

**Anubhav Nehru
Krittika Deshwal
Nikhil Nair
Simran Kaur**



Problem Statement



Market basket analysis is crucial for retailers to understand customer shopping behavior and improve decision-making. Utilizing customer transaction data and overcoming data management challenges, businesses can enhance product placement and customer experience, especially in the online grocery sector like Instacart.

- Understanding Customer Shopping Behavior
- Identify different customer segments based on their purchasing behavior
- Predict which products are likely to be reordered
- Department and aisle analysis
- Understand peak purchasing times and days of the week for targeted promotional activities



Personalization Level Max!

1:19

1802 West Ave ▾

Try "best foods to bring camping"

For you

Grocery

Alcohol

Wholesale

Convenience

H-E-B

Sprouts Express

Walgreens

Target

Family Dollar

Sam's Club

Don's Liquor

Show all 68 stores

Get a \$50 credit

Plus 5% cash back on Instacart

Learn more

Instacart

Stores to help you save

Walgreens

\$5 off • In-store prices • By 8:00pm

Sprouts Farmers Market

In-store prices • By 8:30pm

CVS®

1:20

Search

H-E-B

Search H-E-B...

More from your brands

\$3.79

H-E-B Omega Plus Grade
Aa Cage Free Large Brown...
12 each

\$1.39

H-E-B Frozen Mixed
Vegetables
12 oz

\$4.49

Organic
Driscoll's Org...
Strawberries
16 oz

Many in stock

Many in stock

View more

Produce

Bananas

Organic Wheat Grass

1:20

Search

H-E-B

Search H-E-B...

Recipes based on past orders

Savory Breakfast Wrap

Decadent Homemade Salted Caramel Sauce

Beef Stew Mash

20min

20min

25min

Easy recipes based on past orders

Savory Breakfast Wrap

Decadent Homemade Salted Caramel Sauce

Beef Stew Mash

20min

20min

25min



Personalization Level Max!

1:21

Search H-E-B...

From departments you shop

Sponsored

Kinder Chocolate Smooth milk chocolate candy bar, 12.7 oz \$9.39

Diamondback Chile Sauce, 17 oz \$4.59

JOJO's Jojo's Dark Chocolate Covered Pretzels, 3 oz \$6.59

Gluten-Free
Many in stock

Prepared Foods

kevin's ROASTED GARLIC CHICKEN \$10.89

Okra \$4.69

Kevin's Natural Foods Roasted Garlic Chicken \$11.69 /lb

H-E-B Rotisserie Chicken \$11.69 /lb

View more

1:21

Search H-E-B...

Based on your purchases

Sponsored

ONE Flavored Protein Bar, Peanut Butter Pie, 2.12 oz \$3.09

Pearl Milling Company Syrup, Original Lite, 24 fl oz \$4.09

Febreze Odor Freshener with Lemon Scent, 8.8 fl oz \$3.79

Buy any 2, save \$3
Many in stock

Fresh Fruit

Bananas \$1.93 each (est.)

Bananas \$2.06 each (est.)

Lemons \$4.59

Organic

U.S. Plumes

1:22

Search H-E-B...

Picked For You

Sponsored

Tostitos SCOOPS! Party Size!, 14.5 oz \$6.39

Doritos Tortilla Chips, Nacho Cheese Flavored, 14.5 oz \$6.39

Gatorade Thirst Quencher Lemon Lime, 8 x 20 fl oz \$8.59

Gluten-Free • Vegan
Spend \$20, save \$3
Many in stock

Baking Essentials

Imperial Sugar Granulated Sugar, 8 lb bag \$3.79

Muffin Cake Mix, Whole Wheat \$3.49

Organic



Personalization Level Max!

1:22

Search H-E-B...

Recommended for you

- Mission Carb Balance Soft Taco Flour Tortillas 8x12 oz \$5.79 999+ bought yesterday Keto Many in stock
- H-E-B Cane Sugar 4 lb \$3.79 Many in stock
- Nature's Own Wheat, Whole 20 oz \$4.29 Many in stock

Baking and Cooking

- Pearl Milling Original Corn Syrup \$4.09
- Knorr's Chicken Bouillon in POLLO \$2.89
- Crisco All Purpose Shortening \$4.99

1:23

Got everything you need?

Items you buy often

- Dannon Whole Milk Plain 32 oz \$4.09 Keto Many in stock
- La Tortilla Factory Non-Gmo Project Ve... 32 oz \$4.69
- H-E-B Frozen Vegetables 16 oz 10 ct \$1.79 Many in stock

Save \$1.00

Sparkling Water

- bubly sparkling water orange bubbly 8 fl oz \$1.09
- bubly sparkling water orange bubbly 8 fl oz \$1.09
- White Mountain Spring Water 16.9 fl oz \$1.09

1:23

Got everything you need?

