Top of Form

iNeuron DS Assessment

**aspirecomputers26@gmail.com** [Switch account](https://accounts.google.com/AccountChooser?continue=https://docs.google.com/forms/d/e/1FAIpQLSdifRId30T1vNBjF6BB-KcEgD-EXVFdSDdeOyDUzVECrC1OSQ/viewform?usp%3Dsend_form&service=wise)

Not shared

\* Indicates required question

**Python:**

Total Marks: 40

Each question 10 marks

**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 !!!"]}

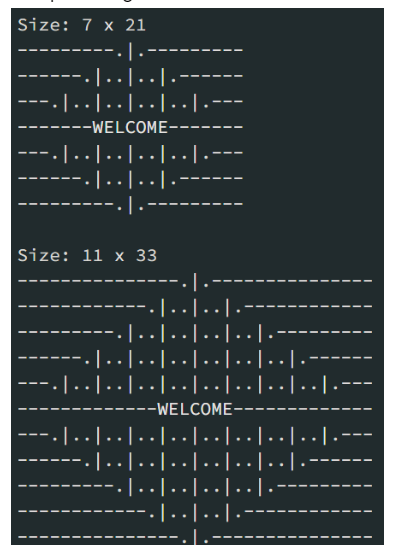
**Question: 2**

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

**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.

**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

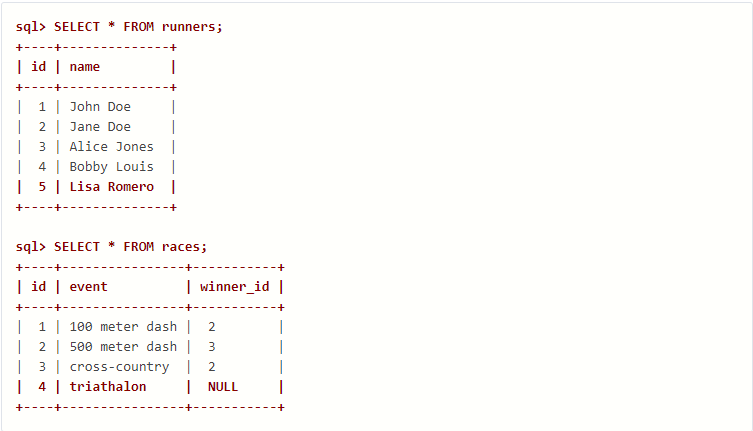
**SQL:**

Total Marks: 40

Each question 10 marks

**Question: 1**

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

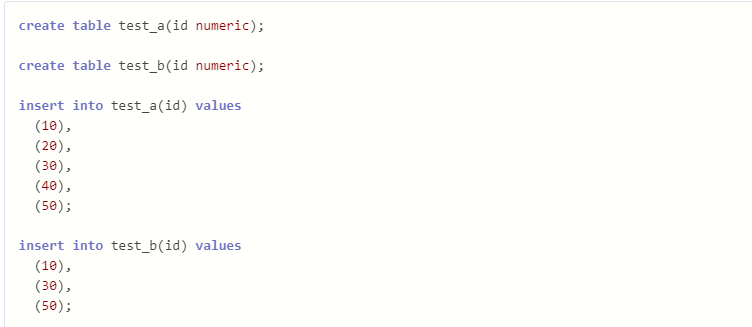
SELECT \* FROM runners

WHERE id NOT IN

(SELECT winner\_id FROM races WHERE winner\_id IS NOT null)

**Question: 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 =>**

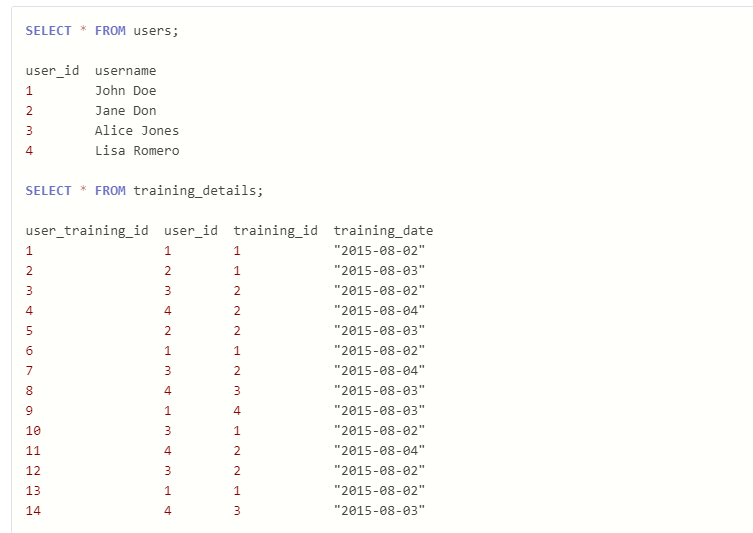
select \* from test\_a

except

select \* from test\_b;

