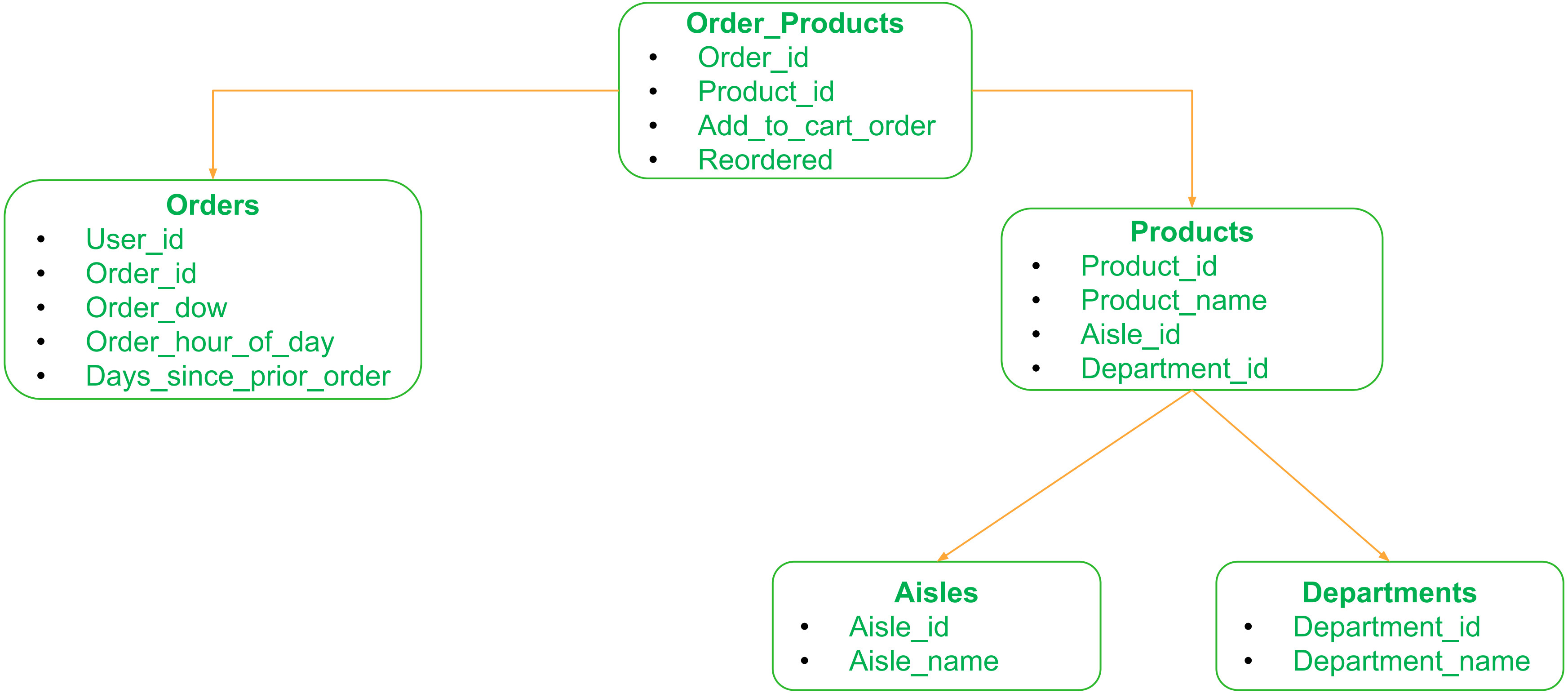
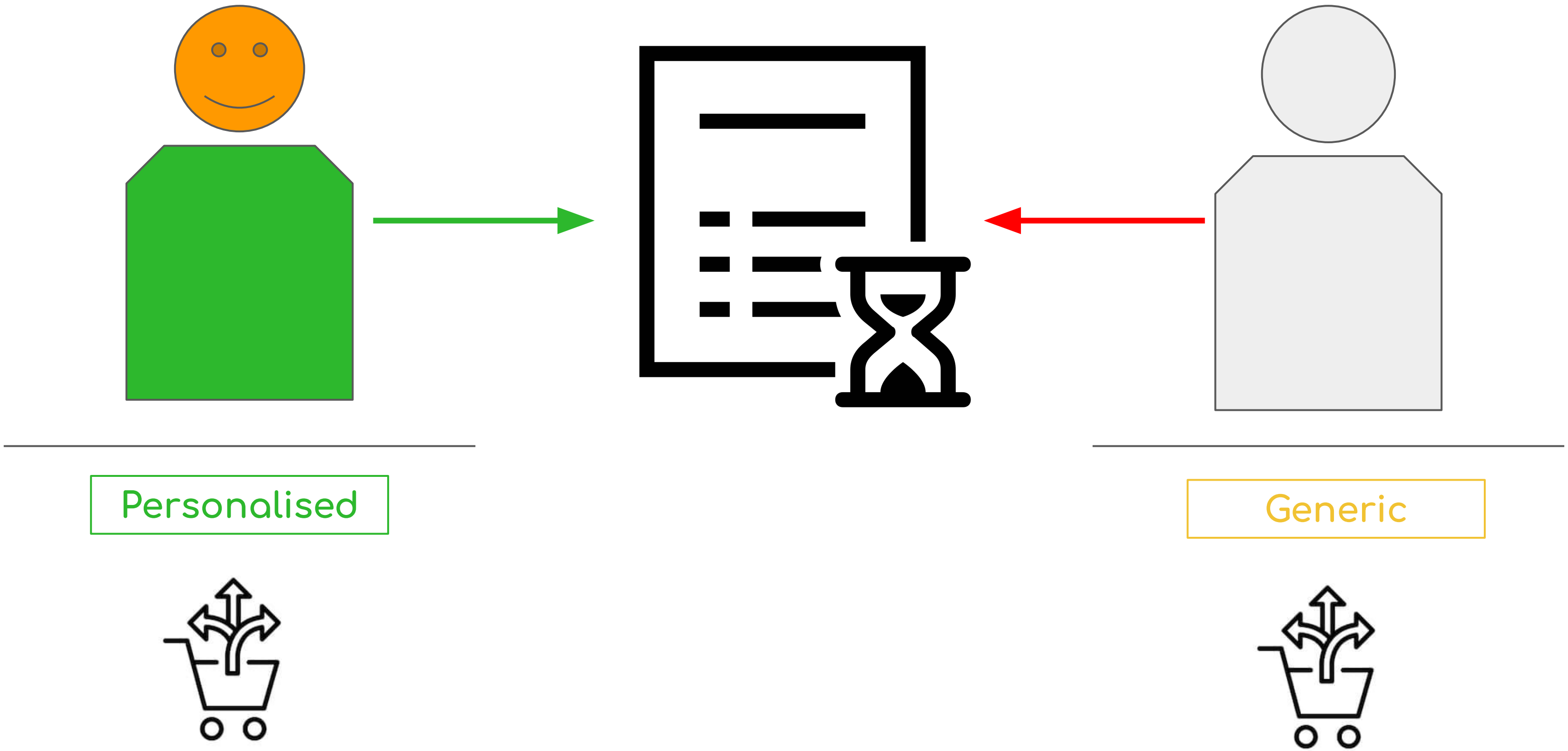
instacart

