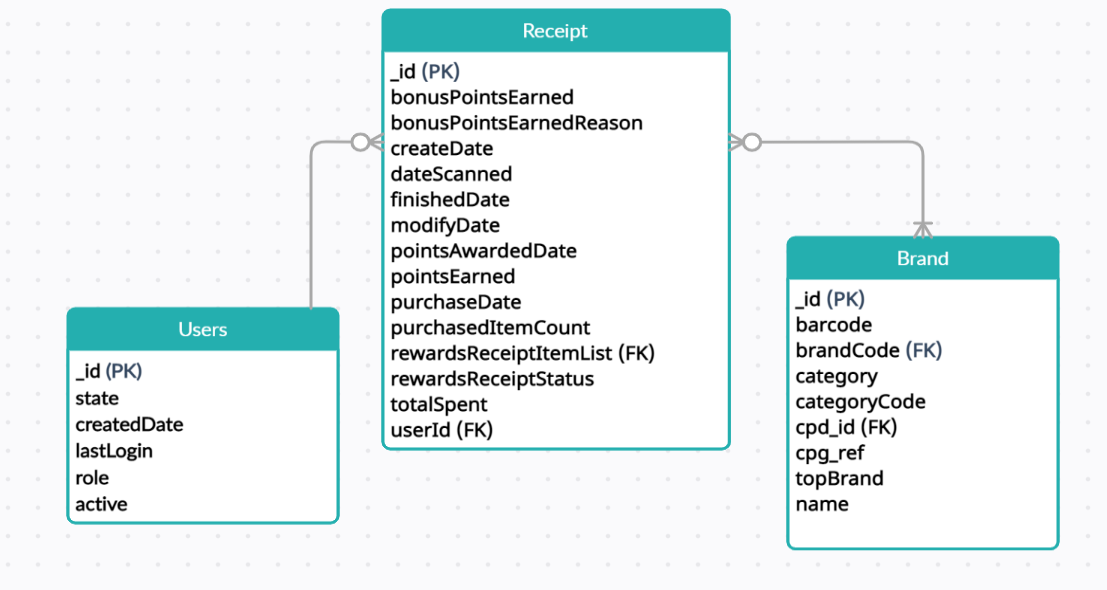
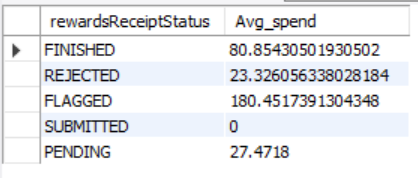
1. **Review unstructured JSON data and diagram a new structured relational data model.**



1. **Generate a query that answers a predetermined business question.**

**2.a) When considering *average spend* from receipts with 'rewardsReceiptStatus’ of ‘Accepted’ or ‘Rejected’, which is g­reater?**

SELECT rewardsReceiptStatus,avg(totalSpent) as Avg\_spend FROM sakila.receipts GROUP BY rewardsReceiptStatus;

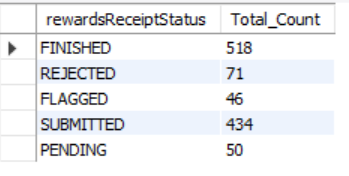


|  |
| --- |
| **Average spends for ‘Accepted’ is greater than ‘Rejected’.** |

Note: Finished are considered as Accepted. Flagged, Submitted and Pending could be rejected at later stages.

**2.b) When considering *total number of items purchased* from receipts with 'rewardsReceiptStatus’ of ‘Accepted’ or ‘Rejected’, which is greater?**

SELECT rewardsReceiptStatus, count(purchasedItemCount) as Total\_Count FROM sakila.receipts GROUP BY rewardsReceiptStatus;



|  |
| --- |
| **Total item count for ‘Accepted’ is greater than ‘Rejected’.** |

Note: Finished are considered as Accepted. Flagged, Submitted and Pending could be rejected at later stages.

1. **Generate a query to capture data quality issues against the new structured relational data model.**

One of the data quality issue could be the way ‘rewardsReceiptItemList’ is getting captured in receipts data. Brands of each item in the receipt is not captured which makes it difficult to the understand the sale distribution of brands. There are pointsPayerId and rewardsProductPartnerId which can be joined to brand table to get the associated brands, but it does not give us all the information and is missing for many items.

1. **Write a short email or Slack message to the business stakeholder.**

Dear Team,

Thank you for providing us with the data file. We would like to quickly set up a meeting with the team to go over the data file and understand a few issues. Here is some more information on the same.

We were able to successfully store the data in the database and gain some useful insights as a result. Although the data appears to be accurate, we may need additional details to fully comprehend the brand selling distribution.

Many of the receipt's items seem to be lacking the brand code or name, causing us to miss out on crucial information about which brands are doing well in the market. It would be fantastic if we could make a mandatory entry in the receipts data for each item's brand.

We will be able to say more about how much a brand is famous among users based on demography and user's industry once we receive all the data from the sales team. This data will be used by the sales and marketing teams to improve their sales and advertising strategies to attract more customers.

Also, please let us know if the business team requires any specific data insights that we might be able to provide.

Looking forward the getting the additional information on receipts data.

Please let me know if you have any questions.

Regards,

Viharika Bharti

Data Engineer

Fetch Rewards