Related to items you buy often

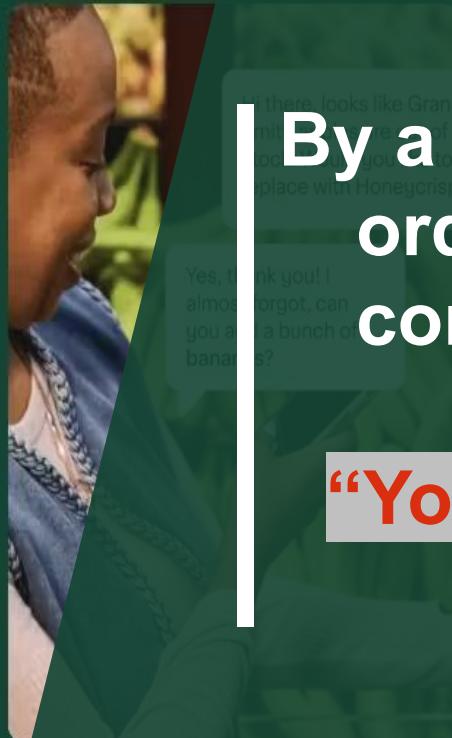
- H-E-B Omega Plus Grade Aa Cage Free Large Brown Eggs 12 each \$3.79 Many in stock
- Hill Country Fare Whole Milk 1 gal \$4.39 Many in stock
- H-E-B Natural Boneless Skinless Chicken Thighs 4 lb \$4.39 Many in stock

Charcuterie board essentials

- White Mountain Spring Water 16.9 fl oz \$14.09
- Love, Corn White Seedless Grapes 5 lb \$2.09
- Giesen 0% Non-Dairy Creamer 32 fl oz \$4.09 Gluten-Free • Vegan



What's the value of a recommendation?



By a customer's 10th order, 25% of all conversions are from the
“Your Items” List!





The Scale of our Data!