Group P- Anshika Ahuja, Apoorv Mehrotra, Arnav Deshwal, Jake Hill, Vikrant Vaidya

First, let's look at our data!

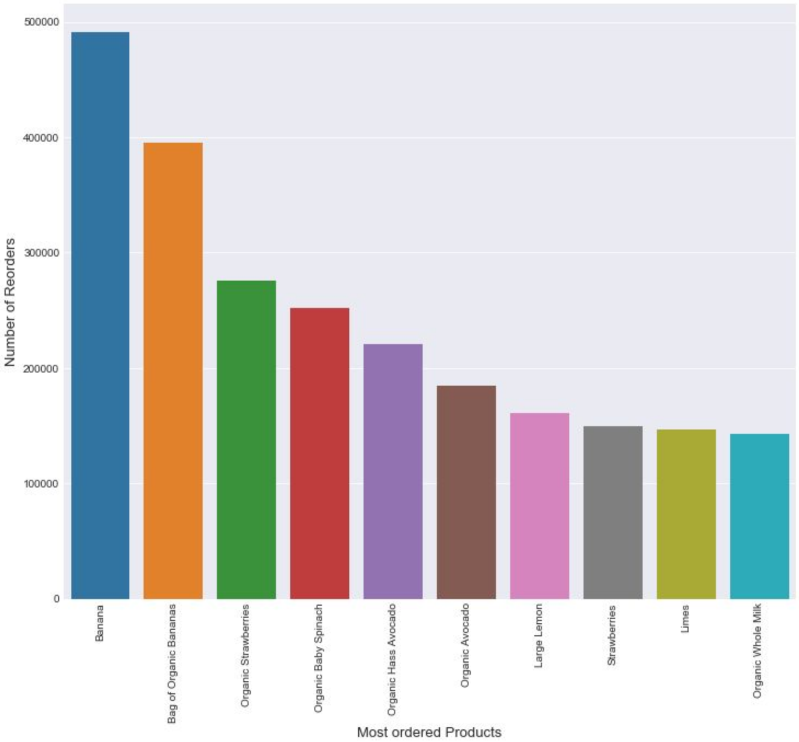


Visitors can either be old or new to the website

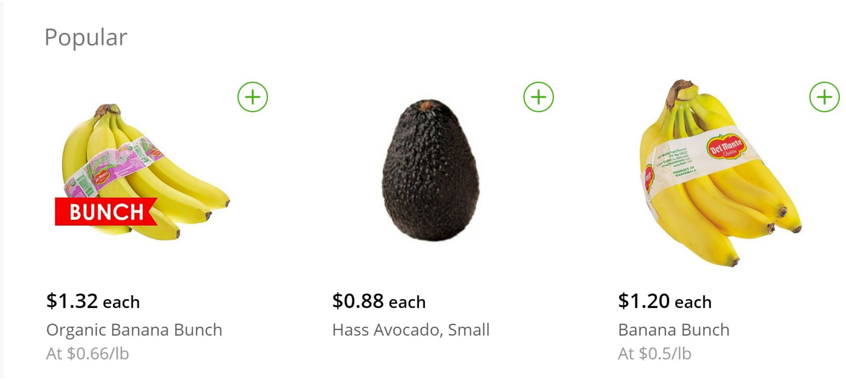


People are going Bananas!

