**Set 1:-**

Use data set to answer the following questions using SQL :-

[**https://raw.githubusercontent.com/rashida048/Datasets/master/movie\_dataset.csv**](https://raw.githubusercontent.com/rashida048/Datasets/master/movie_dataset.csv)

(Read Json data using Python / R Script and create **Movie** table in database)

1. Which is the movie(s) with the smallest runtime? Which is the movie(s) with the highest runtime?

2. Take the top 5 production houses (by budget) and list their top 5 most popular movies, their revenue and vote\_average.

3. List the production house for every year from 2000-2016 which has released the most number of movies in that year.

4. You are going to invest all your life's savings in a production company. You have two choices: "Marvel Studios" and "DC Comics". Which company would you bet on? This is an open ended question. Define your own metrics to measure which one is a better investment opportunity and defend your analysis.

**Set 2:- Solve Below Question Using SQL :-**

**Table 1:- User Profile Data**

| User Id | Install Date | Role | Org ID |
| --- | --- | --- | --- |
| 1 | 2022-05-01 12:30:45 | Site Engineer | O1 |
| 1 | 2023-04-01 00:30:46 | Planning Manager | O2 |
| 1 | 2023-07-01 05:45 | Owner | OAE1 |
| 2 | 2021-04-03 14:23 | SIte Engineer | OO1 |
| 2 | 2022-12-12 03:25 | Site Supervisor | OO2 |
| 3 | 2023-01-01 16:45 | Project Manager | AA1 |
| 3 | 2023-03-03 10:10 | Project Manager | AA2 |
| 3 | 2023-03-03 11:11 | CXO | AA3 |

1. **Write a query to return the latest entry against the user id using SubQuery & Join.**
2. **Write a query to return the latest entry against the user id using the window function.**

**Table 2** ORDER\_TABLE  **:-**

| user\_id | order\_id | purchase\_datetime | product | category | product\_revenue |
| --- | --- | --- | --- | --- | --- |
| 56KHB | 1234 | 2021-01-20 13:33:44 | biscuit | food | 100 |
| 56KHB | 1234 | 2021-01-20 13:33:44 | crocin | medicine | 50 |
| 32HBK | 1235 | 2021-08-20 13:38:55 | chips | food | 100 |
| 67ABC | 1236 | 2021-08-20 15:32:12 | shoes | footwear | 1200 |
| 67ABC | 1236 | 2021-08-20 15:32:12 | shirt | clothing | 500 |
| 67ABC | 1236 | 2021-01-20 15:32:12 | earphones | electronics | 450 |
| 67ABC | 1237 | 2021-08-20 16:18:19 | laptop | electronics | 45000 |
| 67ABC | 1237 | 2021-08-20 16:18:19 | socks | clothing | 150 |

1. Write a query to find the number of products bought in the month of January 2021.

2. Write a query to find the second order\_id for each user\_id ( Without Window Function)

3. Write a query to find the min, max and average time between two orders for any user.

**Set 3 :-**

**Json Data (Meta Data) :-**

mobile:{

“Name”: “xyz”,

“Mobile” :”Iphone”

“Mobileid”:23,

“Mapping”: {

“Imei”:12344567777,

“Specification”: {

Model :- “Apple 12”,

OS Version:-17.0.1,

},

“Country”:”us”,

“Product\_name”:”iphone 12”

}

}

Write a python script to extract following information (without using any inbuilt Function) use only pandas (list, dictionary etc), and Loop (for, while etc)

**Output** :- Imei, model,os version,country,product\_name,mobileid

**Set 4:-**

The attachments are two datasets, one contains organization, project, and user-level event data and another one has the respective event description and their sections on the app. Here, Event means any click or page load, or scroll happening on the app.

Q. The goal is to identify patterns & insights from the dataset. Also, look out for insights/triggers that activated the users and subsequently engaged them.

We understand this is very open-ended, and that's the point. Be creative.

**Note: Please provide the insights in a PPT format.**