***Instructions*:**

● The following test will evaluate abilities based on Python, Machine Learning, SQL

sections. The more questions you answer, the better your scores will be in the evaluation

● Submit your code files and SQL question’s response and upload the folder containing all

the files on GitHub repositories or mail with YOUR\_NAME - EMAIL ADDRESS

● Submit your answers by replying to the DataBeat team (marked in the email) along with

links to your GitHub code repositories or zip files with your code with the subject line

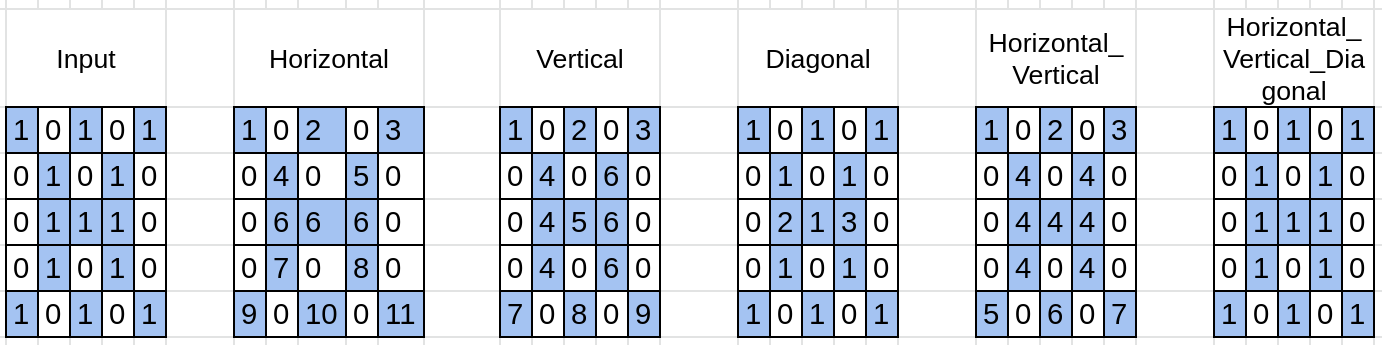
DataBeat Interview: YOUR\_NAME - EMAIL ADDRESS

● You will have 2-3 hours to complete as many questions as you like

***Python:***

Q1. Consider a matrix whose cells are filled with 0’s and 1’s. If two 1’s are adjacent to each other, they are said to be connected. A component is defined as a set of one or more 1’s that are adjacent to each other. An adjacency can be of 5 types - horizontal, vertical, diagonal, horizontal\_vertical and horizontal\_vertical\_diagonal. Given a binary matrix, find the total number of components for a given adjacency type.

An example of a binary matrix with each of the 5 adjacency options is shown below.



| **Adjacency** | **Total Components** |
| --- | --- |
| Horizontal | 11 |
| Vertical | 9 |
| Diagonal | 3 |
| Horizontal\_Vertical | 7 |
| Horizontal\_Vertical\_Diagonal | 1 |

Write a python program to return the total components whose input from stdin would be as follows:

1. Integer representing the adjacency. horizontal, vertical, diagonal, horizontal\_vertical and horizontal\_vertical\_diagonal are represented by 1,2,3,4 and 5 respectively.
2. Contains two integers representing the rows and columns of the input matrix.
3. Each of next line contains 0’s and 1’s representing the respective row of the input matrix.

Sample input and output:

Input: 1

5 5

1 0 1 0 1

0 1 0 1 0

0 1 1 1 0

0 1 0 1 0

1 0 1 0 1

Output: 11

***Machine Learning:***

Q1. Build linear and logistic regression using gradient descent on numpy only (Do not use other libraries like sklearn). Dataset for linear regressions can be found [here](https://archive.ics.uci.edu/ml/datasets/Wine+Quality). Dataset for logistic regression can be found [here](https://drive.google.com/file/d/1CReNesgs3iqZi08rr5a1h0YOci0BmC6H/view?usp=sharing). The main focus here is not the accuracy of the model but the implementation of these algorithms from scratch.

