**PYTHON**

**Question: 1**

**You have an input dictionary given,**

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

**Task: You have to write a Python function that will take this input and print it like that,**

**output = {"abc":["def","ghi","jkl","mno","pqr","stu","vwx","yz"],**

**"def":["ghi","jkl","mno","pqr","stu","vwx","yz"],**

**"ghi":["jkl","mno","pqr","stu","vwx","yz"],**

**"jkl":["mno","pqr","stu","vwx","yz"],**

**"mno":["pqr","stu","vwx","yz"],**

**"pqr":["stu","vwx","yz"],**

**"stu":["vwx","yz"],**

**"vwx":["yz"],**

**"yz":["you are finally here !!!"]}**

**ANS. Check the repository notebook file ->**[**https://github.com/Kashi2723/Test-03-10-2024.git**](https://github.com/Kashi2723/Test-03-10-2024.git)

**(In test.ipynb file)**

**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**

**Explanation: 1st horse at stall 1, 2nd horse at stall 4 and 3rd horse at stall 8**

**ANS. Check the repository notebook file ->**[**https://github.com/Kashi2723/Test-03-10-2024.git**](https://github.com/Kashi2723/Test-03-10-2024.git)

**(In test.ipynb file)**

**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.**

**ANS. Check the repository notebook file ->** [**https://github.com/Kashi2723/Test-03-10-2024.git**](https://github.com/Kashi2723/Test-03-10-2024.git)

**(In test.ipynb file)**

**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**

**ANS. Check the repository notebook file ->** [**https://github.com/Kashi2723/Test-03-10-2024.git**](https://github.com/Kashi2723/Test-03-10-2024.git)

**(In test.ipynb file)**

**SQL QUESTIONS :**

**Q1. Given the following tables:**

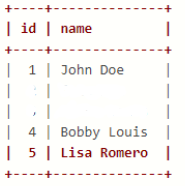


**What will be the result of the query below?**

**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.** This query will return all the runners who have never won a race.

****

id name

1 John Doe

4 Bobby Louis

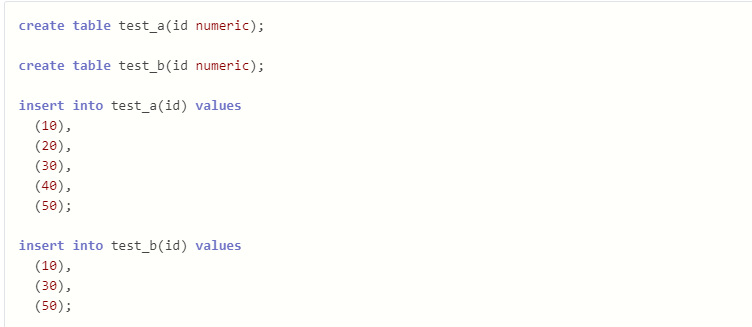
5 Lisa Romero

But this query may show unexpected behaviour as we can see that our races table have a null value.

Alternative solution:

SELECT \* FROM runners WHERE id NOT IN (SELECT winner\_id FROM races where winner\_id is not NULL)

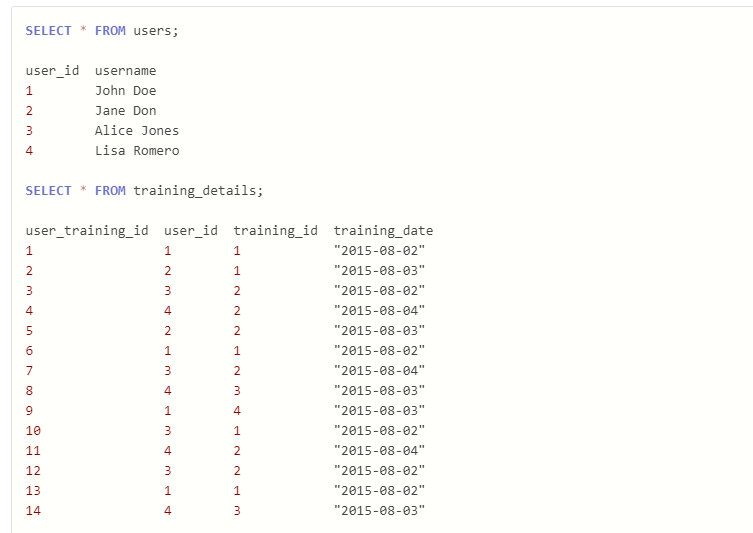
**Q 2. Given two tables created as follows**

