

Fourth Task: Communicate with Stakeholders

Subject: Data Analysis and Description

Dear Stakeholders,

I hope this message finds you well. After conducting a thorough Exploratory Analysis of the item, users, brands, and trans data, We want to share some information with you about our recent data analysis. We have been analyzing the record data of customer receipts to better understand consumer behavior and preferences. Specifically, we have been looking at the top brands by receipts scanned, and comparing the spending and purchasing behavior of receipts with different rewards Receipt Status.

In the process of analyzing the data, we discovered some quality issues of the dataset, which need to be addressed. We found some missing or incomplete data entries, which might lead to inaccurate results. For example, the percentage of duplicate values in the 'OID' column is nearly 90% in the "users" table, and there are 8 columns in the "trans" table nearly 50% of values are null. We also found some inconsistencies in the data format and need to standardize it for better analysis. Furthermore, the lack of unique constraints hindered us at the beginning of the analysis.

To resolve these issues, we implemented a data-cleaning process to ensure the completeness, accuracy, and consistency of the data. In addition, we create unique constraints for each dataset and add foreign keys in each dataset to create the relationship between them. In this case, we ensure that the data in each table is consistent and accurate.

In addition, to optimize the data assets we are trying to create, we need more information on customer demographics and purchase history. Especially those missing and incomplete data. This information will improve the efficiency and accuracy of analysis, helping us create more targeted marketing campaigns, and improve our product offerings.

As we move toward production, we anticipate some performance and scaling concerns. We plan to address these concerns by optimizing our queries, improving our database indexing, and implementing caching mechanisms.

If you have any questions or concerns, please don't hesitate to reach out. We are excited about the potential of this data and look forward to sharing more information with you.

Best regards,
Group 5