Q2. Use pandas to extract insights on this [dataset](https://drive.google.com/file/d/1Fwxy08NgMAVcUvsRVd94KEiqgmYSvOKA/view?usp=sharing).

[movie\_metadata.csv](https://drive.google.com/file/d/1Fwxy08NgMAVcUvsRVd94KEiqgmYSvOKA/view?usp=sharing)

***SQL:***

**Employees**

| **EMP\_ID** | **FIRST\_NAME** | **LAST\_NAME** | **SALARY** | **JOINING\_DATE** | **DEPARTMENT** |
| --- | --- | --- | --- | --- | --- |
| 001 | Manish | Agarwal | 700000 | 2019-04-20 09:00:00 | HR |
| 002 | Niranjan | Bose | 20000 | 2019-02-11 09:00:00 | DA |
| 003 | Vivek | Singh | 100000 | 2019-01-20 09:00:00 | DA |
| 004 | Asutosh | Kapoor | 700000 | 2019-03-20 09:00:00 | HR |
| 005 | Vihaan | Banerjee | 300000 | 2019-06-11 09:00:00 | DA |
| 006 | Atul | Diwedi | 400000 | 2019-05-11 09:00:00 | Account |
| 007 | Satyendra | Tripathi | 95000 | 2019-03-20 09:00:00 | Account |
| 008 | Pritika | Bhatt | 80000 | 2019-02-11 09:00:00 | DA |

**Variables Details**

| **EMP\_REF\_ID** | **VARIABLES\_DATE** | **VARIABLES\_AMOUNT** |
| --- | --- | --- |
| 1 | 2019-02-20 00:00:00 | 15000 |
| 2 | 2019-06-11 00:00:00 | 30000 |
| 3 | 2019-02-20 00:00:00 | 42000 |
| 4 | 2019-02-20 00:00:00 | 14500 |
| 5 | 2019-06-11 00:00:00 | 23500 |

**Designation Table**

| **EMP\_REF\_ID** | **EMP\_TITLE** | **AFFECTED\_FROM** |
| --- | --- | --- |
| 1 | Asst. Manager | 2019-02-20 00:00:00 |
| 2 | Senior Analyst | 2019-01-11 00:00:00 |
| 8 | Senior Analyst | 2019-04 -06 00:00:00 |
| 5 | Manager | 2019-10-06 00:00:00 |
| 4 | Asst. Manager | 2019-12-06 00:00:00 |
| 7 | Team Lead | 2019-06-06 00:00:00 |
| 6 | Team Lead | 2019-09-06 00:00:00 |
| 3 | Senior Analyst | 2019-08-06 00:00:00 |

Please use MSSQL/MYSQL/Oracle for the following:

1. Write a SQL query to create these tables in your database with the following characteristics:

a. Add the primary key “Emp\_ID” to the Employees Table. Also, mention what are the constraints used in SQL.

b. Add foreign key “EMP\_REF\_ID” in Variables Details and Designation Table that references “Emp\_ID” in Employees Table

2. What are the four different types of joins? Give examples of each by performing the joins on the Employees table and Designation Table.

a. Write a query to get the employee details(full name and department) who received the highest and the least variables

b. Write a query to get the designation which has got the highest and second lowest amount (salary + variables) for the whole year of 2019. Get the corresponding amount values.

c. What is cross join? Write a query to give an example of the same by performing it on the Employees table and Designation table.

d. What are the clauses used with Select statements and what are the orders of it? Write a query to get the employee details who got their designations updated in the second half of the year 2019(July to December), sorted by the “variables\_amount” (highest to lowest) where the department name of the Employee has the letter ‘A’ in it.

3. What is a Cursor? Write a query/queries to use the cursor to store the Employees Name( full name) for the HR department into a variable called ‘emp\_names’

a. What is Normalization and explain different forms of normalization with examples. (preferable with the above tables)

b. What is the stored procedure? Write a stored procedure to call the query that you have written for Q2.a