2017



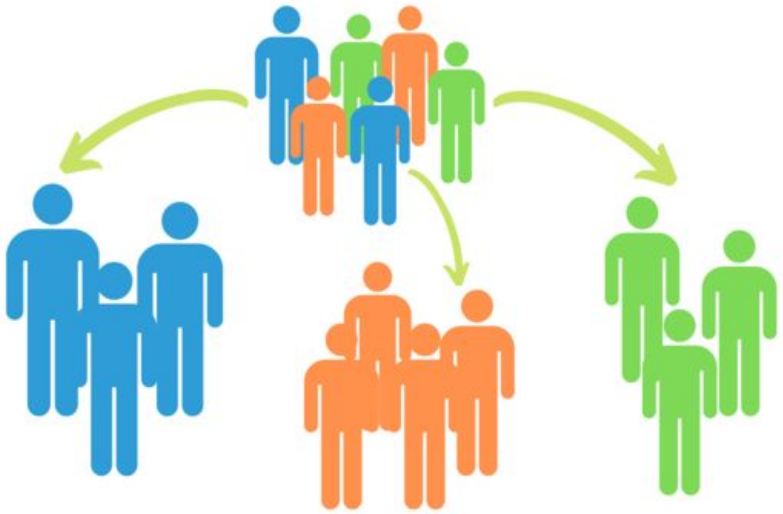
2019



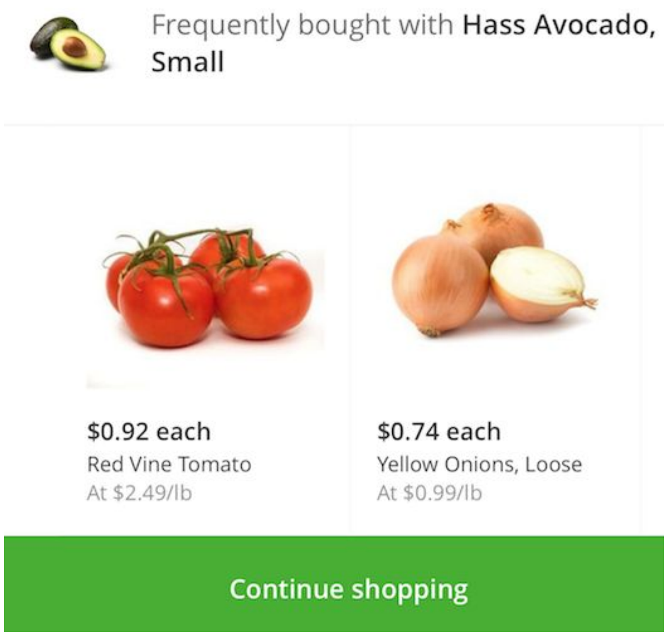
Recommend popular products to new customers

Divide, Conquer and Integrate!

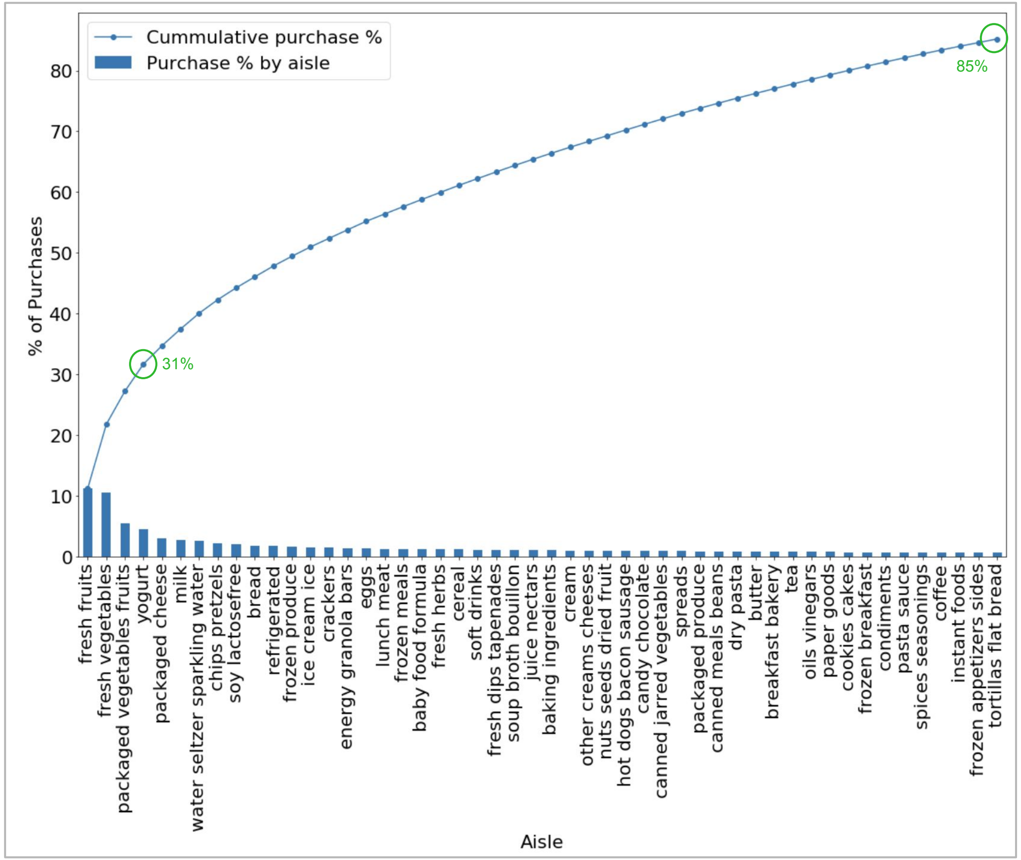
Customer Profiling & Targeting



Recommendation Systems



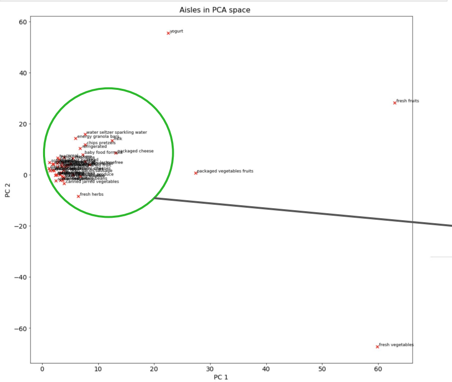
85% of all items purchased come from 35% of the aisles

