

Milestone Report

Problem Statement:

Predict products mixes that will be included in the next purchase order by users. That includes products a user will buy again, try for the first time, or add to their cart during next session.

Benefits

Main audiences: Instacart e-commerce Team

By knowing the user's purchase habits and what user would like to purchase in the next order, Instacart can customize product pages to individuals by showing relevant products, which enables better conversion.

Secondary audiences include:

- **Instacart users:** Users will save time on browsing the needed products, and may find new products of their interests faster, which enhance user experiences.
- **Partner retailers:** Instacart can aggregate prediction of product purchases by store and send that to partner retailers for inventory planning purchases.

Dataset Description

Obtained from Kaggle, data are in form of 4 dataset

- Aisles: product subcategory (size: (134, 2)). Examples: instant food, energy granola bars, etc.
- Department: product department category (size: (21, 2)). Examples: bakery, produce, etc.
- Products: name and id (size: (49688, 4))
- orders: orders id, user_id, order number, order day of week, hours of day, days since prior order (size: (3.4M x 7))
- order_products__prior: product details among order in the “prior” eval set under orders dataset
- order_products__train: product details among order in the “train” eval set under orders dataset
- eval set “test” is the target variable of the problems

Cleaning and Wrangling

- the data has been cleaned by data provider, no missing records, and data are in the right data types

