**Subject**: Data Modeling and Analysis for Fetch Rewards - Brands and Receipt Rewards

Hi,

I hope you're doing well! I wanted to provide an update on the data quality issues we’ve identified and propose the next steps for implementing a data warehouse system.

Current Challenges:

Our data is stored in JSON format with three main files: brands, receipts, and users. The unstructured format makes it difficult to analyze and generate insights.

Proposed Solution:

I recommend creating a data warehouse to store this data in relational tables, which will enable faster queries, better reporting, and more efficient analysis. I’ve attached the proposed data model for your review.

**Next Steps**

Before we move ahead with finalizing the database structure, I would appreciate your feedback on a few key points:

1. **Mapping Brands to Receipts:** The current dataset lacks a direct relationship between brands and receipts, other than a shared cpg\_id stored as rewards\_product\_partner\_id in the receipt item list. Could we establish a more direct mapping between these two entities?
2. **Brand Code Consistency:** The brand\_code in the receipt item list doesn't align with the brand codes in the Brands table. Can we ensure that these codes represent the same data, or would it be possible to standardize them?

**Data Quality Issues Identified**

In the process of analyzing the data, I encountered the following quality concerns:

1. **Missing Values:** Several columns contain missing or null values. We need to assess the importance of these fields and determine the best approach for filling or handling these missing values.
2. **Duplicate Records:** The Users table contains duplicate records. I recommend removing these duplicates before integrating the data into the warehouse.
3. **Receipts Status Imbalance:** There are no receipts with an "Accepted" status, and the distribution of other receipt statuses is uneven. This could affect the reliability of our analytics and predictive models. It would be ideal to gather more varied data to ensure comprehensive analysis.
4. **Receipt Item List Storage:** Currently, the receipt item list is stored within the receipt data itself, which adds processing overhead when trying to normalize this into a separate table. Would it be possible to obtain this information in a separate JSON file to simplify the data processing?

**Optimizing for Analytics**

In designing the data warehouse, it’s crucial to keep in mind the types of queries and reports we expect to run most frequently. By optimizing the data model for common analytical queries, we can ensure faster processing and minimize resource consumption.

**Conclusion**

Once we address the above concerns and implement the necessary changes, I believe the data warehouse will significantly enhance our ability to generate valuable insights and enable better decision-making.

Please feel free to reach out if you have any questions or if there’s anything you'd like me to clarify. I look forward to your feedback.

Best regards,  
Chandana Maruri