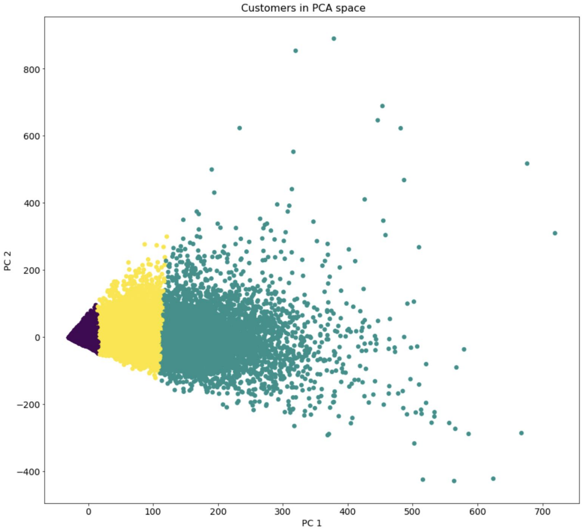


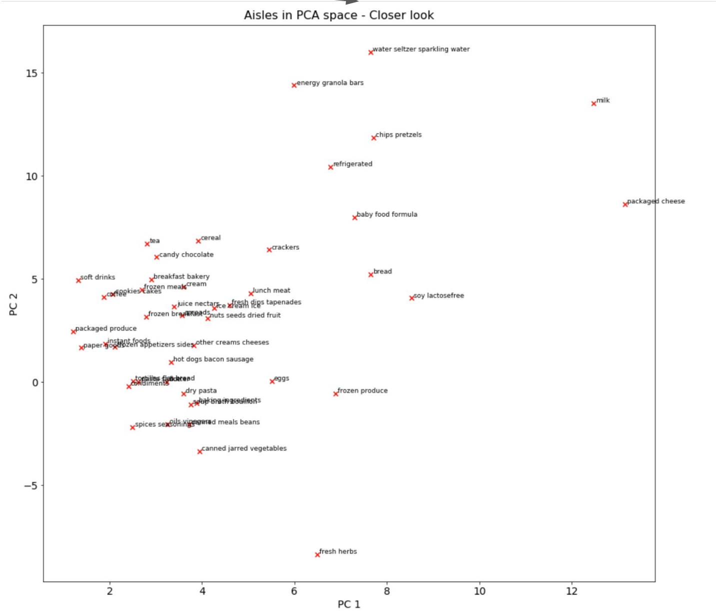
- 31% of all products purchased via Instacart are from 4 aisles - fresh fruits, fresh vegetables, packaged vegetables fruits, yogurt
- 85% of all products purchased via Instacart come from 50 aisles
- We will use these top 50 aisles to observe the behavior of our customers in order to segment them

KNN over PCA output provides well separated clusters

We use PCA to reduce the dimensionality from user behavior in 50 aisles down to 2 components that capture ~60% of the variance







Cluster Statistics

Cluster 0

% Customers: 81%

%Orders: 54%

Customer Stats

Median Orders: 7

Median Order Gap: 16

Median Basket Size (Variety): 6.8

Avg Basket Reorder Share: 50%

Cluster 1

% Customers: 4%

%Orders: 13%

Customer Stats

Median Orders: 56

Median Order Gap: 6

Median Basket Size (Variety): 13

Avg Basket Reorder Share: 77%

Cluster 2

% Customers: 15%

%Orders: 33%

Customer Stats

Median Orders: 29

Median Order Gap: 9.5

Median Basket Size (Variety): 11

Avg Basket Reorder Share: 68%

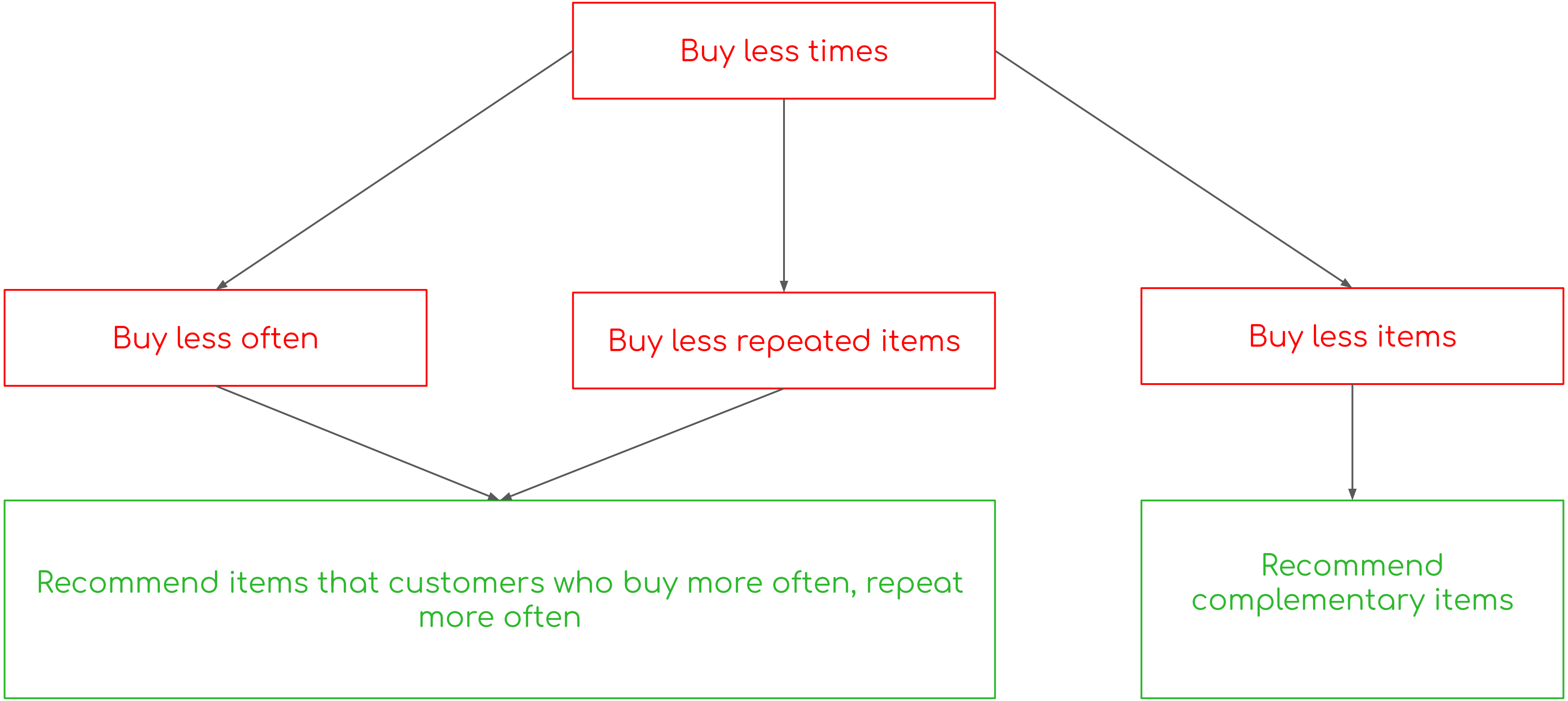
Most of the customers are buying less

Ideal Cluster		Target Cluster
1	Transaction/Customer	0.1x
1	Median order gap	2.7x
1	Basket Variety	0.5x
1	Basket Reorder Share	0.7x

Target Cluster Customers

- Buy less times
- Buy less often
- Buy less items
- Buy less repeated items

Recommendations are the key!



