Subject: Fetch Rewards- Analysis

Dear Stakeholder,

I hope you are having a great week. I am reaching out to update you regarding some recent insights and data quality issues I have come across while going through our company's datasets.

To start with, I am referring to the receipts, users and brands datasets. Here are the concerns I had regarding these datasets:

1. While JSON is an ideal file type to store the data in, considering the unstructured nature of the data, we would need to build a data pipeline to perform essential preprocessing such as normalizing the nested columns, cleaning the columns to be in the correct format, on these datasets and store them in a data warehouse in a normalized and relational format making it easy for our analysts to work on.

2. There is a great prevalence of null values and duplicate rows in the datasets. This can have a major impact on our analysis and the decisions we make as a company. While null values can be interpolated using different techniques, we need to understand what's causing these values to go missing and solve it at its root. Regarding the duplicate rows, this can have a significant impact on our processing times, starting from data ingestion to analysis, presence of duplicate rows can significantly tax our systems and slow down our processes.

3. There seems to be no clear relation between many columns in these datasets. For example, in the brands and the receipts datasets the only possible columns that seem to be related are cpgId in the brands dataset and the rewardsProductPartnerId in the receipts dataset. Having columns that more accurately represent the relation between the datasets can be very beneficial for the company.

These are some of the problems I have come across so far. With your permission I would work towards coming up with solutions for these problems by our next monthly meeting. Additionally, I have run SQL queries answering all the questions you had during our previous conversation. I hope these queries accurately capture the answers you are looking for.

I am attaching all the necessary files here for your reference. Please let me know if you have any questions or need additional data from me. Looking forward to hearing from you.

Thank you.


Regards,
Satya Sai Koppalu