# **HOME TEST 1**

## **SQL (4 questions)**

### **Question 1**

There are 3 tables containing information about the **artists**, **songs**, and **global\_song\_rank**. ([link data](https://docs.google.com/spreadsheets/d/16_Encirn1elIvhiQ0LC18JdV0R1WQHWZmhPH-go5Xf8/edit#gid=1467009589)) Write a query to determine the top 5 artists whose songs appear in the Top 10 of the **global\_song\_rank** table the most.

Assumptions:

* If two artists' songs have the same number of appearances, the artists should have the same rank.
* The rank number should be continuous (1, 2, 2, 3, 4, 5) and not skipped (1, 2, 2, 4, 5).

Expected Output:

| **artist\_name** | **artist\_rank** |
| --- | --- |
| Bad Bunny | 1 |
| Ed Sheeran | 2 |
| Adele | 3 |
| Lady Gaga | 3 |
| Katy Perry | 4 |
| Drake | 5 |

*Suggestions: import data from Google Sheet (link) to Bigquery or any SQL platforms you have and start to write your query.*

### **Question 2 - 4**

There are 3 tables containing information about the **my\_order\_trans**, **dim\_product**, and **my\_buyer\_profile**. ([link data](https://docs.google.com/spreadsheets/d/16_Encirn1elIvhiQ0LC18JdV0R1WQHWZmhPH-go5Xf8/edit#gid=1322321973))

**Question 2**: Find the lifetime **total orders**, **total spent (GMV), unique items bought , earliest purchase date, last purchased date, average amount spent per order (AOV) and average purchase price (APP)** for the following buyer IDs and their purchased products’ categories:

* 1076216361964070
* 3190859517651870
* 3754202390878020

Expected output:

| **customer\_id** | **category** | **total\_orders** | **total\_spent** | **unique\_items** | **earliest\_date** | **last\_date** | **AOV** | **APP** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |

Note:

* total orders: count distinct - order\_id
* unique items bought: count distinct - product\_id
* average amount spent per order: total GMV / total order
* average purchase price: total price/ total order

**Question 3: Find out the top 10 cross border items with the highest quantity sold. The output includes minimum selling price, total spent (gmv) and total orders.**

**Expected output:**

| **product\_name** | **category** | **min\_selling\_price** | **total\_qty\_sold** | **total\_gmv** | **total\_order** | **rank** |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |

**Question 4: Find the average time (in day unit) between their first and second checkout of our customers (for example: the avg time between first & second checkout of customer A is 3.5 days)**

## **Logical Reasoning**

[Link Test](https://docs.google.com/forms/d/e/1FAIpQLSel2rlRoi7IRO53j5syxnZxqRAr0SBFUq5Xy-TOQ3dFeuh3vg/viewform?usp=sf_link)

Note: You should spend only 40 minutes to complete this test.

## **Cohort Analysis**

*What is cohort and cohort analysis?*

A cohort is a collection of users who have something in common. A traditional cohort, for example, divides people by the week or month of which they were first acquired. When referring to non-time-dependent groupings, the term segment is often used instead of the cohort.

Cohort analysis is a tool to measure user engagement over time. It helps to know whether user engagement is actually getting better over time or is only appearing to improve because of growth. Customers are divided into mutually exclusive cohorts, which are then tracked over time. Vanity indicators don’t offer the same level of perspective as cohort research.

Generally, there are three major types of Cohort:

* Time cohorts: customers who signed up for a product or service during a particular time frame.
* Behavior cohorts: customers who purchased a product or subscribed to service in the past.
* Size cohorts: refer to the various sizes of customers who purchase the company’s products or services.

However, we will focus on performing **Cohort Analysis based on Time**. Customers will be divided into acquisition cohorts depending on the month of their first purchase. The cohort index would then be assigned to each of the customer’s purchases, which will represent the number of months since the first transaction.

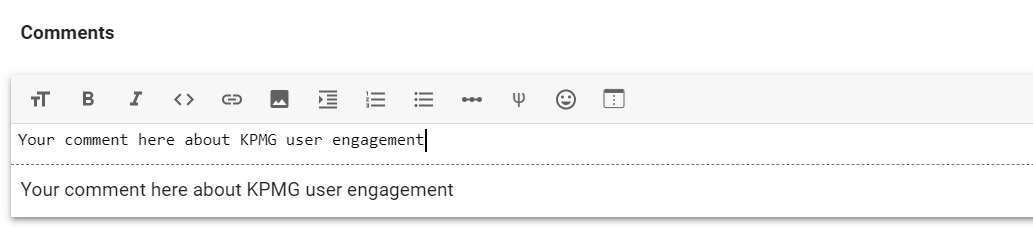
Now, from Python, you will use KPMG transaction data to make a cohort to evaluate user engagement from their first transaction ([Link Dataset](https://docs.google.com/spreadsheets/d/1FgOKnuw66XdYoIWDcvMQ76fot6aVR-VX/edit#gid=1514900709)). Note that you need to filter **order\_status = ‘Approved’** and remove duplicates if they exist.

Expected Output



After that, you need to write down **3 - 5 main** **information/ insights** from the data to evaluate KPMG

user engagement.



**FORM OF SUBMISSION**

**1. SQL & Cohort Analysis**

You will have a total of **3 days** to complete 2 parts. Then, submit your output (Bigquery & Colab) link

via the form: [link](https://docs.google.com/forms/d/e/1FAIpQLSfmxAAZQkSVO57fjUZ9hErmWg1BIDxqOg-_IFXDff6hTngK-A/viewform?usp=sharing)

For Bigquery, you can:

* import data by doing these steps: [link](https://www.youtube.com/watch?v=z6g-Q-IF5jo)
* share the link by doing these steps: [link](https://www.youtube.com/watch?v=TsxYhtJnTkU)