How we built our Recommendation Systems

Association rule based

$$\text{Rule: } X \Rightarrow Y$$

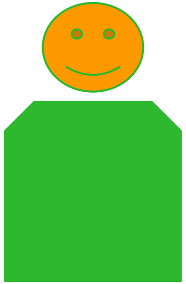
$$Support = \frac{frq(X,Y)}{N}$$

$$Confidence = \frac{frq(X,Y)}{frq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

Repeat based

1. Identified target and ideal audience

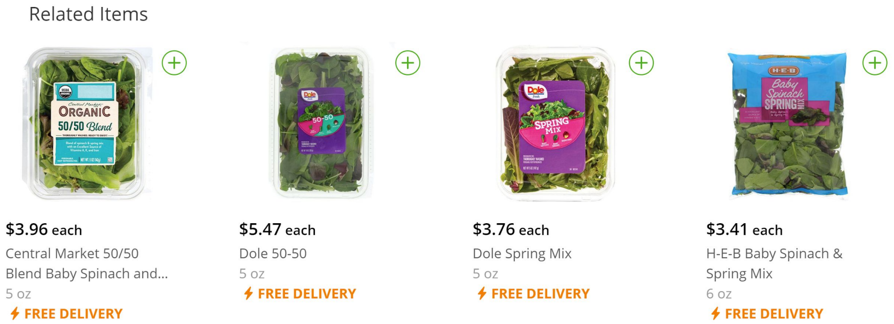


2. Identified similar items that are reordered more by ideal audience

Recommendation system 1: Apriori

- Identified cluster 0 customers as target, cluster 1 customers as ideal
- Cluster 0 customers have basket variety half of that of cluster 1 customers
- Recommend complementary items based on lift scores

Lift score = 5



Limitation: 31% of the transactions were coming from 4 aisles only - most of the products had very low support, and hence most itemsets had low confidence numbers

Recommendation Engine 2: Repeat Based Example

- Identified cluster 0 customers as target, cluster 1 customers as ideal
- For each cluster, for each item, calculate:
    - number of orders where a user ordered the product for the first time (order count)
    - number of orders where a user reordered the product (reorder count)
    - Heuristic: Reorder score = (reorder count)/(order count)
- For target cluster, for each product with low repeat, recommend from ideal cluster:
    - Products from the same department/aisle that have a higher reorder score than that product and reorder score > 1
- Final recommendations: Products from same department/same aisle, which ideal customers tend to repeat more often!

Recommendation Engine 2: Repeat Based

Target cluster Reorder score = 1.325

Ideal cluster Reorder score = 67

Target cluster Reorder score = 0.43

Ideal cluster Reorder score = 4.3

Ideal cluster Reorder score = 6

Ideal cluster Reorder score = 47

Target cluster Reorder score = **0.73**

Target cluster Reorder score = **0**

Future Scope

- Demographic similarity based recommender system:
  - If demographic data becomes available, find lookalike of each target cluster customer in ideal cluster
  - Recommend products that most similar users from ideal cluster buy most repeated/most penetrated items (penetration = # orders where products was purchase/# orders for user)

- Further segment customers based on demographics, purchase behaviour (with qty and $ data):
  - A/B test all 3 recommendations for different customers to see what works best for which type of customers
  - Create logic for composite recommendations based on A/B testing results for different clusters to productionalise

Summary - Recommender comparison

	Apriori	Reorder	User Based
Goal	Increase basket diversity	Increase highly repeated product purchases	Recommend products that similar people buy the most
Scope	Cross-selling	Alternate products to boost upselling	Cross-Selling
Pros	1. Improves basket diversity	1. Increases reorder chances 2. Easy to change level of recommendation (same aisle, same department)	1. Different way to look at similarity, recommendations based on 1-1 behaviour similarity assumptions
Cons	Scope limited to same aisle and high support products due to computational limits	May provide very different recommendations based on recommendation level department	Needs another model to estimate and verify user similarity accurately before recommendations can be made
Data	Available	Available	Not Available

Questions?

Thank you :)

Appendix

```
user_info.groupby('cluster')['avg_basket_size'].agg(['min','max','mean','median'])
```

cluster	min	max	mean	median
0	1.000000	58.25	7.664169	6.800000
1	3.666667	52.50	14.060886	13.068966
2	1.424242	46.68	11.619783	10.720000

cluster	days_since_prior_order
0	16.513949
1	6.277394
2	10.192695

Future Scope

- Target ideal cluster to identify basket size related metrics and try to increase number of customers who transact like them
- See demographics etc. for target cluster to see what works for them
- Look at the monetary data for each transaction on user level
- Analyze the quantity of each product bought in each order by different users to gain more insights into customer's buying patterns