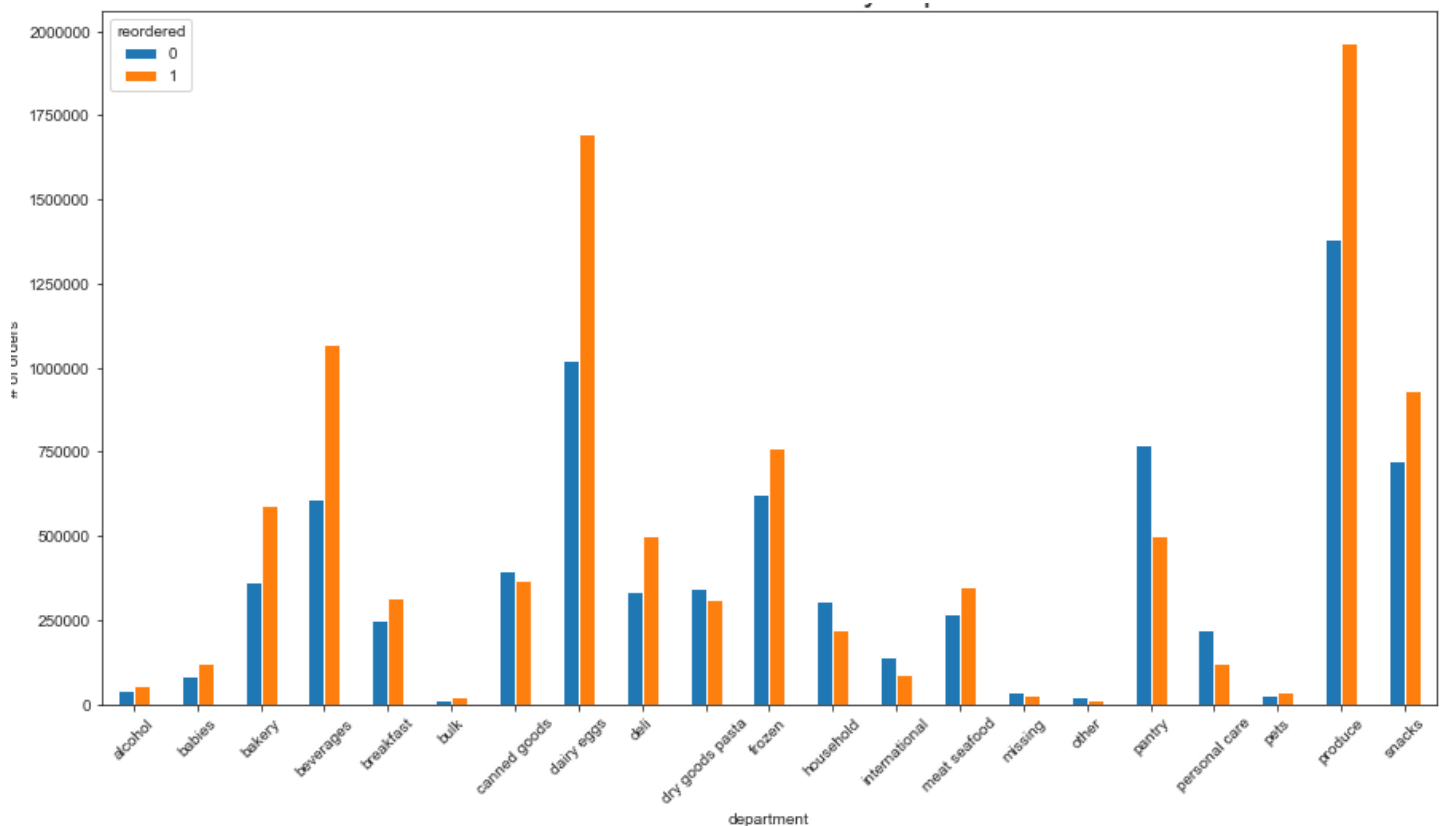
Exploratory Analysis

The goal is to discover features that may have predictive power to target variables (i.e. product id by user) through