****

**Write a query to fetch values in table test\_a that are and not in test\_b without using the NOT keyword.**

**Ans.** SELECT a.id FROM test\_a a LEFT JOIN test\_b b ON a.id = b.id WHERE b.id IS NULL;

**Q 3. Given the following tables:**

****

**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.**

SELECT

u.user\_id,

u.username,

td.training\_id,

td.training\_date,

COUNT(\*) AS num\_lessons\_taken

FROM

users u

JOIN

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

GROUP BY

u.user\_id,

u.username,

td.training\_id,

td.training\_date

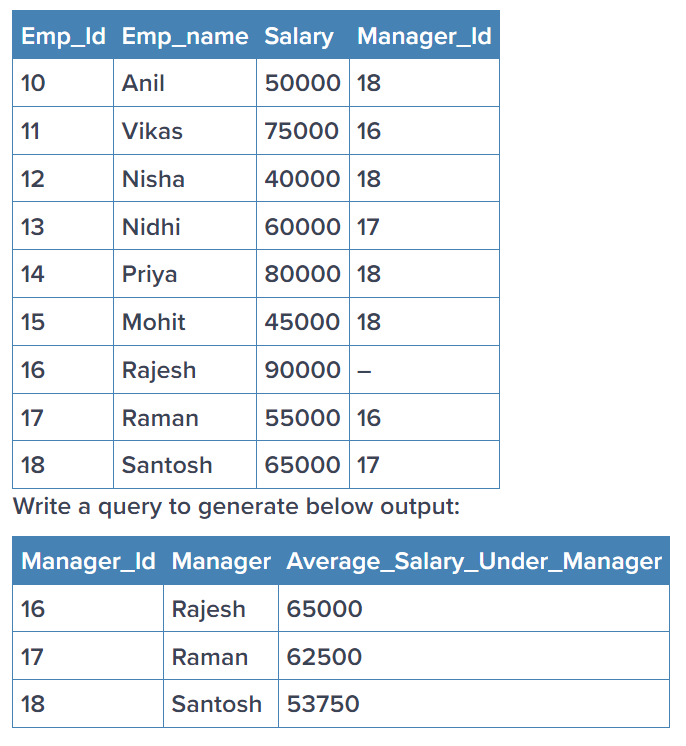
HAVING

COUNT(\*) > 1

ORDER BY

td.training\_date DESC;

**Q4. Consider the Employee table below.**

****

**Ans.**

SELECT

t.Manager\_id,

(SELECT

CASE

WHEN Manager\_id = '-' THEN 'Rajesh'

ELSE Emp\_name

END

FROM employee

WHERE Emp\_id = t.Manager\_id) AS Manager,

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

FROM

employee t

WHERE

t.Manager\_id != '-'

GROUP BY

t.Manager\_id;

**STATISTICS QUESTIONS :**

**Theory Based Answers :**

**Q1. What is the meaning of six sigma in statistics? Give proper example**

**Ans**. Six Sigma is a methodology used by companies to improve the quality of their processes. It aims to minimize defects and variations in production, providing a more consistent and high-quality output. The term "Six Sigma" refers to achieving a very low defect rate, approximately 3.4 defects per million opportunities. The process involves defining goals, measuring current performance, analyzing data, making improvements, and implementing controls to sustain those improvements. The ultimate goal is to deliver products or services that consistently meet or exceed customer expectations.

For Example : Imagine we work in a pizza restaurant, and our task is to ensure that the diameter of each pizza is consistently 12 inches, as per customer expectations.

* Define: Identify the goal - maintaining a pizza diameter of 12 inches to meet customer expectations.
* Measure: Measure the diameters of a sample of pizzas and calculate the standard deviation to understand how much the sizes vary from the target.
* Analyze: Analyze the data to identify the reasons for variations. It could be related to the dough preparation, oven temperature, or other factors.
* Improve: Make necessary adjustments, such as refining the dough preparation process or controlling the oven temperature more precisely, to reduce variations in pizza diameter.
* Control: Implement control measures, like regularly monitoring pizza sizes, to ensure that the improvements are sustained over time.

By applying Six Sigma principles, the goal is to consistently produce pizzas with a diameter very close to the target of 12 inches, resulting in a high-quality product and satisfied customers.

**Q 2. What type of data does not have a log-normal distribution or a Gaussian distribution? Give proper example**

**Ans.** Not all types of data follow a log-normal distribution or a Gaussian distribution (also known as a normal distribution). One common type of data that does not have a log-normal or Gaussian distribution is skewed data.

Skewed data is characterized by an asymmetric distribution, where the tail on one side of the distribution is longer or fatter than the other. There are two types of skewness:

* Positive Skewness (Right Skewed): The right tail of the distribution is longer or fatter than the left. The majority of the data points are concentrated on the left side.
* Negative Skewness (Left Skewed): The left tail of the distribution is longer or fatter than the right. The majority of the data points are concentrated on the right side.

Example: Income Distribution

One classic example of skewed data is the distribution of income. In many societies, a small percentage of the population earns a very high income, creating a right-skewed distribution. The majority of people earn moderate to low incomes, and the income distribution extends with a long tail to the right due to a few individuals or households with exceptionally high earnings.

In this scenario, applying logarithmic transformations may not result in a normal or log-normal distribution because the fundamental structure of the data is inherently skewed.

