Subject: Data Quality Issues and Insights from Fetch Application Data

Dear John/Jane Doe,

I hope this message finds you well. I am reaching out to provide an overview of some key findings and challenges regarding the Fetch application data, as per our recent discussion. After a preliminary quality analysis, we noticed that the data is mostly clean and structured well. The ER diagram helped understanding how the data is connected. There were, however, issues with the quality of the individual data sets:

A. Products Data
1. Null/Missing fields, especially the barcode field. This could be problematic because this field is the **foreign key** that connects to the Transactions table
2. Number of items in the top three major categories aggregate to more than 99% of the total products available in the data set. This could lead to under-representation of other categories in any targeted analyses.
3. Missing actual manufacturer or brand information. This also could potentially hamper any brand performance analysis.

B. Users Data
1. Missing birth date values might make it difficult to perform generational analysis to understand the consumer demographic by age.
2. Sign-ups the top 5 states make up close to 40% of the data, which could skew state-based demographic analysis and indicate potential regional disparities in product awareness and/or adoption.
3. There gender options contain redundant values which could cause confusions in DEI analysis, with choices such as:
   i. *prefer_not_to_say* and *Prefer not to say*,
   ii. *non_binary* and *Non-Binary*
   iii. *not_listed*, *not_specified*, and *My gender isn't listed*

C. Transactions Data
1. Similar to Products data, there are missing values in the barcode field which could disrupt the foreign key connection between the Products and Transactions tables per the ER diagram.
2. Over 50% of the total 50000 transaction records are duplicates, which could lead to unreliable results for any aggregate analysis.

Going past these data quality issues, we uncovered several noteworthy trends that offer valuable insights into Fetch's market performance.

Fetch saw exponential growth in user sign-ups from 2016 to 2019, with a massive **820%** surge in 2017. The momentum of growth kept up through 2021, but started slowing, and by 2023 and 2024, sign-ups dropped by close to 40% and 25%. Even with the recent decline, Fetch's rapid growth in its early years shows just how much it connected with users and took off.

Based on product utilization, we identified two key categories of power users.
1. App Engagement - Users who scanned at least 7 receipts (5x more than the median user). We identified 6 power user accounts in this group.
2. Purchase Spend - Users who scanned receipts worth over $100 to $1,000 (top 0.1% of users). We identified 18 power user accounts in this category.

To dig deeper into the analysis, we'll need to address the data quality issues mentioned earlier. Specifically, fixing the missing values, cleaning up the inconsistent gender entries, and removing duplicate transaction records will make the dataset more reliable. It would also be helpful if we could get a bit more clarity on certain data points (like complete manufacturer and brand details) to ensure everything is in order.

Once we get these pieces sorted, we'll be in a much better position to establish stronger data validation rules, which will help us provide more accurate insights into user behaviors and market trends.

Thanks for your time, and we look forward to your guidance on how we can move forward!

Best,
Niha

Niharika Chunduru (She/her)
Product Data Analyst
Fetch Rewards
+1 234-567-8901
email@fetch-rewards.com