Part 2: Answer business questions

Q1

Total Spend by Brand (Top 5)

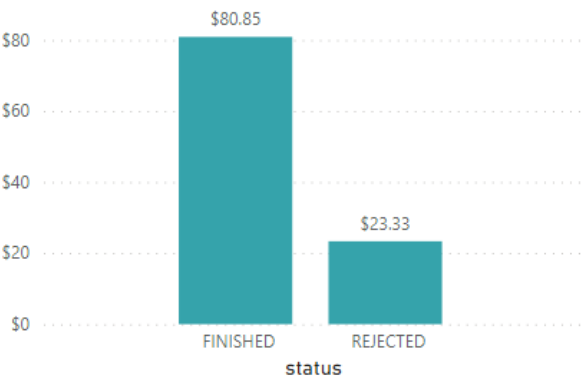| Brand	Total Spend
BEN AND JERRYS	$32.1K
PEPSI	$29.0K
FOLGERS	$27.9K
KRAFT	$27.0K
BIGELOW	$26.2K

Total Spend
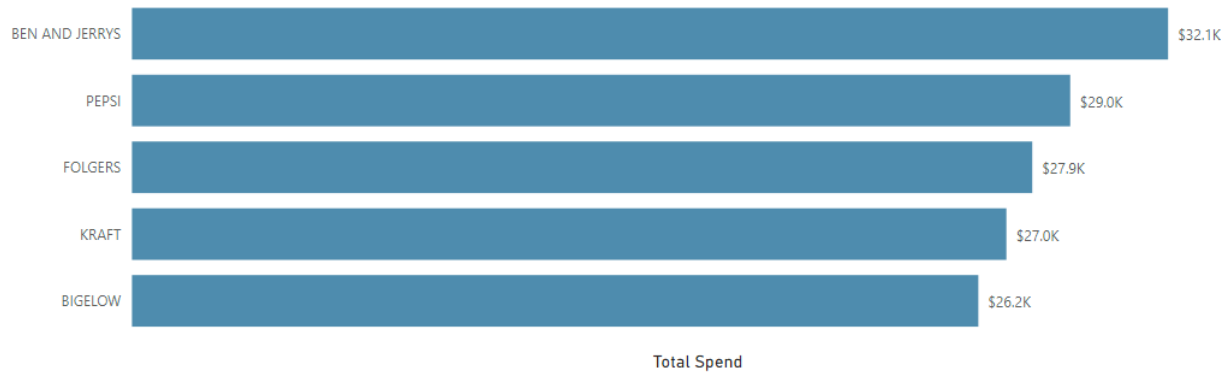
Receipt Count by Brand (Top 5)

Brand	Receipt Count
BEN AND JERRYS	31
PEPSI	23
KRAFT	22
KLEENEX	21
FOLGERS	20

Receipt Count

Q3 & Q4

Average Spend

status	Average Spend
FINISHED	$80.85
REJECTED	$23.33

Item Count

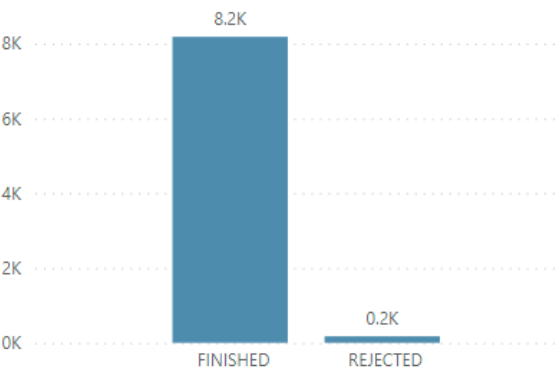| status	Item Count
FINISHED	8.2K
REJECTED	0.2K
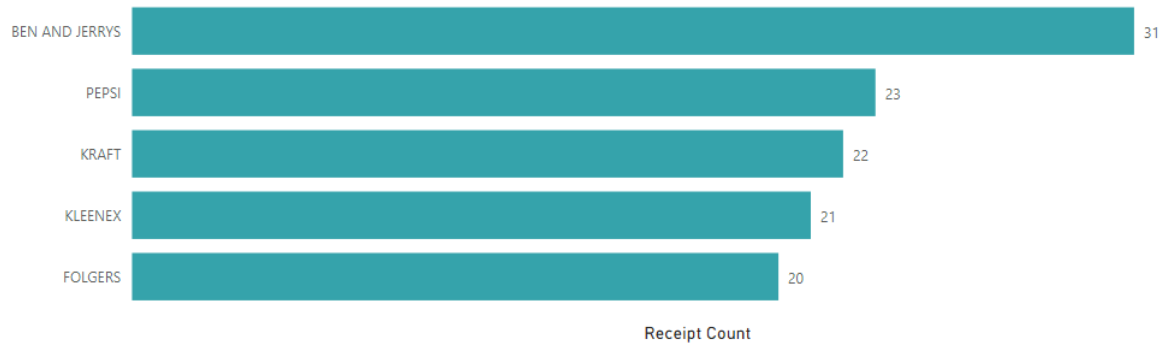
Part 4: Communicate

Hi Barb,

I have a bunch of questions for you regarding the receipts data I've been digging into to help answer the questions you are interested in. I'll share them with you below but it may be easier to have a short meeting to go over them. Let me know if that is the case and I'll set it up.

1. Can you explain the difference between Points Payers and Reward Product Partners? How are they related to brands?
2. In a previous discussion you mentioned a receipt status of 'Accepted'. I'm not seeing that in the data. The statuses we have are 'Submitted', 'Finished', 'Flagged', 'Pending', and 'Rejected'. Is it possible that you meant one of these statuses? If not, I'll research why I don't see 'Accepted'.
3. What does 'CPG' stand for and what is a 'CPG collection'?

I also noticed some holes in the data that could cause the results to not be accurate and understate things like total spend. I wanted to get your input on whether this is expected or a problem that needs to be looked into.

1. Should all items on a receipt be associated with a brand? If not, how do you want us to report the spend related to those items?
2. There are no receipts scanned in December 2020. Is there an explanation for that I should know or do I need to dig into that?
3. There are receipts with a status of 'Submitted' but all other fields are empty. Is this expected or do I need to research why the data is being lost?

Finally, I just wanted to address the concerns you expressed with regard to performance. We are using best practices in how we structure the data. The structure is optimized for doing the type of analytics you are interested in doing. The receipts line data will grow at a faster rate than the other data. We've done some forecasting on the rate at which the data will grow and the usage of the data. This will help us to select the appropriate database size and budget for the growth. We also have tools in place that will help us monitor usage and alert us if anything goes wrong.

Another possibility is to archive the data as it ages or partition it once it grows massively. Archiving depends on your requirements and we have some time to discuss the pros and cons and make a decision. Essentially, you'll want to consider how far back you'll want to pull data. If you only ever need to report on 3 years of data, for example, then we can archive data older than that. You'd still be able to access it but the process would be slightly different. If we decide to partition the data, it will happen in the database and the only change for you to see will be that the reports are running faster.

Does that ease your mind or is there anything I can clarify or answer for you regarding performance?