Marketing Analytics - Week 3

Transaction Analysis 1

In week 3 we will:

- Explore Dataset quality
- Clean the data
- Date Partitions
- Run Transaction analysis

Scenario

You are a marketing analyst at Myer. From your team, you have been provided with customer and date-level transaction data for the past few years. In this dataset, you will find the average monthly spending by households in each of these department stores and their postcodes. It can be used for the demonstration of any analysis that requires transaction information, like RFM. RFM analysis enables you to use a simple rubric to discover who your best customers really are and to make them even more of an asset. The data also provides response information of customers to a promotion campaign. Based on this dataset, your stakeholders are seeking the following questions.

The highlight of this dataset is that you can evaluate the effectiveness of the RFM group by checking one of the business metrics. Transaction data provides the index of customers (customer_id), date of transaction (trans_date), and the Amount of purchase (tran_amount).

You are going to present the RFM analysis to the team:

1. Upload the data table into your Google Cloud Platform
2. Check For Null Values for focal variables.
3. Calculation of Recency Score, Frequency Score, Monetary Score – raw data calculation
    - Recency: How recent was the customer's last purchase? (recent_order)
    - Frequency: How often did this customer make a purchase in this given period? (total_orders)
    - Monetary: How much money did this customer spend in this given period? (total_sales)
4. Segmentation using RFM: Convert raw data to actionable RFM analysis.
    a. While you can slice the data as finely as you'd like, using quintiles mirrors the Pareto principle by indicating which of your customers are in the top 20% for all three metrics.

i. To create the quintiles, we'll use an NTILE window function that requires just a few lines of additional code at the top of your raw RFM Values query.
ii. Concatenate the quintile rank for each metric to create a 3-digit number, also known as a "cell."
iii. When concatenating the data this way, your best customers would have a cell value of 555 because they're in the top 20% for all three metrics, while a cell value of 111 would be given to customers in the bottom 20% for all three. Try to sort them based on your requirement.

b. Find the Customer who bought most recently and most often and spent the most.

5. Group Discussion: In what ways can Myer leverage the results of RFM to personalize marketing campaigns and improve customer retention?