20M + Total Orders



134 Aisles



50k+ Products



21 Departments



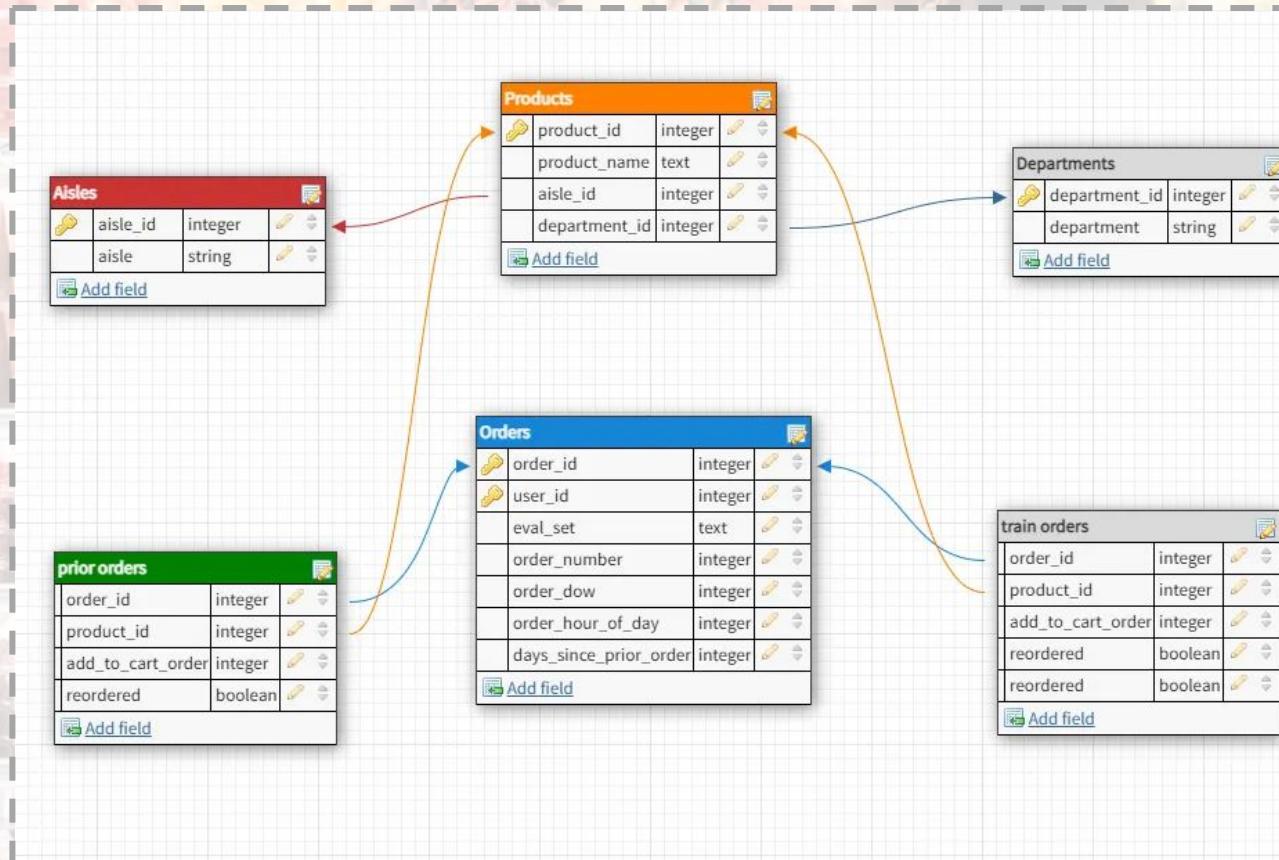
Data at Glance

20M + Total Orders

