Subject: Data Analysis Summary: Key Findings and Next Steps

Hello Manger,

I've completed an analysis of the data from our recent investigation and wanted to summarize the key findings, trends, and outstanding questions that need resolution.

Key Data Quality Issues

1. **Sales Data Completeness**: The FINAL_SALE column contains a significant number of blank or invalid values, which reduces the accuracy of any sales-related metrics. We filtered these out to proceed, but it's unclear why they exist.
2. **Inconsistent Date Formats**: The PURCHASE_DATE field required cleaning due to inconsistent formats, which could lead to errors in time-based analysis.

Interesting Trend

From our analysis, we identified **Fetch's power users**—those contributing significantly to our platform:

- Coco-cola brand has the most number of sales, while Zacca has 0 sales.
- Age groups like **25-34** or **35-44** might contribute the most to transactions or sales.
- Dove and Trident make it list of Top 5 in both a) user who are above 21 and b) user who has an account for atleast 6 months.
- Certain brands might perform better among specific age groups or account age brackets
- average activity. This suggests a small group of highly engaged users drives significant value.

Additionally, year-over-year growth in the number of scanned receipts highlights Fetch's increasing engagement, with a **30% YoY increase** for 2024 compared to 2023.

Outstanding Questions

1. **Sales Data Issues**: Can we confirm why many FINAL_SALE values are blank? Are these errors in data entry, missing data, or intentional (e.g., free transactions)?
2. **Receipt and User Matching**: There are a few discrepancies in the linking of receipts to users. Could we validate the integrity of USER_ID mappings across datasets?

Request for Action

1. **Data Validation**: Assistance from the data engineering team to understand and address the data quality issues, particularly for the FINAL_SALE column.

2. **User Segmentation Feedback**: Input from the product team on refining the definition of "power users" based on business goals. Should the thresholds for receipts or sales be adjusted to align with our strategic priorities?

Let me know if we can set up a brief sync to discuss next steps or if there's additional data context that could help improve our findings.

Best regards,

Shankar R

Sr Data Analyst