visualizing their historical pattern.

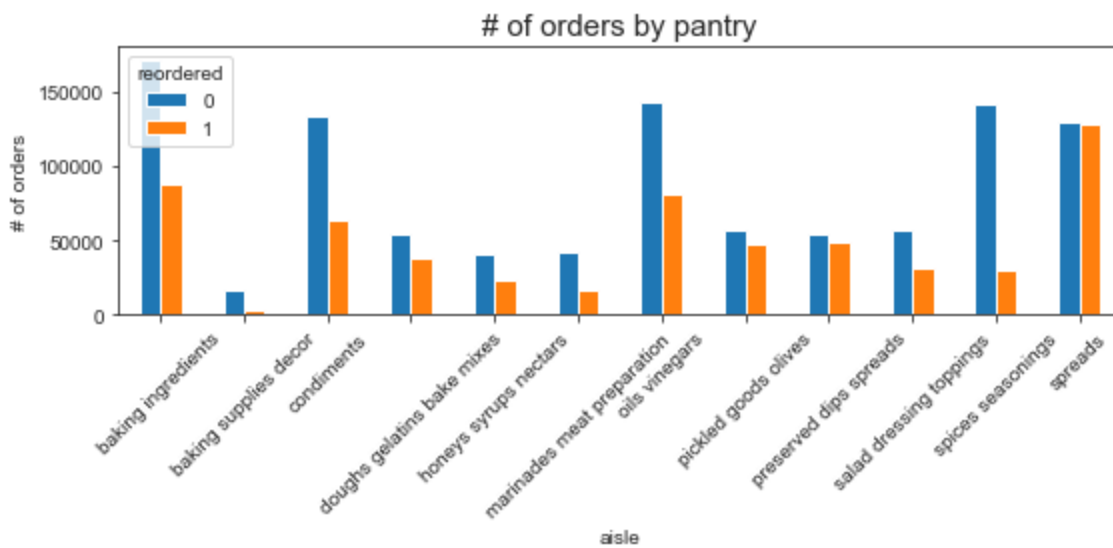
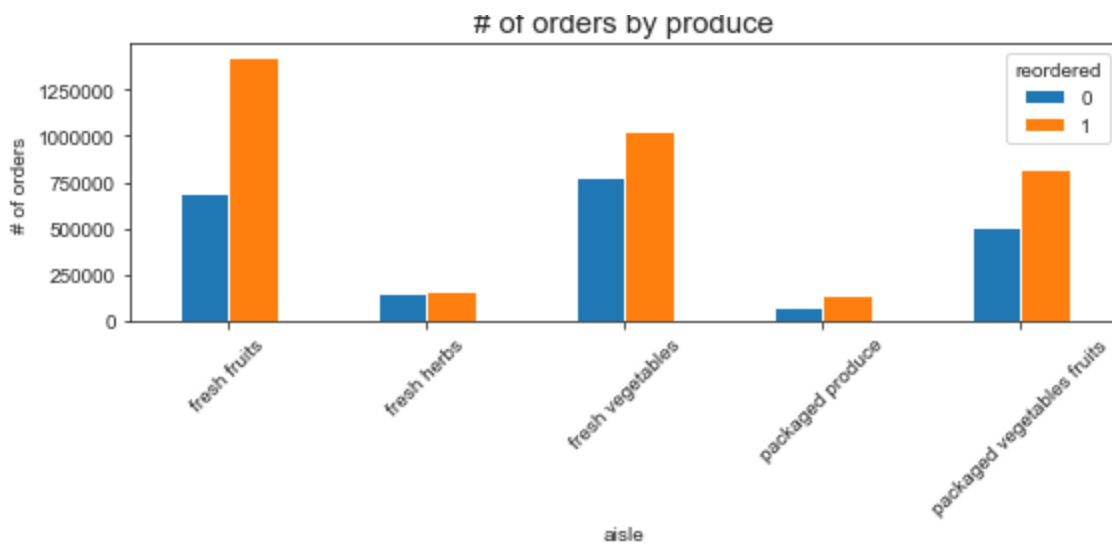
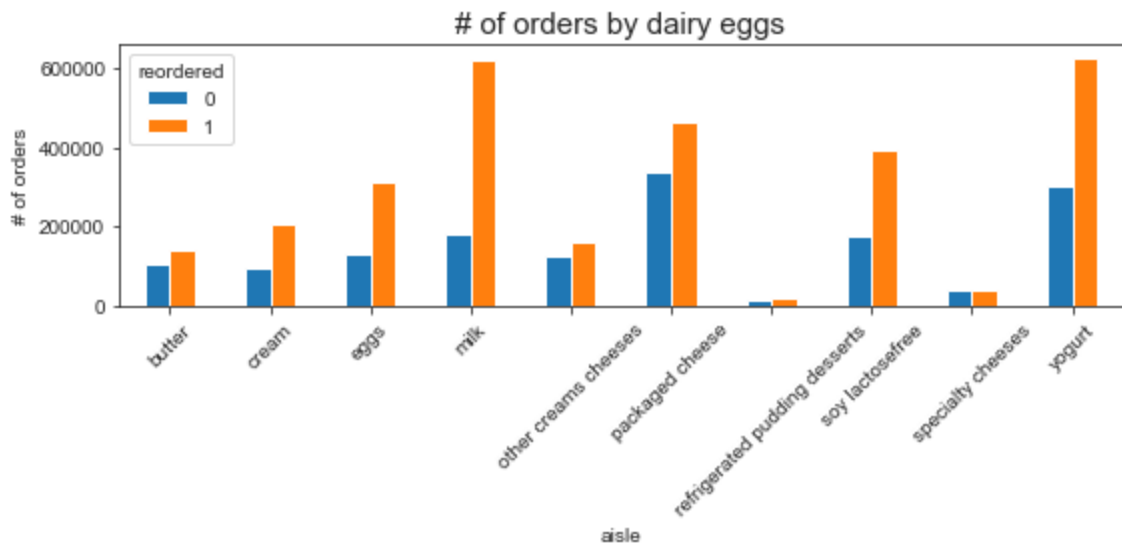
Areas of focus:

