I trust this email finds you well. Following a comprehensive review and analysis of the data, I have identified a few areas that require clarification and further information to enhance our data modeling efforts. Your insights on the following questions would be invaluable:

**1. Users File:**

The presence of a "states" field has caught my attention. Could you kindly confirm whether this data is exclusively for the USA region? If not, is it possible to provide the corresponding country code along with the state code information?

**2. Brands File:**

I am seeking more information on the meaning of "cpg" in the Brand files and an overview of what the CPG collection entails. This clarification is essential for structuring the CPG table effectively.

**3. Receipts File:**

While examining the Receipts data, I observed that the "rewardsReceiptItemList" attribute, which contains a list of items on the receipt, has several sub-attributes. Could you specify which attributes of "rewardsReceiptItemList" should be included in the data model, along with their definitions? We would like to capture only the attributes that hold significant value to our business.

Furthermore, during my analysis of the data, I noticed an issue with the dates in both the Users and Receipts files. None of the date fields, such as dateScanned, purchaseDate, and finishedDate, seem to be in the correct format. If my assumption is correct, could you please ensure that the files are sent with the correct date format? If the dates are indeed accurate, could you guide me on how to utilize the date fields to derive the correct format?

In addition, for optimal utilization of data, I propose including the following data points in the Users and Receipts files:

* Name, gender, age, and address for better insights into customer purchase patterns and trends.
* Another crucial data point for driving ROI would be the purchase location mentioned in the receipts. This information can aid in identifying the best-selling locations, enabling Fetch to establish more partnerships in those high volume areas.

Looking ahead, considering Fetch's growth and increased data volume, I recommend initiating discussions about migrating to a cloud data lake solution such as ADLS on Azure or S3 on AWS. This move would address potential scaling and performance issues, with the added benefit of leveraging Parquet file format for efficient handling of large datasets. Azure Synapse Analytics, with its Spark capabilities, could serve as an ideal platform for processing and analyzing data, facilitating faster insights for our analysts and data scientists.

I hope you find these findings and proposals aligned with our long-term goals. Your input on these matters and consideration of the cloud system proposal would be highly appreciated.

Thank you for your time and collaboration.

Best regards,

Sandeep Rani