of 18 and 24.

3. Median Q2: The middle value is 32.

4. Third Quartile Q3:The median of the upper half is 42 average of 36 and 48.

5. Maximum Max: The largest value is 52.

The five-number summary for this dataset is 12, 20, 32, 42, 52.

The five-number summary is particularly useful for providing a concise summary of the dataset's central tendency and spread, allowing for a quick understanding of its distribution. Box-and-whisker plots are often created using the five-number summary to visualize the spread and identify potential outliers in the data.

**Question: 4**

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

**Answer:**

Correlation is a statistical measure that describes the degree of association between two variables. It indicates whether and how strongly pairs of variables tend to change together. The most common measure of correlation is the Pearson correlation coefficient, denoted by *r*. It ranges from -1 to 1, where:

*r*=1: Perfect positive correlation

* *r*=−1: Perfect negative correlation
* *r*=0: No correlation

Git hub link-----><https://github.com/prachikabra121/Inueron_Assessment/blob/main/Stats_Question4.ipynb>

**Machine learning**

**Question: 1**

Imagine you have a dataset where you have different Instagram features like u **sername , Caption , Hashtag , Followers , Time\_Since\_posted , and likes ,** now your task is to predict the number of likes and Time Since posted and the rest of the features are your input features. Now you have to build a model which can predict the number of likes and Time Since posted.

[Dataset](https://www.kaggle.com/datasets/rxsraghavagrawal/instagram-reach) This is the Dataset You can use this dataset for this question.

**Answer:** Github link→<https://github.com/prachikabra121/Inueron_Assessment/blob/main/ML_Question1.ipynb>

**Question: 2**

1. Train an SVM regressor on : [Bengaluru housing dataset](https://www.kaggle.com/datasets/amitabhajoy/bengaluru-house-price-data)

**Must include in details:**

- EDA

- Feature engineering

**Answer:Github link→**[**https://github.com/prachikabra121/Inueron\_Assessment/blob/main/ML\_Question2.ipynb**](https://github.com/prachikabra121/Inueron_Assessment/blob/main/ML_Question2.ipynb)

**Question: 3**

Train and fine tune a decision tree using the wine dataset by following the following steps:-

1. Use load\_wine() to generate wine dataset

2. Split the dataset into train and test dataset

3. Use random search CV to hyperparameter tune the Decision Tree

4. Try to achieve an accuracy of at least 85%

Grow a random forest using the following steps:-

1. Continuing the previous question, create 10 subsets of the training dataset. You can use the ShuffleSplit class for it.

2. Train 1 decision tree on each subset, using the best hyperparameter values found in the previous question.

3. Evaluate all the trees on the test dataset. Are they performing better than the tree created in the previous question?

**Answer**: Github link→https://github.com/prachikabra121/Inueron\_Assessment/blob/main/ML\_Question3.ipynb

**Deep Learning**

**Question: 1**

**(a) Explain how you can implement DL in a real-world application.**

**(b) What is the use of Activation function in Artificial Neural Networks? What would be the problem if we don't use it in ANN networks.**

**Answer:**

**(a) Implementing Deep Learning in a Real-World Application:**

**-To implement Deep Learning (DL) in a real-world application, follow these general steps:**

**1. Define the Problem:**

**- Clearly understand the problem you want to solve and determine if DL is suitable for it.**

**2. Collect and Preprocess Data:**

**- Gather relevant data for training and testing. Ensure the data is clean, labeled, and diverse.**

**- Preprocess the data by handling missing values, scaling, encoding categorical variables, etc.**

**3. Choose a DL Framework:**

**- Select a DL framework based on your expertise, project requirements, and community support. Common frameworks include TensorFlow, PyTorch, and Keras.**

**4. Design the Neural Network Architecture:**

**- Define the architecture of your neural network, including the number of layers, types of layers, and the activation functions.**

**5. Model Training:**

**- Split the data into training and validation sets.**

**- Train the model using appropriate optimization algorithms and loss functions.**

**- Adjust hyperparameters, like learning rate and batch size, based on performance.**

**6. Evaluate and Fine-Tune:**

**- Assess the model's performance on a separate test set.**

**- Fine-tune the model based on evaluation results, considering overfitting and underfitting.**

**7. Deploy the Model:**

**- Once satisfied with the model's performance, deploy it to a production environment.**

**- Choose deployment options based on the application, like cloud-based services or on-premise deployment.**

**8. Monitoring and Maintenance:**

**- Implement monitoring mechanisms to track the model's performance in real-time.**

**- Regularly update the model with new data to ensure its relevance.**

**9. Scale and Optimize:**

**- Optimize the model for efficiency and scale if needed.**

**- Explore hardware accelerators (GPUs, TPUs) to speed up computations.**

**10. Security and Ethics:**

**- Implement security measures to protect models and data.**

**- Address ethical considerations related to data privacy and biases.**

**(b) Activation Function in Artificial Neural Networks (ANN):**

**Activation functions introduce non-linearity to the neural network, allowing it to learn complex patterns and relationships in the data. The key purposes of activation functions in ANNs are:**

**1. Introduce Non-Linearity:**

**- Most real-world data is non-linear, and activation functions help neural networks model these non-linear relationships. Without non-linearity, a neural network would be equivalent to a linear model, limiting its expressive power.**

**2. Enable Learning Complex Patterns:**

**- Activation functions enable the network to learn intricate patterns and representations from the input data. Non-linear activation functions are essential for capturing the hierarchical features present in complex datasets.**

**3. Enable Backpropagation:**

**- During backpropagation, gradients are calculated and used to update weights. The non-linearity introduced by activation functions ensures that gradients can flow backward through the network, facilitating learning.**

**Common activation functions include ReLU (Rectified Linear Unit), Sigmoid, Tanh, and Softmax.**

**Problem if Not Using Activation Function:**

**If activation functions were not used in ANNs, the entire network would behave like a linear model, regardless of its depth. This limitation would severely restrict the model's ability to learn complex representations, rendering it ineffective for tasks requiring non-linear mappings. Activation functions are crucial for the expressive power and effectiveness of deep neural networks.**

**Question: 2**

**Train a Pure ANN with less than 10000 trainable parameters using the MNIST Dataset**

**Answer:** Git hub link-<https://github.com/prachikabra121/Inueron_Assessment/blob/main/DL_Question2.ipynb>

**Question3:**

Perform Regression Task using ANN

Note: You are feel free to use any Regression ML dataset

**Answer:** Git hub link→<https://github.com/prachikabra121/Inueron_Assessment/blob/main/DL_Question3.ipynb>