SQL Quiz

#### **Question 1 (1 point)**

Write SQL script to check if there is a duplicate record in table **“shopee.test\_candidate\_03”** and how many candidates have duplicated records?

**Hint :**

* Please refer to the table below
* **One candidate** must have only **1 record**

**Table** (shopee.test\_candidate\_03) :

| candidate\_id | candidate\_name | candidate\_working\_exp | candidate\_score | ingestion\_timestamp |
| --- | --- | --- | --- | --- |
| 1111 | Somjai | 10 | 1 | 2022-01-02 |
| 1112 | Somphong | 10 | 2 | 2022-01-02 |
| 1113 | Somsri | 5 | 4 | 2022-01-02 |
| 1113 | Somsri | 5 | 4 | 2022-01-04 |
| 1114 | Somchai | 2 | 4 | 2022-01-04 |
| 1115 | Sompong | 3 | 2 | 2022-01-05 |
| 1116 | Somwang | 3 | 2 | 2022-01-05 |
| 1117 | Somwung | 10 | 4 | 2022-01-05 |
| 1116 | Somwang | 3 | 2 | 2022-01-06 |
| 1117 | Somwung | 10 | 4 | 2022-01-06 |

**Expected output :**

| candidate\_id | num |
| --- | --- |
| 1113 | xx |
| 1116 | xx |
| 1117 | xx |

##### 

##### **Answer 1**

With base as (Select candidate\_id

, count(\*) as num

from shopee.test\_candidate\_03

group by candidate\_id)

Select \* from base where num >1

#### **Question 2 (2 points)**

Write SQL script to find how many daily new buyers have made their FIRST purchase **with Fashion Cluster** in April 2021?

**Hint :**

* **Need to join 2 tables** to get first purchase transaction of each buyer

Tables :

* order\_item : This table keeps order transaction at order level on a daily basis
* first\_purchase : This table keeps the first purchase date of each shopee users (1 buyer can have only 1 record)

Key: buyer\_id

Fields :

* order\_item.create\_datetime is when buyers make their purchase
* first\_purchase.result\_first\_purchase\_time is when buyers make their first purchase

**Expected output :**

| create\_date  (in date format) | num\_of\_first\_purchase\_buyers |
| --- | --- |
| 2022-04-01 | xxx |
| 2022-04-02 | xxx |
| … | xxx |
| 2022-04-30 | xxx |

##### 

##### 

##### **Answer 2**

WITH

base AS (

SELECT

Distinct

oi.create\_datetime,

oi.order\_id,

oi.buyer\_id,

oi.cluster\_orders\_item,

fp.result\_first\_purchase\_time

FROM

order\_item oi

LEFT JOIN

first\_purchase fp

ON

oi.buyer\_id = fp.buyer\_id),

first\_purchase\_fashion AS (

SELECT

date\_format(result\_first\_purchase\_time,

“%Y-%m-%d”) AS first\_purchase\_date,

buyer\_id

FROM

base

WHERE

cluster\_orders\_item = “Fashion”

AND create\_datetime = result\_first\_purchase\_time)

SELECT

first\_purchase\_date,

COUNT(\*)

FROM

first\_purchase\_fashion

GROUP BY

first\_purchase\_date

#### **Question 3 (2 points)**

Write SQL script to find top 10 shops who gain high Gross Merchandize Value (GMV) in April 2021

**Hint :**

* Use order\_item table
* Use sum(gmv\_usd) to find total gmv
* Use window function to get the top 10 shops with highest GMV in April 2021

**Expected output :**

| month | shop\_id | total\_gmv | ranking |
| --- | --- | --- | --- |
| April | 12345 | xxx | 1 |
| April | 23456 | xxx | 2 |
| … | … | … | … |
| April | 56789 | xxx | 10 |

##### **Answer 3**

WITH ShopGMV AS (

SELECT

shop\_id,

SUM(gmv\_usd) AS total\_gmv,

RANK() OVER (ORDER BY SUM(gmv\_usd) DESC) AS ranking

FROM

order\_item

WHERE

EXTRACT(MONTH FROM create\_datetime) = 4 -- April

AND EXTRACT(YEAR FROM create\_datetime) = 2021 -- 2021

GROUP BY

shop\_id

)

SELECT

'April' AS month,

shop\_id,

total\_gmv,

ranking

FROM

ShopGMV

WHERE

ranking <= 10

ORDER BY

ranking;