Hello Product Leader,

We’ve been analyzing the product, user, and transaction data and wanted to provide and update on our findings and what issues we have run into that can potentially be fixed going forward.

One of the main issues with the dataset we are working with is in the product and transaction subsets. The Barcode column is a crucial connecting piece between the datasets and being able to run a full analysis on key metrics such as brand sales, types of products purchased, and overall sales across different industries.

There are other columns we are missing data for such as language, gender, state that is hindering some analysis and grouping among our target customers however the amount missing on gender and state are insignificant to cause a major impact on the analysis but would be ideal to have a complete picture. Other issues within the data relates to how the data is stored, we should be using a consistent date across our datasets for easy comparison and avoiding unnecessary conversions.

From our analysis we were able to draw some meaningful metrics for our top brands showing users who are over 21 have brand loyalty to them, furthermore we can break this down into generational subsets that will allow us to cater campaigns specifically to our main target audience or a desire to attract a new one. For example for users over 21 we found these brands to have the top 5 receipts COCA-COLA, ANNIE'S HOMEGROWN GROCERY, DOVE, BAREFOOT, ORIBE.

Some outstanding questions on the data would relate back to the missing values and if those are expected or an indication of a larger issue in our data collection. Fixing this would greatly increase the effectiveness of our analysis. Standardizing our dataset with naming conventions and data types would also be a large help in allowing us to do more analysis and less data cleaning. For some numeric columns such as Final Quantity purchased we are receiving values such as “zero” while obvious what this means it is causing an extra step in the processing to rectify this issue.

I look forward to discussing the findings more in depth with you and going through actionable solutions in our campaigns to better target our customers with more relevant products according to their age, gender, and locations. As well as possible other fields we should consider capturing for even more in depth analysis on who our power users are.

Thanks,

Brennan O’Brien