50k+ Products

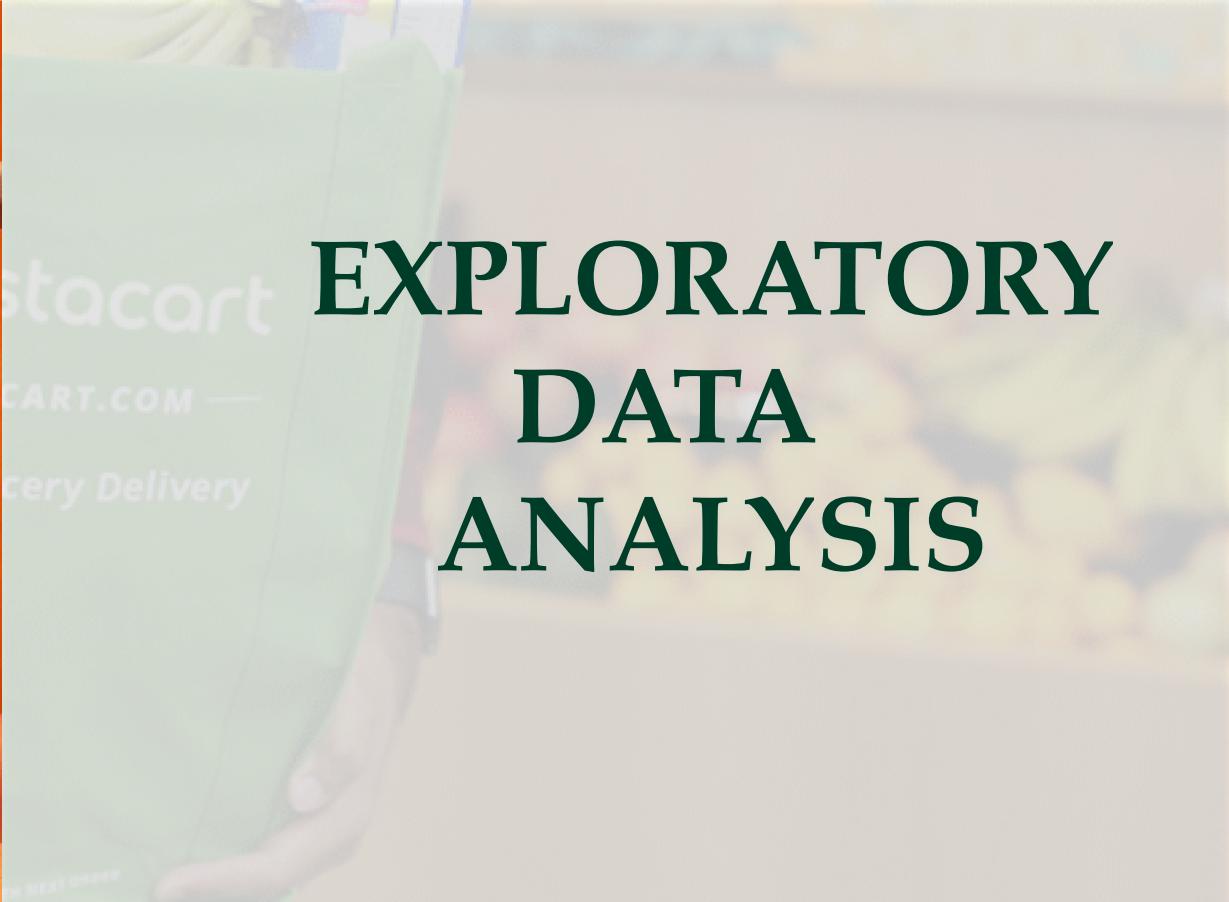
134 Aisles

21 Departments





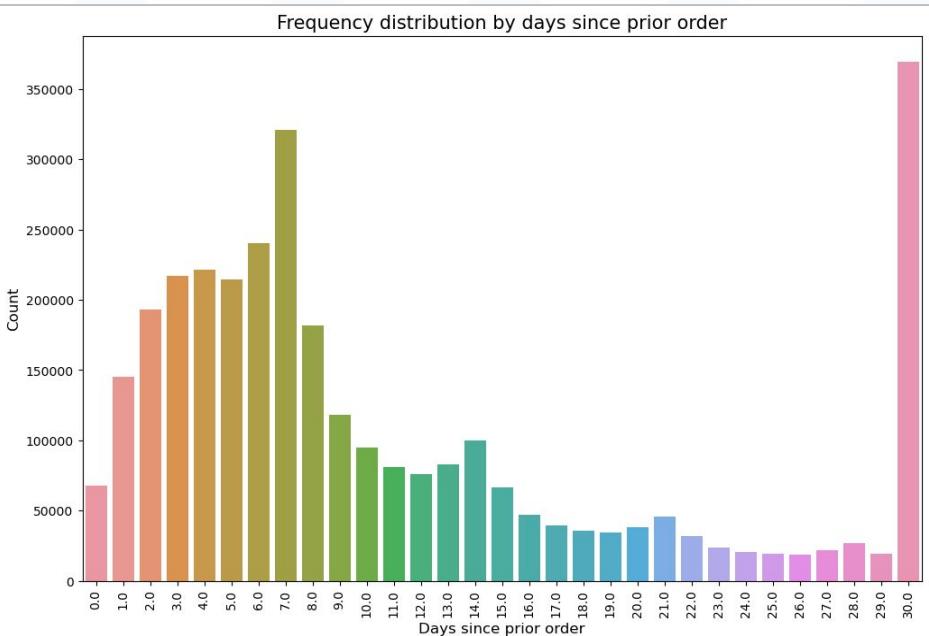
Let's Deep Dive!