FPM Results

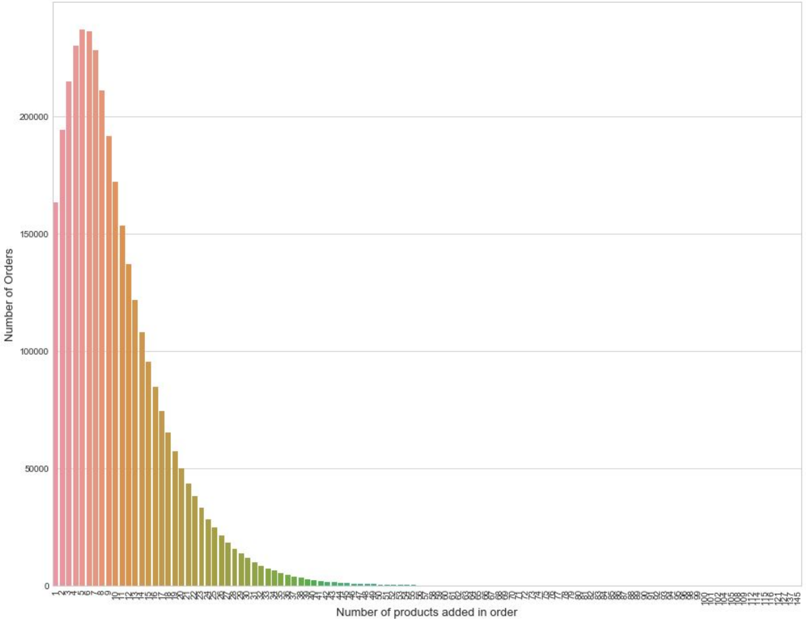
- https://drive.google.com/file/d/1q1g4Bv1nB25hxCj1efsK_Tv4KT7J00Iw/view?usp=sharing

Overview

- Exploratory Data Analysis
- Two Types of customers - New and existing
- New Customers - Recommend popular products
- Existing Customers are the Real Deal!!!
- Problem Statement: How to recommend products to the existing customers of Instacart in the most optimum way?
- Approach-
  - How do our customers behave? CUSTOMER SEGMENTATION
  - Which customers do we target? CUSTOMER TARGETING
  - What products do we recommend? RECOMMENDATION SYSTEM

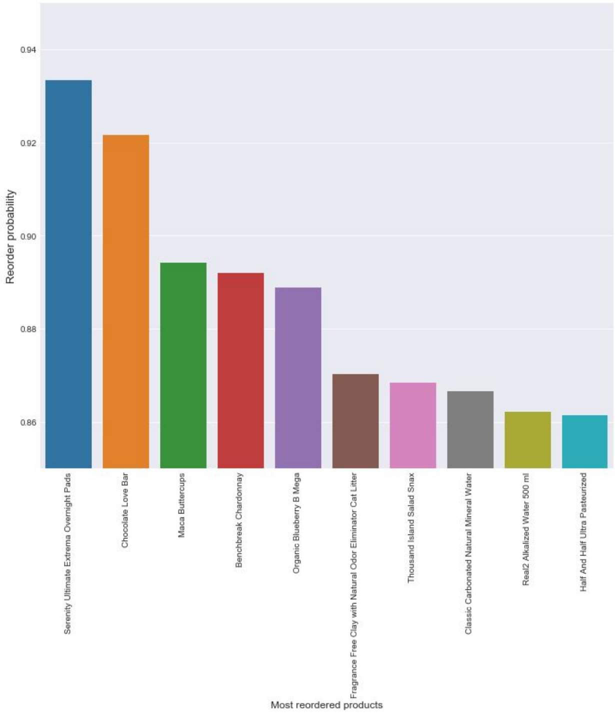
Exploratory Data Analysis

No of orders vs No of products in each order

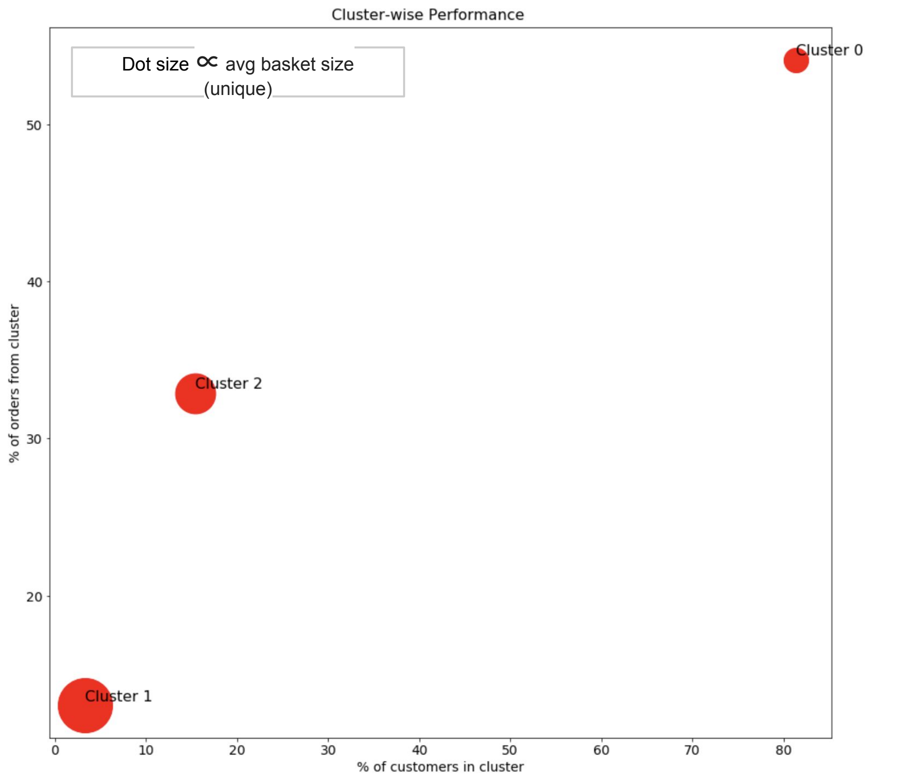


Most people order around 5 products

Top reordered products



Cluster 0 is a gold-mine for cross-selling

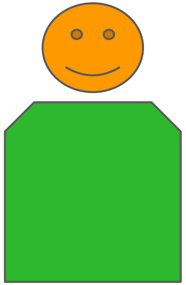


80% of our customers and 54% of our orders come from cluster 0, which has the lowest basket size among all 3. Since we do not have data for quantity, we conclude that Cluster 0 has the least diverse baskets among all 3 clusters.