**Question: 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,

username,

training\_id,

training\_date,

count( user\_training\_id ) AS count

FROM users u JOIN training\_details t ON t.user\_id = u.user\_id

GROUP BY u.user\_id,

username,

training\_id,

training\_date

HAVING count( user\_training\_id ) > 1

ORDER BY training\_date DESC;

**Output=>**

user\_id username training\_id training\_date count

4 Lisa Romero 2 August, 04 2015 00:00:00 2

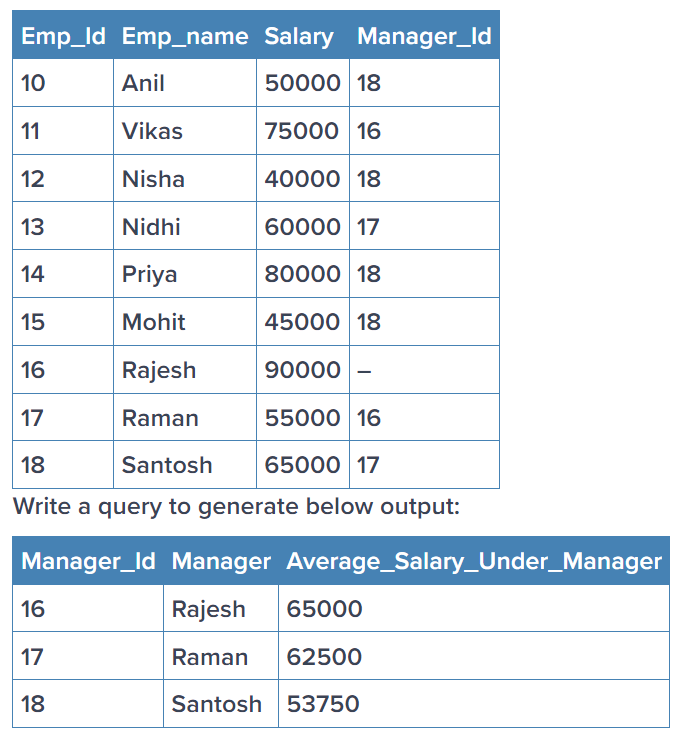
4 Lisa Romero 3 August, 03 2015 00:00:00 2

1 John Doe 1 August, 02 2015 00:00:00 3

3 Alice Jones 2 August, 02 2015 00:00:00 2

**Question: 4**

Consider the Employee table below.



Ans=>

SELECT name, salary

FROM employees

WHERE salary > (

SELECT AVG(salary)

FROM (

SELECT salary

FROM employees

ORDER BY salary

LIMIT 2 - (SELECT COUNT(\*) FROM employees) MOD 2

OFFSET (SELECT (COUNT(\*) - 1) / 2 FROM employees)

) AS subquery

);

**Statistics:**

Total Marks: 40

Each question 10 marks

**Question: 1**

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

Six Sigma is based on the idea that all business processes can be measured and optimized.

The term Six Sigma originated in manufacturing as a means of quality control. Six Sigma quality is achieved when long-term defect levels are below 3.4 defects per million opportunities (DPMO).

Six Sigma has since evolved into a more general business concept, focusing on meeting customer requirements, improving customer retention, and improving and sustaining business products and services.

**The 5 Steps of Six Sigma**

The Six Sigma method uses a step-by-step approach called DMAIC, an acronym that stands for Define, Measure, Analyze, Improve, and Control. According to Six Sigma adherents, a business may solve any seemingly unsolvable problem by following these five steps.

### **Define**

A team of people, led by a Six Sigma expert, chooses a process to focus on and defines the problem it wishes to solve.

### **Measure**

The team measures the initial performance of the process, creating a benchmark, and pinpoints a list of inputs that may be hindering performance.

### **Analyze**

Next the team analyzes the process by isolating each input, or potential reason for any failures, and testing it as the possible root of the problem.

### **Improve**

The team works from there to implement changes that will improve system performance.

### **Control**

The group adds controls to the process to ensure it does not regress and become ineffective once again.

