

# Project instructions

## Athena queries

1. Create a table called **order\_products\_prior** by using the last SQL query you created from the previous assignment. It should be similar to below (note you need to replace the s3 bucket name "imba" to yours own bucket name):

```
CREATE TABLE order_products_prior WITH (external_location =  
's3://imba/features/order_products_prior/', format = 'parquet')
```

```
as (SELECT a.*,  
        b.product_id,  
        b.add_to_cart_order,  
        b.reordered  
FROM   orders a  
        JOIN order_products b  
        ON a.order_id = b.order_id  
WHERE  a.eval_set = 'prior')
```

2. Create a SQL query (user\_features\_1). Based on table **orders**, for each user, calculate the max order\_number, the sum of days\_since\_prior\_order and the average of days\_since\_prior\_order.
3. Create a SQL query (user\_features\_2). Similar to above, based on table **order\_products\_prior**, for each user calculate the total number of products, total number of distinct products, and user reorder ratio(number of reordered = 1 divided by number of order\_number > 1)
4. Create a SQL query (up\_features). Based on table **order\_products\_prior**, for each user and product, calculate the total number of orders, minimum order\_number, maximum order\_number and average add\_to\_cart\_order.
5. Create a SQL query (prd\_features). Based on table **order\_products\_prior**, first write a sql query to calculate the sequence of product purchase for each user, and name it product\_seq\_time (For example, if a user first time purchase a product A, mark it as 1. If it's the second time a user purchases a product A, mark it as 2). Below are some examples:

User_id	Order_number	Product_id	Product_seq_time	....
123	1	Egg	1	
123	1	Apple	1	
123	1	Fish	1	
123	2	Egg	2	
123	2	Mop	1	
123	2	Banana	1	
123	3	Egg	3	
123	3	Orange	1	
123	3	Fish	2	
123	3	Salad	1	

Then on top of this query, for each product, calculate the count, sum of reordered, count of product\_seq\_time = 1 and count of product\_seq\_time = 2.