

## Part 1: Forming SQL Queries

The following question(s) will be based on the two tables shown below:

**Table Name:** *my\_order\_trans*

**Description:** *Table records all orders/transactions at product variation level from the beginning of time*

columns	description	type
buyer_id	unique identifier of the buyer	bigint
gmv	gross merchandise value, a.k.a. total sales	double
is_brand	1 indicates product is a brand product	integer
is_cross_border	1 indicates product is a cross border product	integer
is_lowest_price	1 indicates product is a lowest price product	integer
is_new_purchase	1 indicates it is a first purchase order	integer
item_rating	product ratings	double
l1_cat	product's main category	varchar
l2_cat	product's sub category	varchar
on_campaign	1 indicates product was in campaign(s)	integer
order_date	date at which order was created	date
order_id	unique identifier of the order	bigint
order_status	status of the order	varchar
price	price at which item was transacted at	double
price_before_discount	price before discount	double
product_id	unique identifier of the product	bigint
product_listed_on	date as of when product was first listed	date
product_name	product name	varchar
promo_code	promotion code if any	varchar
qty_sold	number of units sold	double
return_status	1 indicates it is a returned order	integer
shipping_fee_paid	shipping fee paid by the buyer	double
store_id	unique identifier of the shop	bigint
var_id	unique identifier of the product variation	bigint

**Table Name:** *my\_buyer\_profile*

**Description:** *Table records all buyer details*

columns	description	type
buyer_id	unique identifier of the buyer	bigint
birthday	buyer's date of birth	date
gender	buyer's gender; 1 indicates male, 2 indicates female	integer

**Q1:** Find the lifetime total orders, total spent (gmv), unique items bought, earliest purchase date, last purchased date, average amount spent per order and average purchase price for the following buyer IDs and their purchased products' main categories:

- 576123
- 123152

Expected output:

Userid	Main Category	Total Orders	Total Spent (RM)	Unique Items Bought	Earliest Purchase Date	Last Purchased Date	Avg Spending per Order (RM)	Avg Purchase Price (RM)

**Q2: Find out the top 10 cross border items with the highest quantity sold last month, together with their tier \*\*, minimum selling price, total spent (gmv) and total orders.**

**\*\* Tier is a user-defined item attribute dimension with 3 unique values:**

- Short Tail (>20 average daily orders)
- Mid Tail (between 10 - 20 average daily orders)
- Long Tail (< 10 average daily orders)

Expected output:

Rank	Product ID	Item Name	Category	Tier	Min. Selling Price	Total Qty Sold	Total GMV	Total Orders

**Q3: The following buyers purchased on Shopee on separate days and on several occasions:**

- 123456
- 987654
- 34567

Find the average time (in hrs) between their first and second checkout in the last 120 days.

Expected output:

Average time between 1st and 2nd purchase (hour)

## Part 2: Writing Pseudocodes & Python Codes

1. Write a function that will be able to return the smallest positive integer missing from an unsorted array.

**For example:**

**Input:** [2, 3, -7, 6, 8, 1, -10, 15, -35]

**Output:** 4

2. Write a function that will be able to return the Nth element from the following sequence:

0, 3, 3, 6, 9, 15, 24, 39, 63, 102, 165,...

For e.g. if  $N = 3$ , output = 3 ; if  $N = 7$ , output = 24

### 3. Write a function that will be able to remove duplicates from a given array.

Output requirements:

- a. Array consists of a unique list of strings
- b. Ascending order
- c. First letter of each array element (string) capitalized
- d. All trailing and leading spaces removed

**For example:**

**Input:** ['Panasonic', 'personic', 'panasonic', 'Haier', 'electrolux', 'Electrolux']

**Output:** ['Electrolux', 'Haier', 'Panasonic', 'Personic']

## Part 3: Open-Ended Question

How would you develop data mart and data flow for rider incentive which will be used for the rider to check the incentive on daily basis.

*Suggestions/Tips:*

1. *A rider will have 4 factors: delivery point, pickup point, return point, and delivery successful point (% of monthly delivery successful)*
2. *Incentives will be calculated cumulatively based on the sum of 4 factors*
3. *Think about the daily and monthly maintenance (detail data and summary) and flexibility of monthly base rate changes*
4. *Keep your answers structured*