**Lead Python Developer Task**

Helpful pointers:

1. Having high code readability is a big plus. Usage of docstrings, proper naming conventions, and clean code is recommended.
2. Use object oriented Python. Keep object and function names relevant, with no space left for ambiguity.
3. A single Python file is preferred (You may submit SQL codes inside a docstring, or in a separate txt file). If you’re using Jupyter Notebooks, please send a clean notebook (without unneeded cells).
4. It is preferable to make a single object which has methods for each question / subquestion.

Questions:

1. Assume a database table with the following schema:

events(visitor\_id, session\_id, session\_start, session\_end, page\_visited)

session\_start and session\_end have format YYYY-MM-DD HH:MI:SS.

page\_visited has the pages visited by that user in that session.

Write a single SQL query that finds out details of users who have visited ‘pricing’ + 1 other page, and have a session time > 45 seconds.

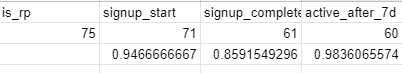
1. Find the excel sheet attached with this doc. Refer to ‘user\_data’ and ‘session\_data’ columns

Use any programming tools necessary to get only visitor\_id, session\_id, country, event\_date and signup\_start onto a dataframe. Sort by ascending event\_dates. Select only those rows where Experiment Number=10000, and event\_date lies in 2021/05/03-2021/05/09 (inclusive) range. Export this dataframe onto a csv

Refer to the “session\_data” sheet in the Excel sheet given. Create a function that counts the number of rows where value=1 **on a per day basis**, *for each column.*

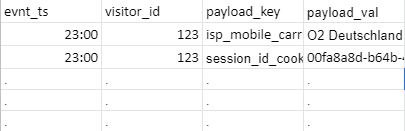
1. Write a python program that:

* Let’s assume that after some data manipulation, you get data which is present in the sheet “session\_sum”. Write a function, which finds out the ratio of a column with respect to its previous one (for example, col2/col1, col3/col2 and so on). Export the result into a separate csv, with appropriate naming. Your final output should look like:



* Write a function that plots a graph (using any module of your choice) which shows the number of visitors per day.

1. Refer to the ‘event\_data’ sheet given. The payload column contains keys=value pairs separated by ‘&’. Make a function that exports a csv which expands the data in the following format:



and so on for **all key-value pairs** in the payload