- What types of products have higher reorder frequency?
- Do users buy the same type product over time (i.e. does product mix by user stay relatively constant?)
- Would the day of the week (e.g. Monday vs Friday) influence the number/ type of product being purchased?
- When do users tend to buy from instacart (day of week and hour of day)?
- How many days are between each order?
- What's user attrition rate?

Order/Reorder Frequency by Department:

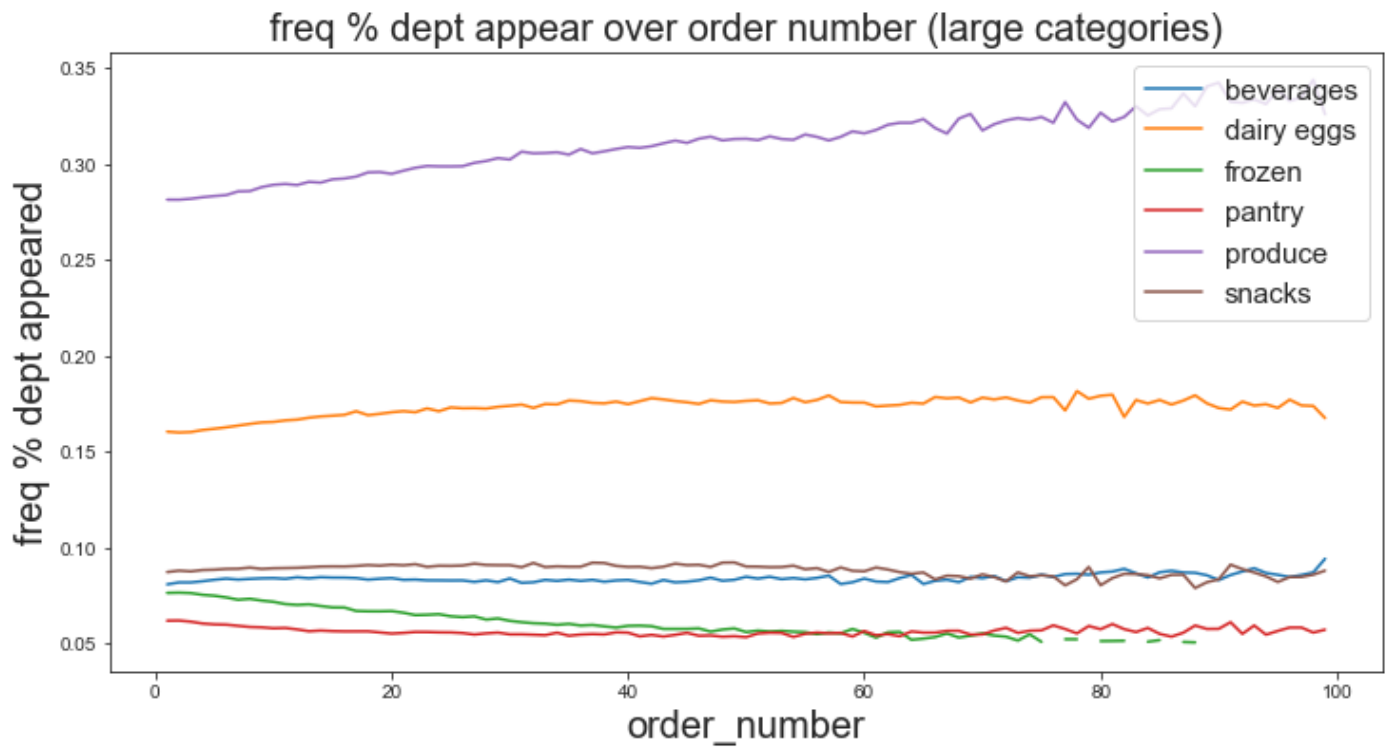


- At the department level, beverages, dairy and produce appear to have the highest re-purchasing rates. Pantry does not seem to be popular items for re-purchasing.