Cluster 0 becomes the ideal population for cross-selling focus

How we built our Recommendation System

Identified target audience



$$Rule:\ X \Rightarrow Y$$

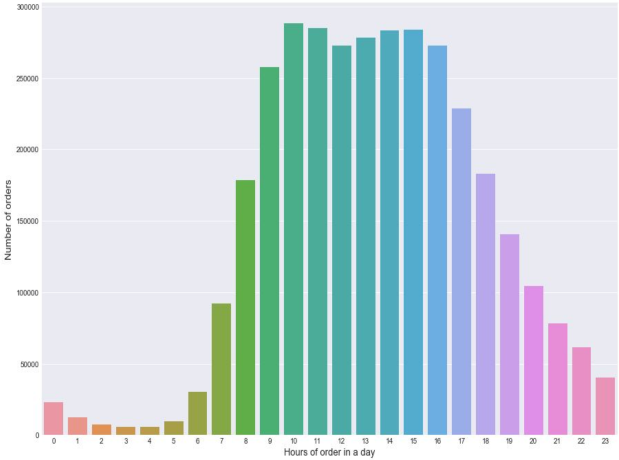
$$Support = \frac{frq(X,Y)}{N}$$

$$Confidence = \frac{frq(X,Y)}{frq(X)}$$

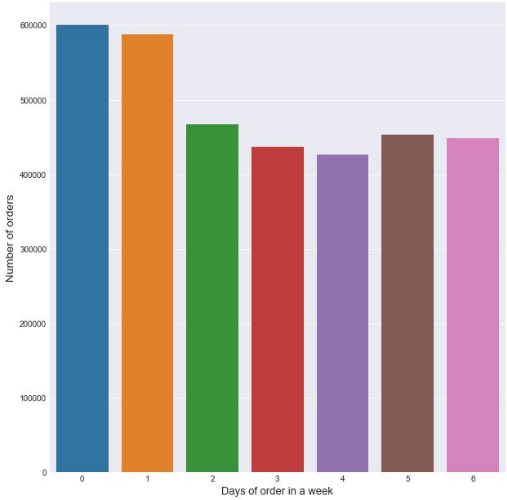
$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

Exploratory Data Analysis

No of orders vs hours of the day



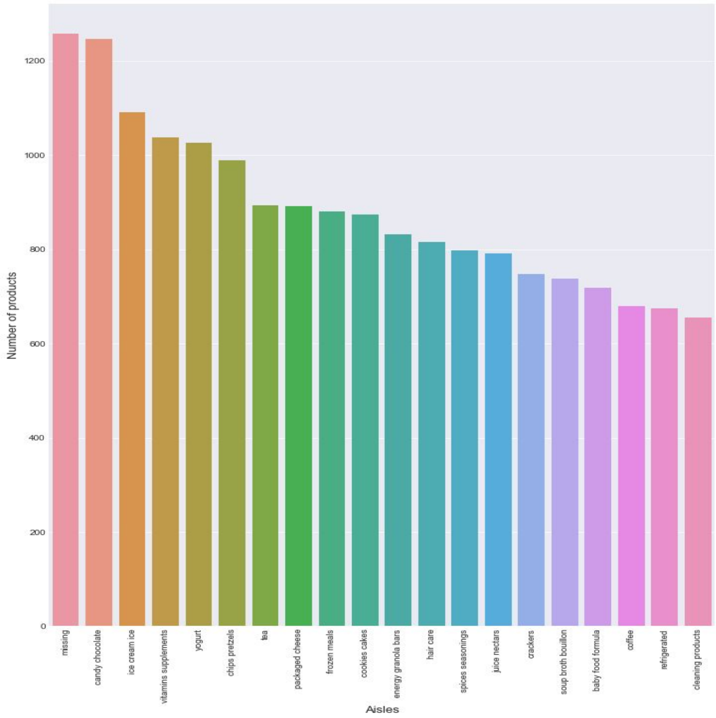
No of orders vs day of the week



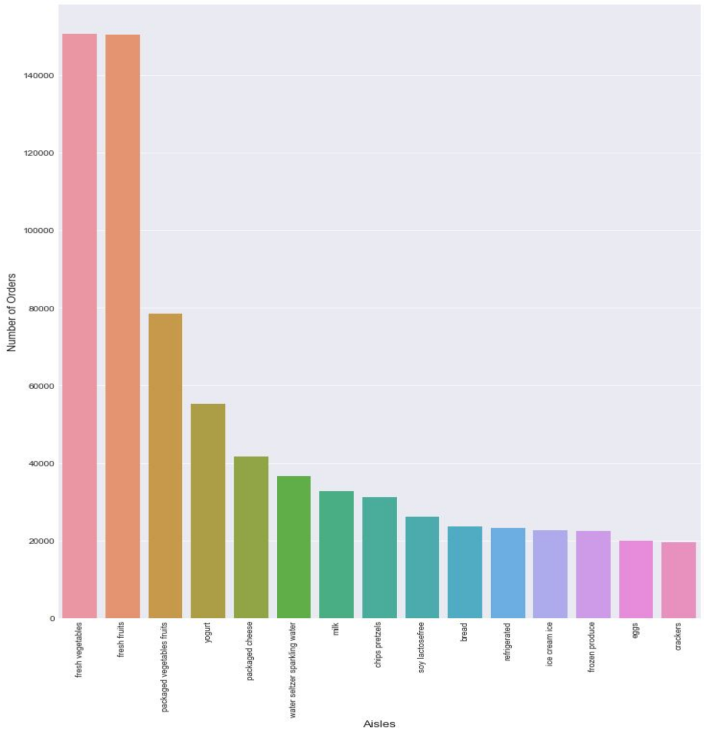
No of orders vs day of the week & hour of the day