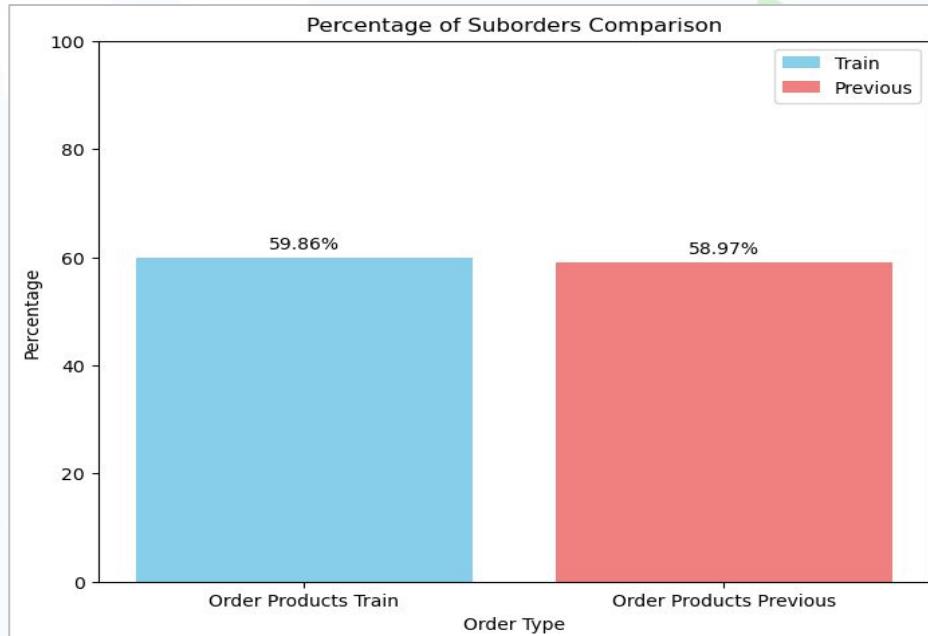
EXPLORATORY DATA ANALYSIS



Order Frequency Analysis



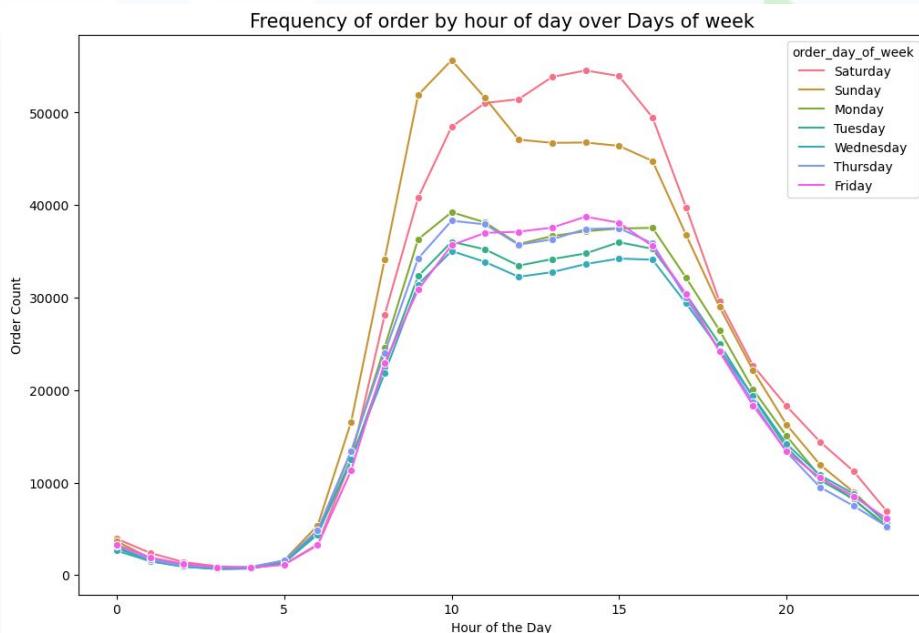
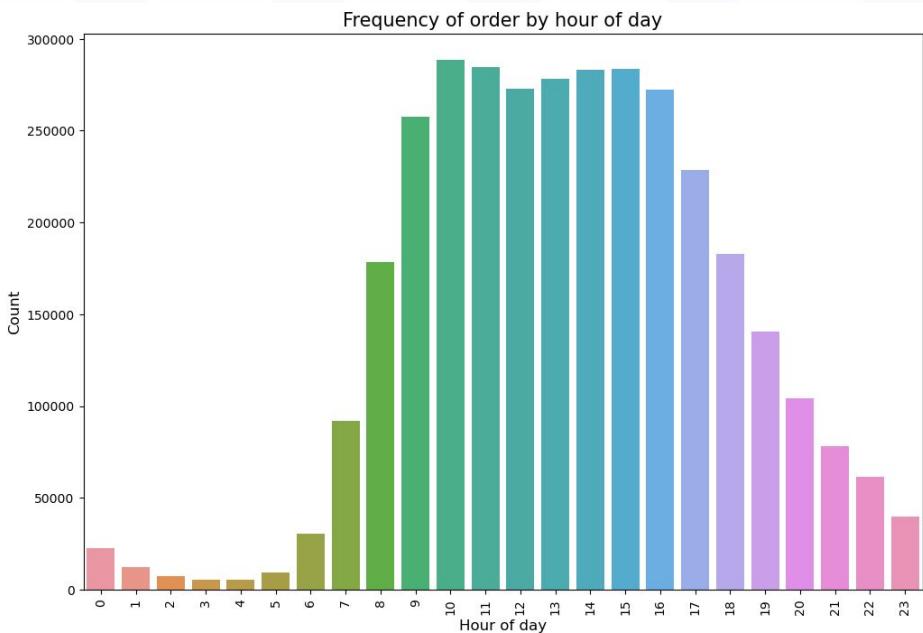
Most orders are re-ordered either within 7 days or after 30 days, implying frequent buyers and bulk buyers