- Aisles within each department vary on re-ordering rates

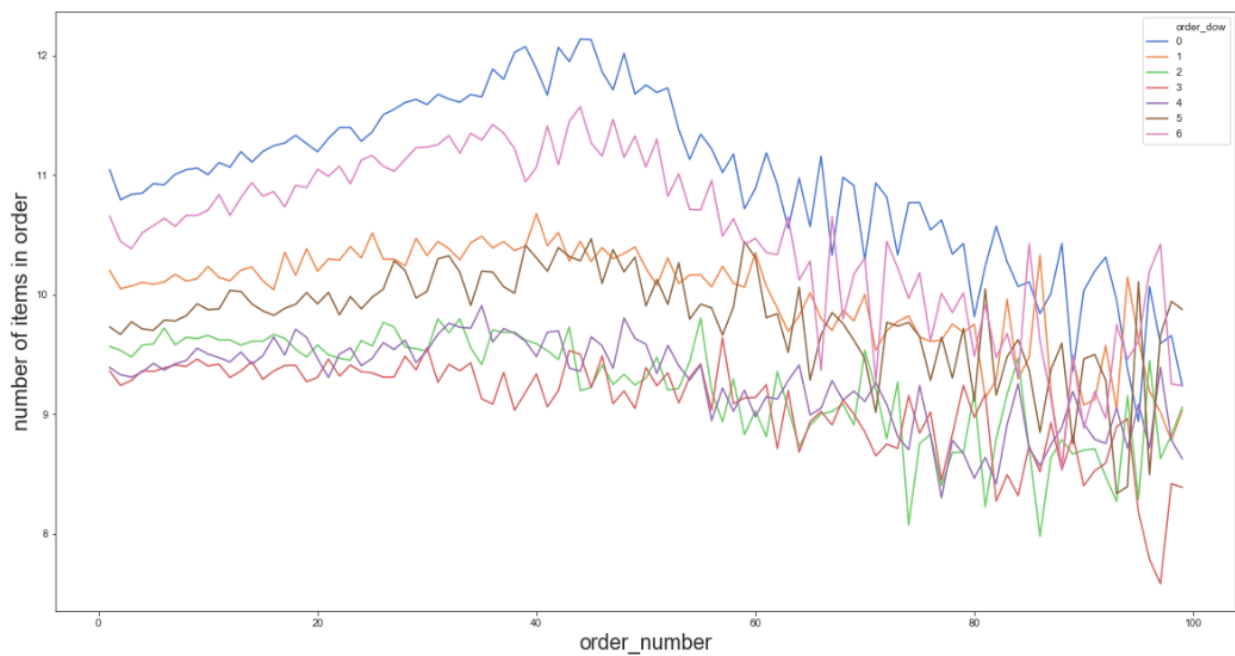
Order Mix % by Department



- Department mix appear to be stable over time

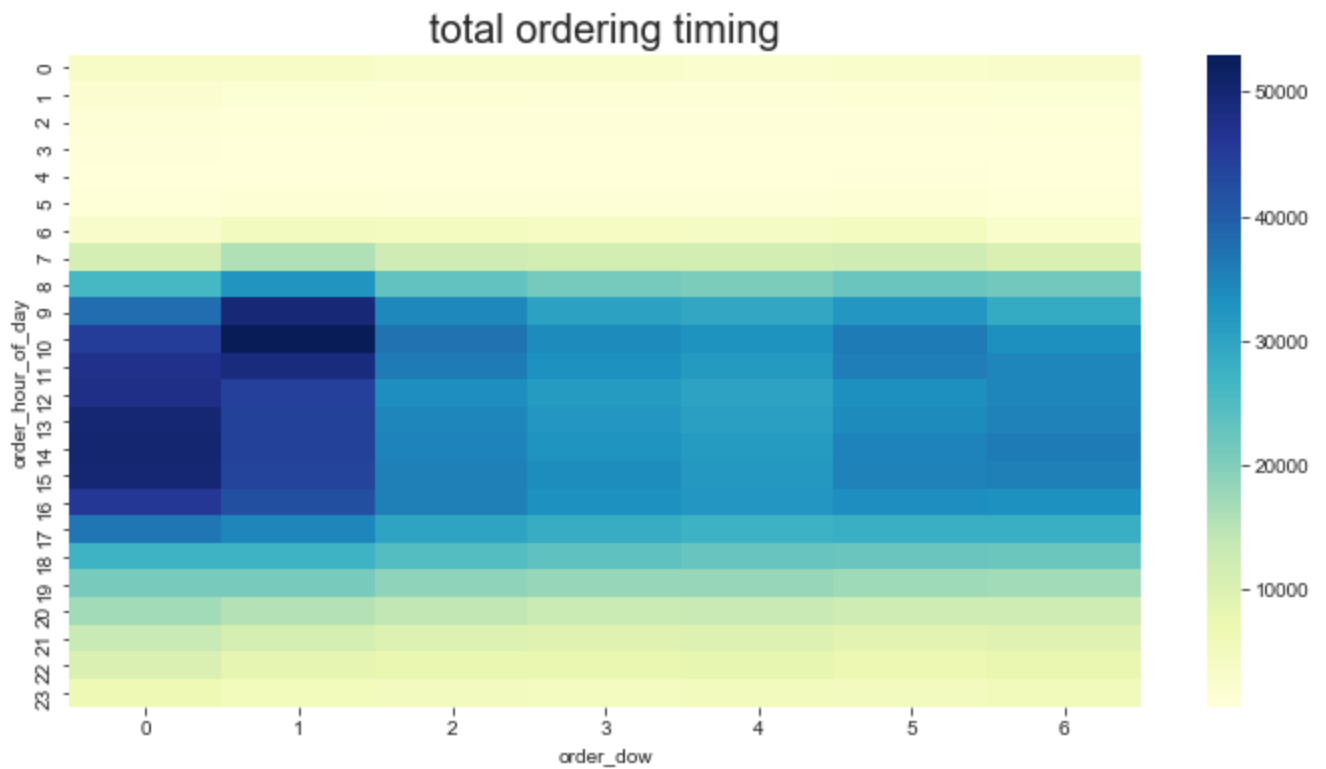
Day of Week

- Users tend to buy more items on Day 0 (Sunday) and Day 6 (Saturday). But overall orders are between 9-11 items



