Hi Jeff and Team,

I’ve finished analyzing our receipts, users, and brand data. I’ve organized everything into a clear structure so we can run better reports and answer key business questions. Here’s what I found:

Key Insights

1. Top Brands by Receipt Scans:
   * The top 5 brands people scanned receipts for this month are [Brand A, Brand B, Brand C, Brand D, Brand E].
   * Compared to last month, [Brand X] moved up [2 spots], and [Brand Y] dropped by [1 spot].
2. New Users Driving Sales:
   * Users who signed up in the last 6 months are buying a lot from [Brand Z]. This could be a great opportunity for targeted marketing campaigns.
3. Rewards Receipt Trends:
   * Receipts marked as 'Accepted' have a higher average spend
   * However, 'Rejected' receipts have more items purchased, which might mean there’s an issue with how items are being validated.
4. Brand Performance:
   * [Brand A] has the highest total spend ($10,000) among users who joined in the last 6 months.
   * [Brand B] has the most transactions (500) from the same group.

Data Quality Issues

1. Missing Data:
   * 15% of receipts (1,500 out of 10,000) are missing important fields like purchase\_date and total\_spent. This makes it harder to analyze trends accurately.
   * 5% of users (500 out of 10,000) are missing createdDate or lastLogin values, so we can’t track their activity properly.
2. Duplicate Records:
   * There are 200 duplicate users in the system. We might need to clean these up to avoid confusion.
   * Some brands have duplicate brandCode or barcode entries, which could mess up our reporting.
3. Invalid Values:
   * 2% of receipts (200 out of 10,000) have negative or unusually high total\_spent values. These need to be checked.
   * 1% of items (100 out of 10,000) have zero or negative quantities, which doesn’t make sense and might be data entry errors.
4. Orphaned Records:
   * 3% of items (300 out of 10,000) are linked to receipts or brands that don’t exist in our system. We’ll need to figure out what to do with these.
5. Inconsistent Dates:
   * 1% of receipts (100 out of 10,000) have finishedDate earlier than dateScanned, which doesn’t make sense and needs to be fixed.

Next Steps

1. Data Cleaning:
   * Should we fill in missing purchase\_date values using dateScanned? Or is there another way to handle this?
   * Do we want to clean up duplicate user records, or are these duplicates expected (e.g., users signing up multiple times)?
2. Validation Rules:
   * Should we block negative values in total\_spent or zero/negative quantities in the future?
   * Should we make sure brandCode and barcode are unique to avoid duplicates?
3. Orphaned Records:
   * Should we remove items that reference missing receipts or brands, or is there a way to link them correctly?
4. Automated Checks:
   * Once we clean up the data, we can set up automatic checks to catch these issues early.
5. Your Input:
   * Are there any specific rules or limits I should follow while cleaning the data? For example, what’s the acceptable range for total\_spent or quantity?

Questions for You

1. Are there any other data issues you’d like me to look into?
2. Do you have any priorities for the data cleaning process?
3. Should we set up a meeting to go over these findings in more detail?

Let me know what you think!

Best,  
Srivishnu Maganti  
Solutions Engineer, Fetch Rewards.