60% of the products are re-ordered



Order Frequency by Time and Day of the Week

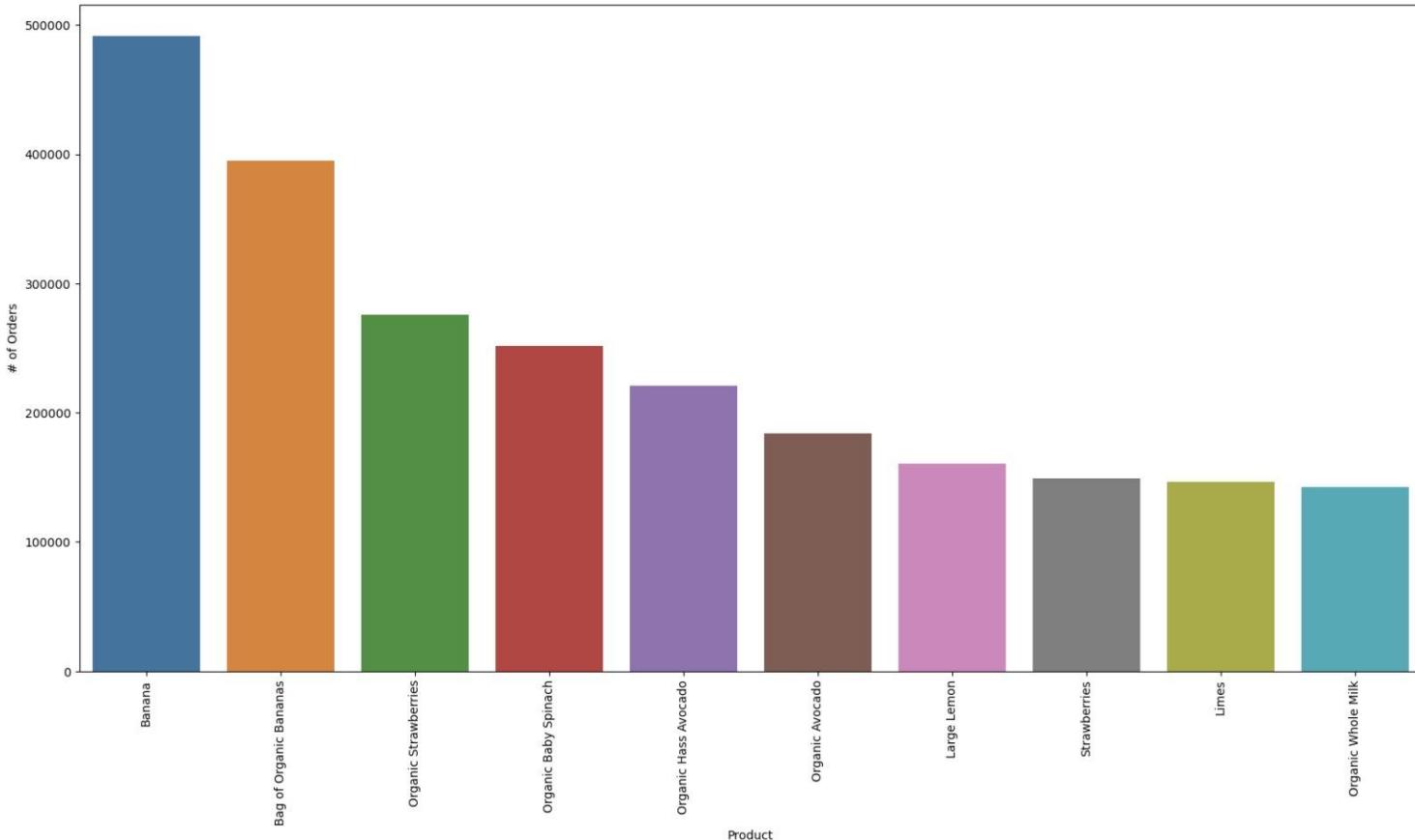


Most of the orders are placed during 9AM to 5PM

There is a prominent peak of Order Count on the weekends



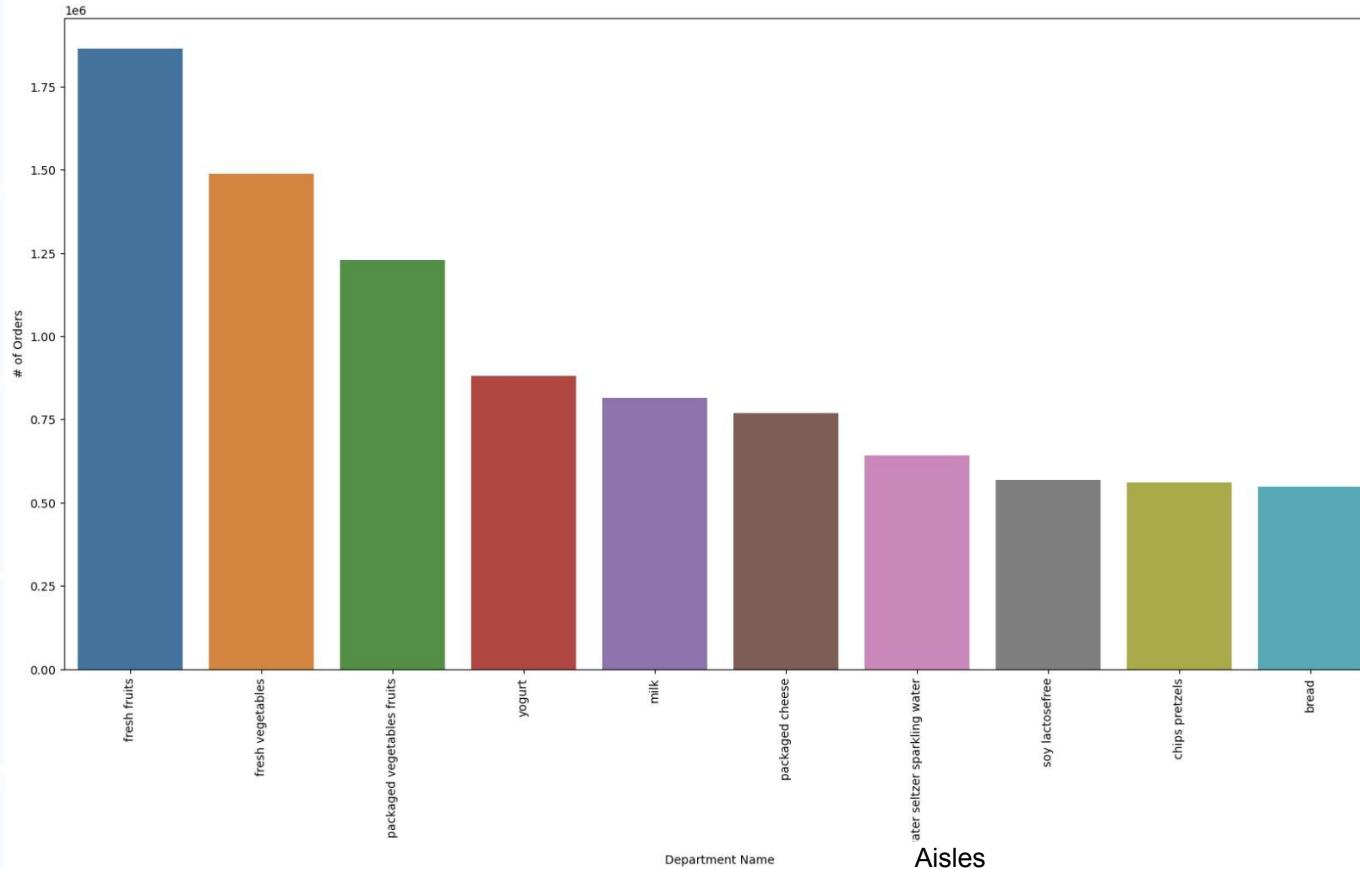
Best Selling Products



Most of the best selling products are organic fresh fruits and vegetables.



