

This Notebook

In this notebook, we'll answer requirement number 2 from the Data Evaluation Notebook:

2. Query the Database using SQL to provide business intelligence

The business questions we can answer for our stakeholder include:

1. What are the top 5 brands by receipts scanned for most recent month?
2. How does the ranking of the top 5 brands by receipts scanned for the recent month compare to the ranking for the previous month?
3. When considering average spend from receipts with 'rewardsReceiptStatus' of 'Accepted' or 'Rejected', which is greater?
4. When considering total number of items purchased from receipts with 'rewardsReceiptStatus' of 'Accepted' or 'Rejected', which is greater?
5. Which brand has the most spend among users who were created within the past 6 months?
6. Which brand has the most transactions among users who were created within the past 6 months?

Referencing our understanding gained in the Data Evaluation notebook and the constructed SQL Schema, we will use the following query to answer questions 3 and 4 using PostgreSQL 9.6. This query assumes that the data in its current form has been cleaned based on communication between the analyst and the stakeholder, and that the table structure within the database is as in the submitted schema. For example, I will take the liberty of assuming that 'FINISHED' in the `rewardsReceiptStatus` column means the same thing as and has been changed to 'ACCEPTED' in the database.

```
In [102]: # SELECT rewardsReceiptStatus,
#           AVG(totalSpent) AS average_spend,
#           COUNT(rewardsReceiptStatus) AS total_items
# FROM receipts
# WHERE rewardsReceiptStatus = 'ACCEPTED' OR rewardsReceiptStatus = 'REJECTED'
# GROUP BY rewardsReceiptStatus
```

This will return a table with rows 'ACCEPTED' and 'REJECTED' and columns 'average_spend' and 'total_items', ignoring rows from `receipts` that do not fall into either category, which will allow us to compare the quantities for each between the two groups and thus answer questions 3 and 4. For answers to these questions, see the bottom of the data evaluation notebook.