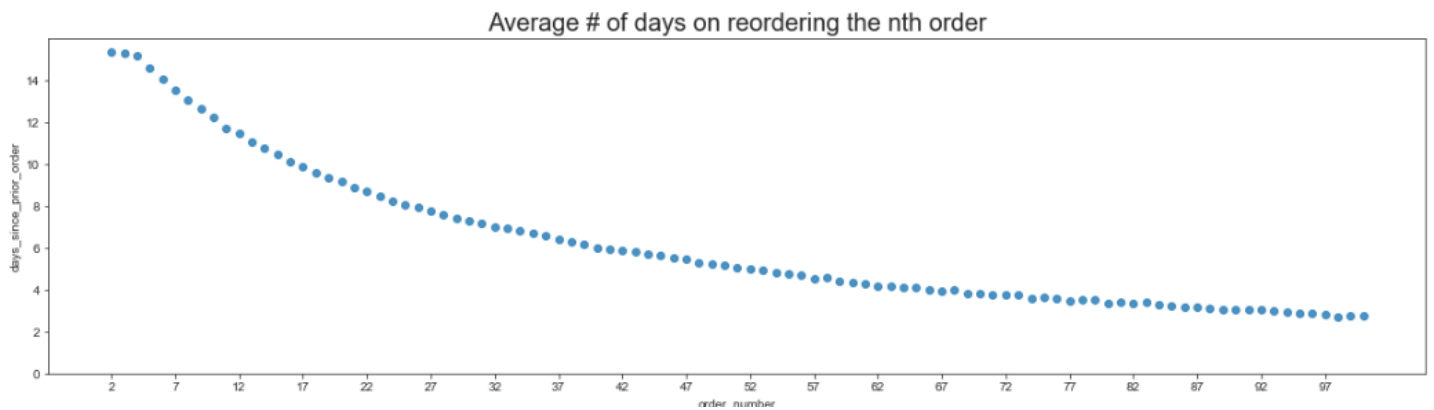
Hour of Day

- Generally, most orders are placed between 8am to 6pm



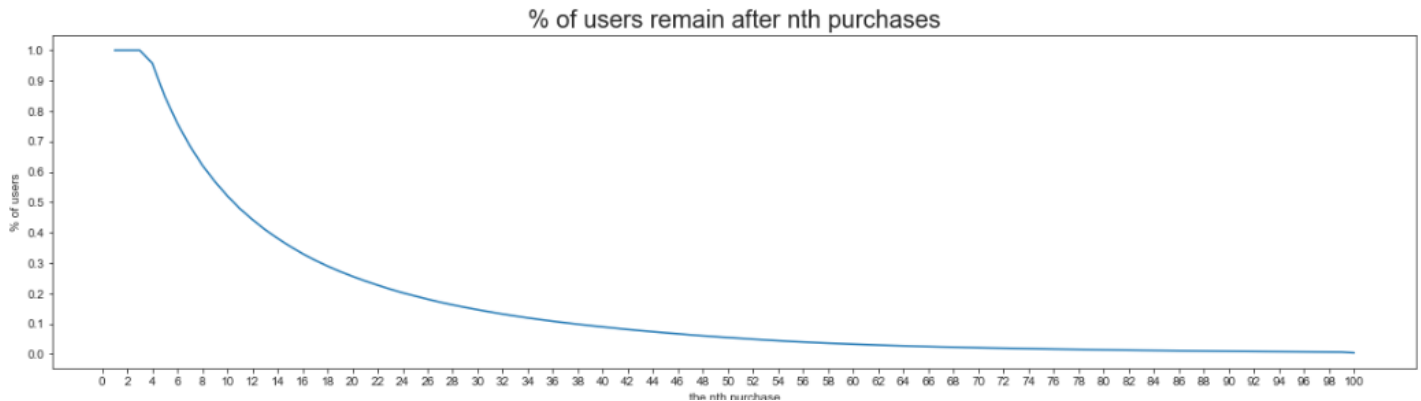
Order Lags

- For the first 3 orders, the day separation is about 15 days, and gradually decrease as the user re-order more often



User attrition

- After the 4th purchase, users attrition rates is at ~5% per additional one re-purchase. By the 10th purchase, only 50% of users remain using the services. About 10% of the users making 38th purchase or more



Summary

- Product categories have different reorder rates, among all beverages (water), dairy eggs (milk, yogurt) and produce (fruits) are most popular items for re-ordering. Pantry items are the least likely to be re-ordered
- Product categories mix does not vary significantly over time. Users will gradually trade off one item favor the other, but the reorder items appear to be similar to the previous order that user made
- Users tend to purchase on day 0 and 1, from 9am to 4pm
- Some products are being purchased on different days
- Most users who signed up will use the service 3-4 times, on average 15 days between each other, then users' attrition starts. Perhaps that's an opportunity to improve user loyalty. 10% of the users become very loyal to the service and purchase almost every week.
- In each order, around 9-10 items are added to carts.

Next steps will be to convert the above observation into features to predict the likelihood of a product to be re-ordered by the user.