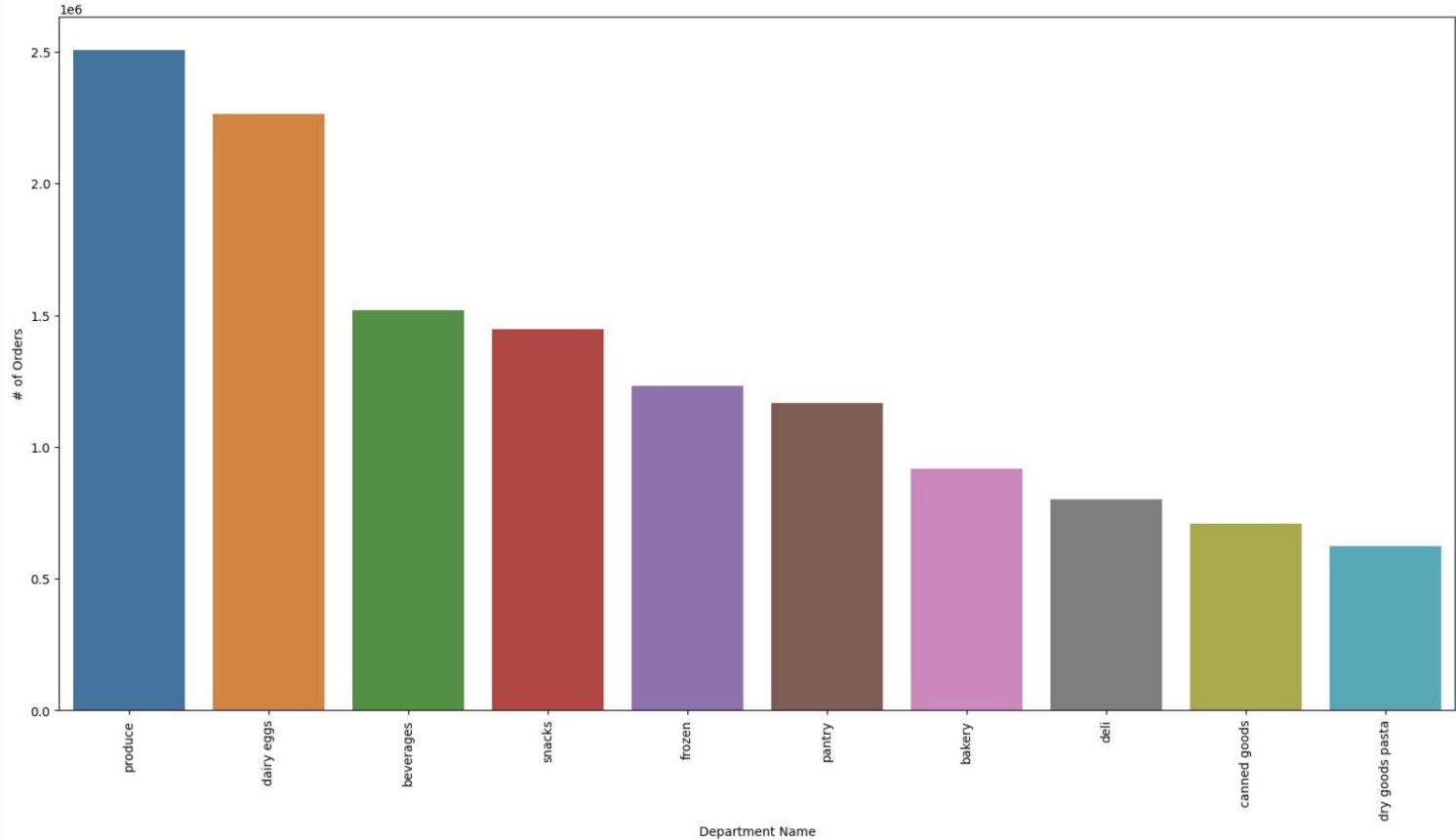
Best Selling Aisles



Most of the best selling aisles are organic fresh fruits and vegetables, and daily consumables like dairy products



Best Selling Departments



Most of the best selling departments are produce, and daily consumables like dairy products, snacks, and frozen items.



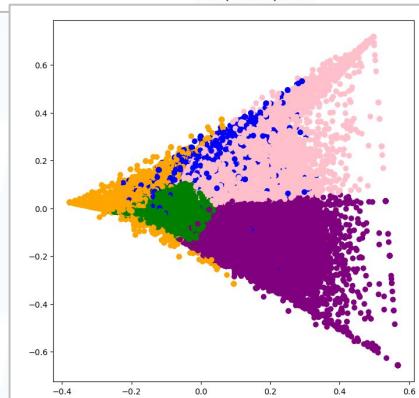
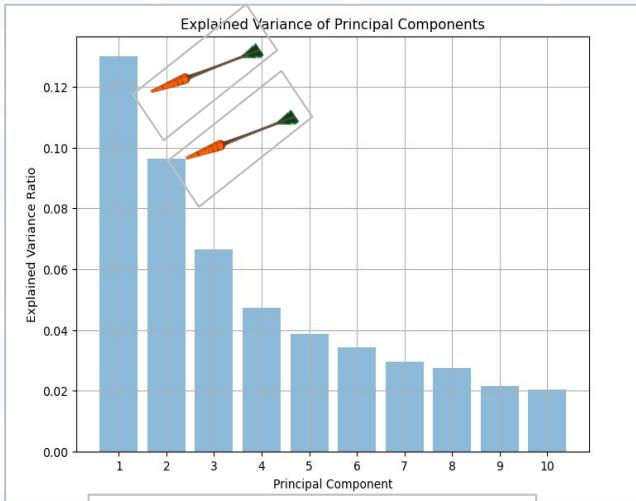
Let's Deep Dive!



CUSTOMER SEGMENTATION



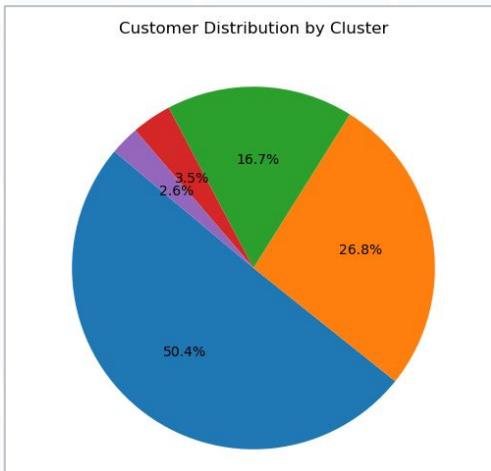
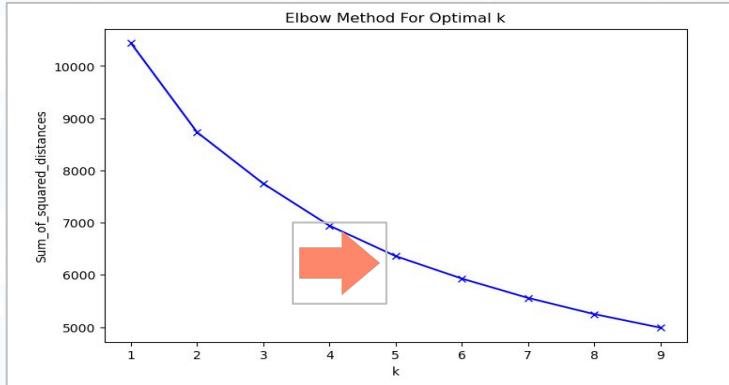
K-Means over PCA



- Since there are thousands of products, we rely on aisles, which represent categories of products.
- Even with aisles, we reduced features and used Principal Component Analysis to find new dimensions along which clustering would be easier.
- 5 Distinguishable Clusters by purchasing Behaviour
- Visualization of clusters among first 2 components.