Exploratory Data Analysis

Top aisles by the no of products

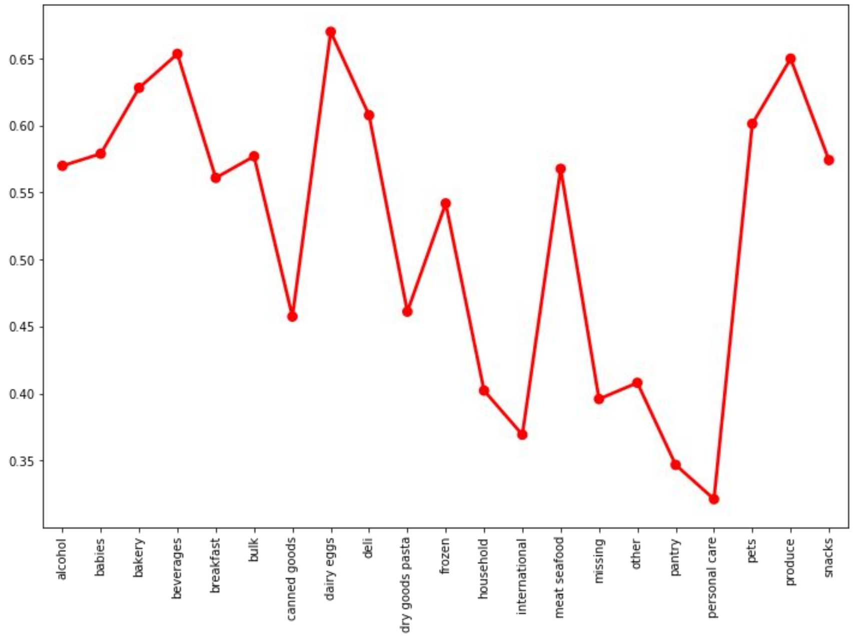


Top aisles by the no of orders

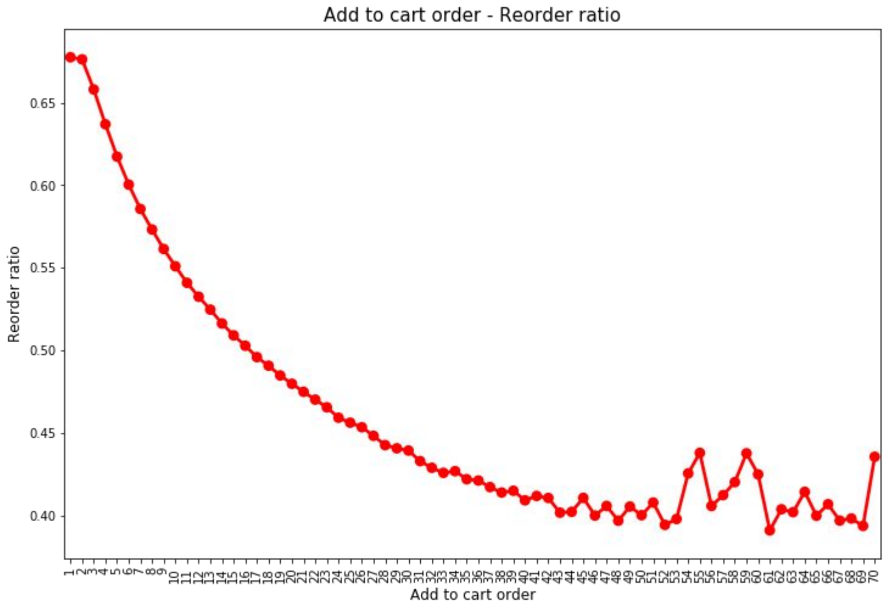


Exploratory Data Analysis

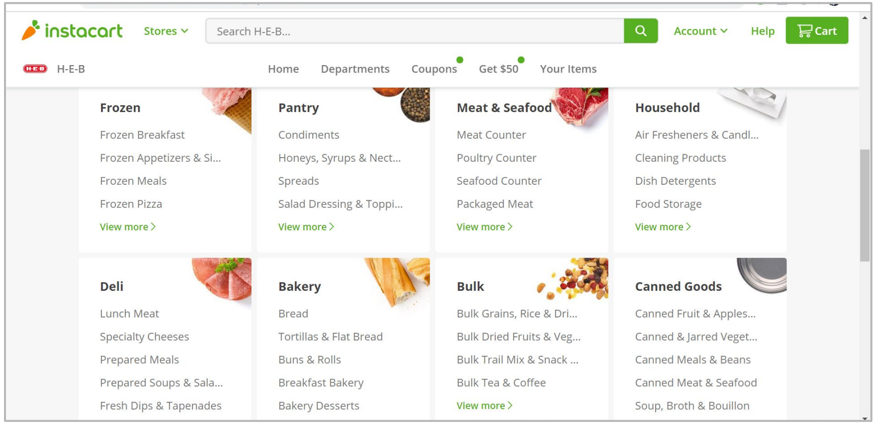
Reorder Ratio vs Department



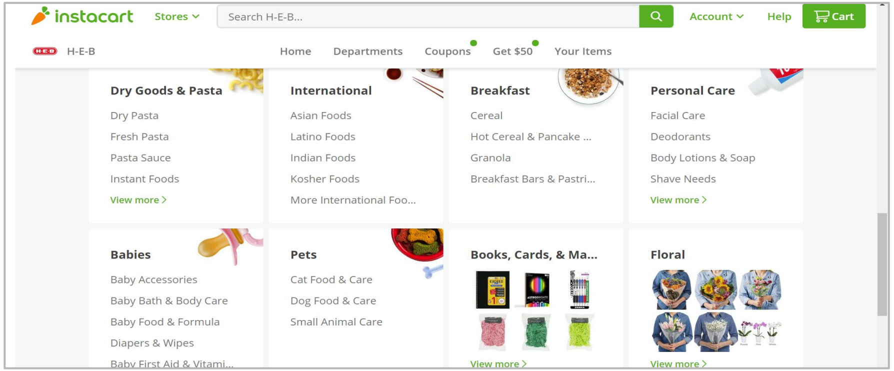
Add to cart Order vs the Reorder Ratio



Digital real estate opens the door to “personalised aisles”



Instacart does a good job at utilising their real-estate by creating “aisles” of similar products.



For existing customers, we aim to optimise aisle positioning that is designed to promote cross-selling and up-selling by placing aisles and products with high lifts near each other.