**Examples of Six Sigma**

Six Sigma is used by many companies, local governments, and other institutions. Here are two examples of how Six Sigma improved operational efficiency, saved money, and increased customer satisfaction.

**Microsoft**

Microsoft is one of the largest software producers in the world. It used Six Sigma to help eradicate defects in its systems and data centers and systematically reduce IT [infrastructure](https://www.investopedia.com/terms/i/infrastructure.asp) failures.

The company first established standards for all of its hardware and software to create a baseline measurement for detecting defects.

### Example 2)

### **Ventura County, California, Government**

Ventura County, California, credited the use of Lean Six Sigma for a savings of $33 million. The county government began to use the program in 2008 and has trained more than 5,000 employees in the methodology. The county says the savings are due in part to the introduction of more efficient new systems and the elimination of unnecessary, but time-consuming, steps from its prior processes

**Question: 2**

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

For example the Poisson Distribution, which describes (among other things) the number of unlikely events occurring after providing a sufficient opportunity for a few events to occur. It is pretty non-Gaussian unless the mean number of events is very large. The mathematical form of the distribution is still Poisson, but a histogram of the number of events after many trials with a large average number of events eventually looks fairly Gaussian.

Another example is the location of the centers of raindrop ripples on a pond; they are not uniformly spaced in (say) the east-west direction, but they are uniformly distributed.

The simplest example is the distribution of numbers that show up on the top of a fair die after a large number of throws. Each number from 1 to 6 will occur with approximately equal frequency. Increasing the number of throws will not tend to produce a bell-shaped histogram, in fact the fractional occurrence will approach a constant 1/6 over the possible numbers.

**Question: 3**

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

A five-number summary is especially useful in descriptive analyses or during the preliminary investigation of a large data set. A summary consists of five values: the most extreme values in the data set (the maximum and minimum values), the lower and upper quartiles, and the median. These values are presented together and ordered from lowest to highest: minimum value, lower quartile (Q1), median value (Q2), upper quartile (Q3), maximum value.

These values have been selected to give a summary of a data set because each value describes a specific part of a data set: the median identifies the centre of a data set; the upper and lower quartiles span the middle half of a data set; and the highest and lowest observations provide additional information about the actual dispersion of the data. This makes the five-number summary a useful measure of spread.

A five-number summary can be represented in a diagram known as a box and whisker plot. In cases where we have more than one data set to analyse, a five-number summary with a corresponding box and whisker plot is constructed for each.

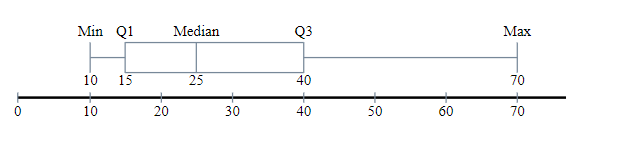
**Examples**  
**1. Calculate Five number summary from the following data  
10,50,30,20,10,20,70,30**  
  
**Solution:**  
**Five number summary :**  
10,50,30,20,10,20,70,30  
  
Steps of Five-Number Summary  
  
**Step-1: Arrange the numbers in ascending order**  
10,10,20,20,30,30,50,70  
  
**Step-2: Find the minimum value**  
Minimum =10 (the smallest number)  
  
**Step-3: Find the maximum value**  
Maximum =70 (the largest number)  
  
**Step-4: Find the median**  
The median is the middle number in a sorted data set and N is the total number of elements  
If N is odd then the median is a single middle number, and if N is even then the median is the average of the two middle numbers.

**Question: 4**

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

10,10,20,20,30,30,50,70  
  
*N*=8 is even, so median is the average of the two middle numbers at position 4 and 5  
  
We have 20+302=25  
  
∴ Median =25  
  
**Step-5: Place parentheses around the numbers above and below the median.**  
{10,10,20,20},{30,30,50,70}  
  
**Step-6: Find *Q*1 by finding the median for lower half of data(left of the median)**  
  
10,10,20,20  
  
*N*=4 is even, so median is the average of the two middle numbers at position 2 and 3  
  
We have 10+202=15  
  
∴*Q*1=15  
  
**Step-7: Find *Q*3 by finding the median for upper half of data(right of the median)**  
  
30,30,50,70  
  
*N*=4 is even, so median is the average of the two middle numbers at position 2 and 3  
  
We have 30+502=40

∴*Q*3=40  
  
**Step-8: Summary found in the above steps.**  
Minimum =10  
  
*Q*1=15  
  
Median =25  
  
*Q*3=40  
  
Maximum =70



**Machine learning:**

Total Marks: 60  
Each question 20 marks

**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.

**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

**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?

**Deep Learning :**

Total Marks: 60  
Each question 20 marks

**Question: 1**

1. Explain how you can implement DL in a real-world application.

Machine learning and deep learning are both subsets of artificial intelligence. Deep learning is the evolved and advanced phase of machine learning. In machine learning, human programmers create algorithms that learn from the data and derive analyses.

