

MSDS 430 - Final Project - Alison Au

Comparing Instacart Frequent vs. Infrequent Customers

Introduction

The goal of this project is to compare buying habits of the most frequent versus less regular customers of online grocery store Instacart. This can help guide strategy and marketing tactics to drive people to become more regular users.

My idea for this project stems from my general interest in consumer behavior and food. I was also curious to look deeper at this newer industry of e-commerce grocery shopping. Lastly, this goes hand-in-hand with my career in marketing/advertising. We continually use data to determine the most effective advertising messages for different audience segments.

This data is sourced from Kaggle:

<https://www.kaggle.com/c/instacart-market-basket-analysis/data>

Instacart had publicly released a sample of over 3MM orders in 2017 from more than 200K users. For each user, between 4 and 100 of their orders are provided.

Variables Explored:

- Order volume: used to segment users into high/medium/low-frequency customers
- Product volume
- Products purchased
- Reordered products: 1 if this product has been ordered by this user in the past, 0 otherwise
- Day of the week the order was placed on
- Hour of day the order was placed on
- Days since the last order, capped at 30

Solution Process

The analysis follows an EDA approach. The datasets are first merged into two main dataframes, one with user and order-level data and another with order and product-level data. Note that these are not combined into one large dataset, as this is not necessary and caused the code to take much longer to run.

Next, the distribution of users by order volume is pulled. Users are then segmented into equally sized buckets of high, medium, and low users based on orders. This resulted in the following breakout:

- Low users: 4-7 orders
- Medium users: 8-16 orders
- High users: 17-100 orders

These segments are then sliced by the various variables noted above in order to compare the audience segments and understand the overall customer base.

Results

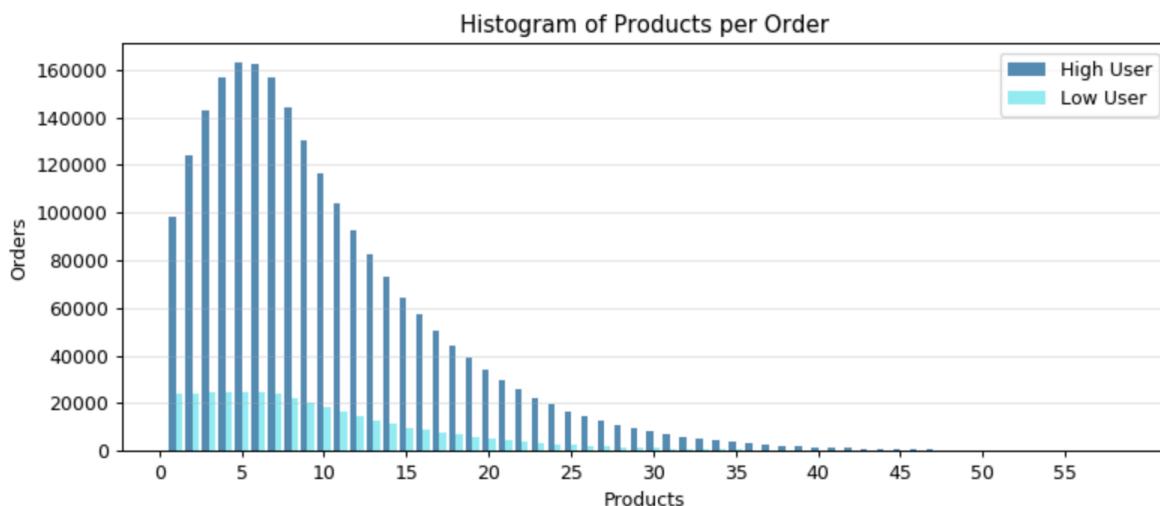
Overall, the two customer segments have similar buying habits, aside from order volume and frequency of orders. They purchase the same amount of items per order, have similar tastes in fresh produce and organic products, and shop once a week.

However, a key difference is that high-frequent customers reorder the same products significantly more than low users. Further exploration can be done beyond this dataset to see if other factors drive this, e.g. cost, quality of produce, or delivery times. Targeted ads can also be run showing products that low customers have purchased before. These ads can more heavily run on Sunday afternoons, when low users typically shop on Instacart.

Some of the main results are shown below.

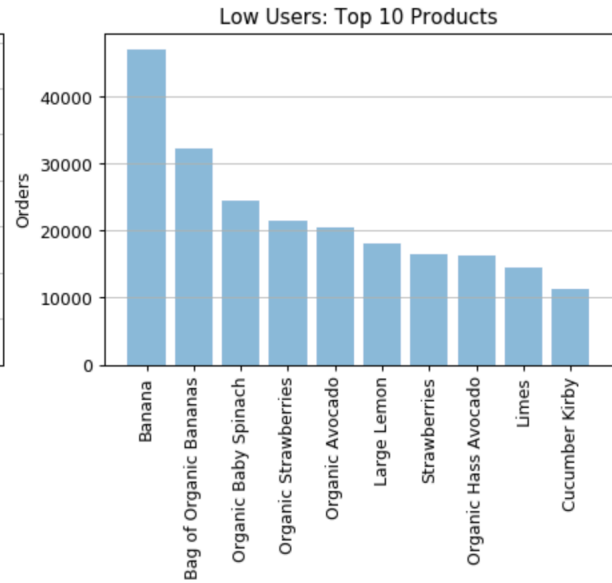
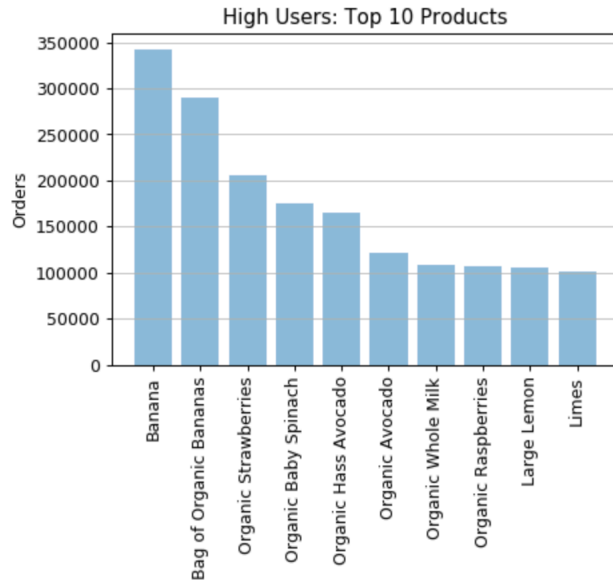
Products Per Order

On average, both high and low users purchase 10 products per order. This does become a bit more differentiated when looking at the distribution. High users most commonly buy 5-6 products, while low users do have more variance with 1-7 products per order. This shows there is a small opportunity to grow the number of products low users buy per order.



Top Products

Top sellers are fresh produce—bananas, strawberries, and spinach in the top 3 or 4 for both segments. Organic products are also popular. The Instacart audience, both high and low customers, seem to be a more health-conscious consumer.



Reordered Products

A main difference between low and high users is the percentage of products ordered that had been bought previously (31% for low users vs. 67% for high). This shows opportunity to encourage users to re-buy what they already have before.

		reorder_pct
freq_cust	reordered	
low	0	69.5%
	1	30.5%
med	0	53.7%
	1	46.3%
high	0	32.6%
	1	67.4%