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).

```
SELECT user_id,

    Max(order_number)         AS user_orders,

    Sum(days_since_prior_order) AS user_period,

    Avg(days_since_prior_order) AS user_mean_days_since_prior

FROM   orders

GROUP  BY user_id
```

3. Create a SQL query (user_features_2).

```
SELECT user_id,

    Count(*)                    AS user_total_products,

    Count(DISTINCT product_id)        AS user_distinct_products ,
```

```
    Sum(CASE WHEN reordered = 1 THEN 1 ELSE 0 END) / Cast(Sum(CASE WHEN
order_number > 1 THEN 1 ELSE 0 END) AS DOUBLE) AS user_reorder_ratio
FROM   order_products_prior
GROUP  BY user_id
```

4. Create a SQL query (up_features).

```
SELECT user_id,
    product_id,
    Count(*)            AS up_orders,
    Min(order_number)     AS up_first_order,
    Max(order_number)     AS up_last_order,
    Avg(add_to_cart_order) AS up_average_cart_position
FROM   order_products_prior
GROUP  BY user_id,
       product_id
```

5. Create a SQL query (prd_features).

```
SELECT product_id,
    Count(*)      AS prod_orders,
    Sum(reordered) AS prod_reorders,
    Sum(CASE WHEN product_seq_time = 1 THEN 1 ELSE 0 END)  AS prod_first_orders,
    Sum(CASE WHEN product_seq_time = 2 THEN 1 ELSE 0 END)  AS prod_second_orders
FROM   (SELECT *,
           Rank()
             OVER (
               partition BY user_id, product_id
               ORDER BY order_number) AS product_seq_time
        FROM   order_products_prior)
```

```
GROUP  BY product_id
```