Deep learning is different from machine learning in that it works on an artificial neural network which closely represents a human brain. The same network allows machines to analyze data just the way humans do. Such machines with deep learning capacities do not require to act upon the instructions of human programmers.

Deep learning is made possible through the ginormous amounts of data that we create and consume daily. Every deep learning model makes extensive use of data to facilitate data processing.

**Real World Applications of Deep Leaning**

**1. Computer vision**

High-end gamers interact with deep learning modules on a very frequent basis. Deep neural networks power bleeding-edge object detection, image classification, image restoration, and image segmentation. So much so, they even power the recognition of hand-written digits on a computer system. To wit, deep learning is riding on an extraordinary neural network to empower machines to replicate the mechanism of the human visual agency.

**2. Text generation**

Machines now have the power to generate new text from the scratch. They can learn the punctuation, grammar, and style of a piece of text and pen down effective news pieces. Robo-journalists riding on deep learning models have been producing accurate match reports for at least three years now. And the skill isn’t limited to match report writing exclusively.

AI-based text generation is fully equipped to handle the complexity of opinion pieces on issues concerning you and myself. As of now, text generation has helped create entries on just about everything from children’s rhymes to scholarly topics.

**3. Language identification**

At this point, we are looking at a preliminary stage where deep learning machines can differentiate between different dialects. For example, a machine will make the decision that someone is speaking in English. It will then make a distinction based on the dialect. Once the dialect has been established, further processing will be handled by another AI that specializes in the particular language. Not to mention, there is no human intervention in any of these steps.

(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.

Activation functions are really important for a Artificial Neural Network to learn and make sense of something really complicated and Non-linear complex functional mappings between the inputs and response variable.They introduce non-linear properties. Their main purpose is to convert a input signal of a node in a A-NN to an output signal.

If you do not use any activation function in a neural network, it would become a giant linear regression model as (X\*W1\*W2\*W3) + (b1\*W2\*W3 + b2\*W3 + b3) that can be simplified into (X\*W) + B. Moreover, the hidden layers would be useless and the model will not learn any non-linear relationship in the data.

If we do not apply a Activation function then the output signal would simply be a simple linear function.A linear function is just a polynomial of one degree. a linear equation is easy to solve but they are limited in their complexity and have less power to learn complex functional mappings from data. A Neural Network without Activation function would simply be a Linear regression Model, which has limited power and does not perform good most of the times.

**Question: 2**

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

**Question: 3**

Perform Regression Task using ANN

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

**Submission Process**

There are Two Types of Questions Theory based Question and Project-based (where you actually have to code)

First and foremost, set up a github repository and save all of your answers there.

Create a Google Doc with written answers to all questions, explaining theory. Keep it in the GitHub repository for easy access.

When coding is required, create a folder in the same GitHub repository, upload your code, and share the repository link for submission.

*Eg. Answer. 6 Python  - > GitHub repo link*

*Note:*

* *If you are building any End to end project try to write code in .py file*
* *If you are only analyzing or doing EDA use .ipynb file*

Submit the final GitHub repo link containing all answers, ensuring completeness and easy access for evaluation.

Your Name: \*

Mrs. Shubhangi Ravindra Mankar

Your Email:\*

Aspirecomputers26@gmail.com

Your Phone Number: \*

8888322568

Give your submission link:  
\*

Your answer

Submit

Clear form

Never submit passwords through Google Forms.

Bottom of Form

This form was created inside of iNeuron Intelligence Private Limited. [Report Abuse](https://docs.google.com/forms/u/0/d/e/1FAIpQLSdifRId30T1vNBjF6BB-KcEgD-EXVFdSDdeOyDUzVECrC1OSQ/reportabuse?source=https://docs.google.com/forms/d/e/1FAIpQLSdifRId30T1vNBjF6BB-KcEgD-EXVFdSDdeOyDUzVECrC1OSQ/viewform?usp%3Dsend_form)

[Forms](https://www.google.com/forms/about/?utm_source=product&utm_medium=forms_logo&utm_campaign=forms)