**Q 3. What is the meaning of the five-number summary in Statistics? Give proper example**

**Ans.** 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, the first quartile (Q1), the median (Q2 or the second quartile), the third quartile (Q3), and the maximum. These values help to understand the center, spread, and overall shape of the data.

The five numbers are often denoted as:

1. Minimum: The smallest value in the dataset.
2. Q1 (First Quartile): The value below which 25% of the data falls.
3. Median (Q2 or Second Quartile): The middle value of the dataset. If there is an even number of observations, it is the average of the two middle values.
4. Q3 (Third Quartile): The value below which 75% of the data falls.
5. Maximum: The largest value in the dataset.

Example: Exam Scores

Let's consider the exam scores of a class:

50,65,70,75,80,85,90,95,100

* Minimum: 50 (the smallest score)
* Q1 (First Quartile): 70 (25% of the scores fall below 70)
* Median (Q2 or Second Quartile): 80 (the middle score)
* Q3 (Third Quartile): 90 (75% of the scores fall below 90)
* Maximum: 100 (the largest score)

The five-number summary helps to understand the overall distribution of the scores. For instance, it shows that the middle half of the scores (from Q1 to Q3) ranges from 70 to 90, giving an idea of the spread and concentration of scores in that interval. Additionally, it indicates the range of the dataset from the minimum to the maximum.

**Q 4. What is correlation? Give an example with a dataset & graphical representation on jupyter Notebook**

**Ans.** Correlation is a statistical measure that quantifies the degree to which two variables change together. In other words, it assesses the strength and direction of a linear relationship between two variables. The correlation coefficient ranges from -1 to 1:

* 1 indicates a perfect positive correlation,
* -1 indicates a perfect negative correlation,
* 0 indicates no correlation.

**For example : check the notebook** [**https://github.com/Kashi2723/Test-03-10-2024.git**](https://github.com/Kashi2723/Test-03-10-2024.git)

**(In test.ipynb file)**

**Machine Learning Questions :**

**Q1. 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.**

**Ans. Check the repository notebook file ->** [**https://github.com/Kashi2723/ML-Q1.git**](https://github.com/Kashi2723/ML-Q1.git)

**Q2.**

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**

**Ans. check the repository** [**https://github.com/Kashi2723/ML-Q2.git**](https://github.com/Kashi2723/ML-Q2.git)

**Q3. 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?**

**Ans. Check the repository** [**https://github.com/Kashi2723/ML-Q3.git**](https://github.com/Kashi2723/ML-Q3.git)

**Deep Learning Questions :**

**Q1. (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.**

**Ans. (a). Implementing Deep Learning (DL)** in a real-world application involves several steps:

* Define the Problem: Clearly define the problem you want to solve and determine if DL is the right approach. DL is suitable for tasks such as image and speech recognition, natural language processing, and more.
* Collect and Preprocess Data: Gather a sufficient amount of labeled data for training and testing. Preprocess the data to ensure it is in a suitable format and is representative of the real-world scenarios.
* Choose a DL Framework: Select a deep learning framework such as TensorFlow, PyTorch, or Keras. These frameworks provide a set of tools and abstractions to simplify the implementation of neural networks.
* Design the Neural Network Architecture: Define the architecture of your neural network. This includes the number and type of layers, the activation functions, and the connections between neurons.
* Train the Model: Split your dataset into training and testing sets. Train the model on the training set using an optimization algorithm, adjusting the weights and biases of the network to minimize the error
* Validate and Tune: Evaluate the model on the validation set to ensure it generalizes well to new data. Fine-tune hyperparameters and architecture based on performance
* Deploy the Model: Once satisfied with the model's performance, deploy it to the real-world environment. This could involve integrating it into a web application, a mobile app, or an embedded system.
* Monitor and Update: Regularly monitor the model's performance in the real-world environment. If necessary, update the model with new data and retrain it to adapt to changing conditions.

**(B). Use of Activation Function**:

Introducing Non-Linearity: Activation functions introduce non-linearities into the network, allowing it to model and understand complex patterns and relationships in the data.

Learning Complex Representations: Non-linear activation functions enable the neural network to learn hierarchical and intricate representations of the input data, which is essential for capturing features at different levels of abstraction.

Gradient Descent Optimization: Activation functions help in the optimization process during training by providing gradients that allow the network to adjust its parameters through backpropagation.

**(C). Problem Without Activation Function:-** If neural networks had no activation functions, they would fail to learn the complex non-linear patterns that exist in real-world data

**Q2 Train a Pure ANN with less than 10000 trainable parameters using the MNIST Dataset**

**Ans. check the repository :** [**https://github.com/Kashi2723/DL-Q2.git**](https://github.com/Kashi2723/DL-Q2.git)

**Q3. Perform Regression Task using ANN**

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

**Ans.** [**https://github.com/Kashi2723/DL-Q3.git**](https://github.com/Kashi2723/DL-Q3.git)