K-Means over PCA

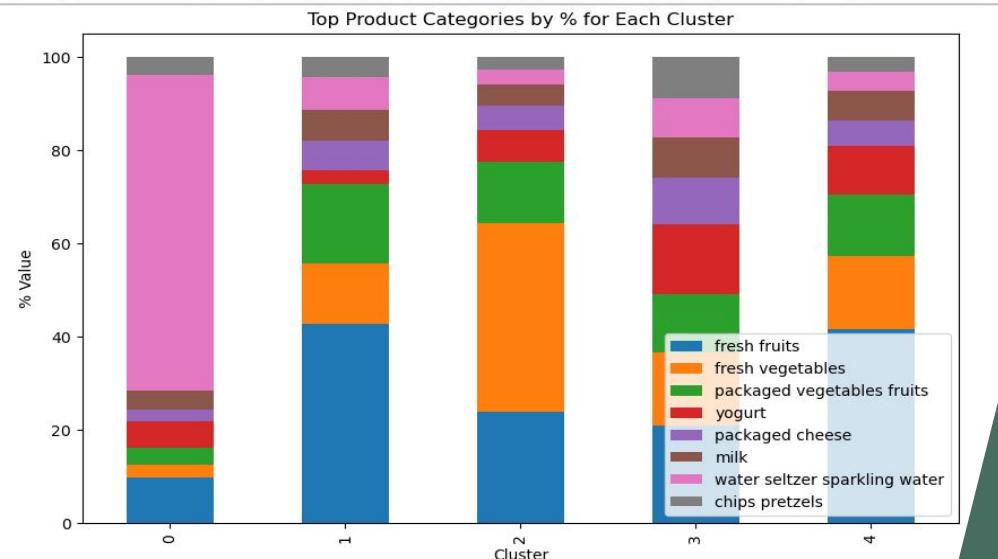


- Optimal k value identified at 5. Distinguishable Clusters by purchasing Behaviour

- 50% customers in the largest cluster accounting for over 45% of the overall orders.



K-Means over PCA



- Each cluster respectively results into certain consumers having a very strong preference for (with some overlap):

0: water seltzer sparkling water aisle

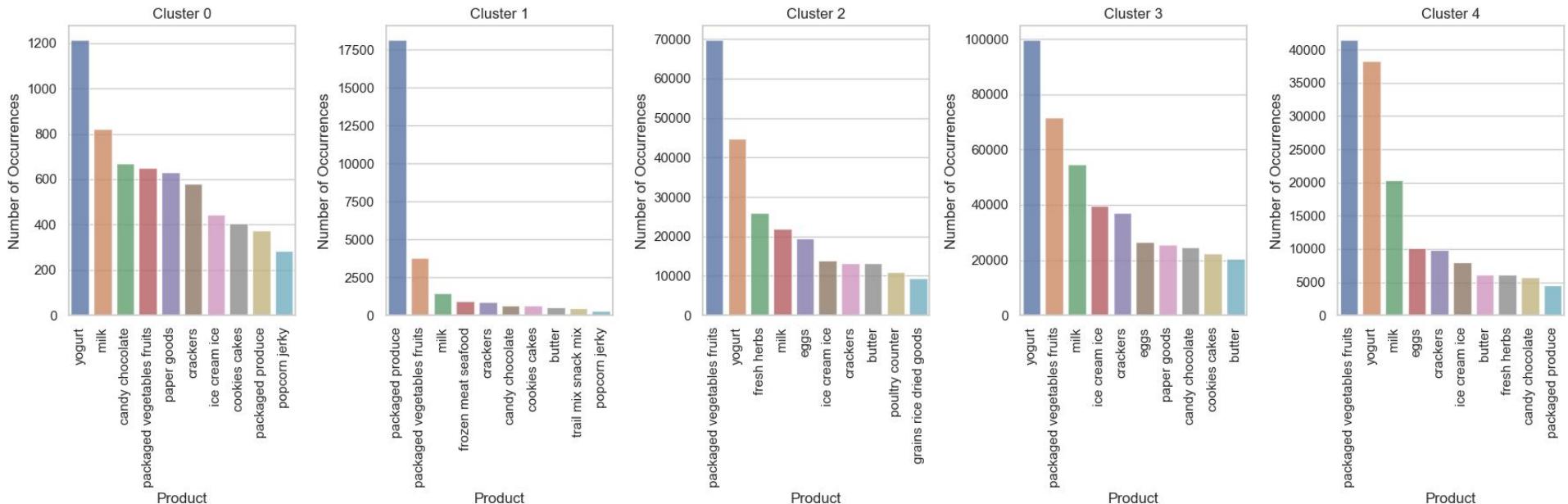
1: fresh fruits & vegetables

2: fresh fruits & vegetables

3: yoghurt and milk



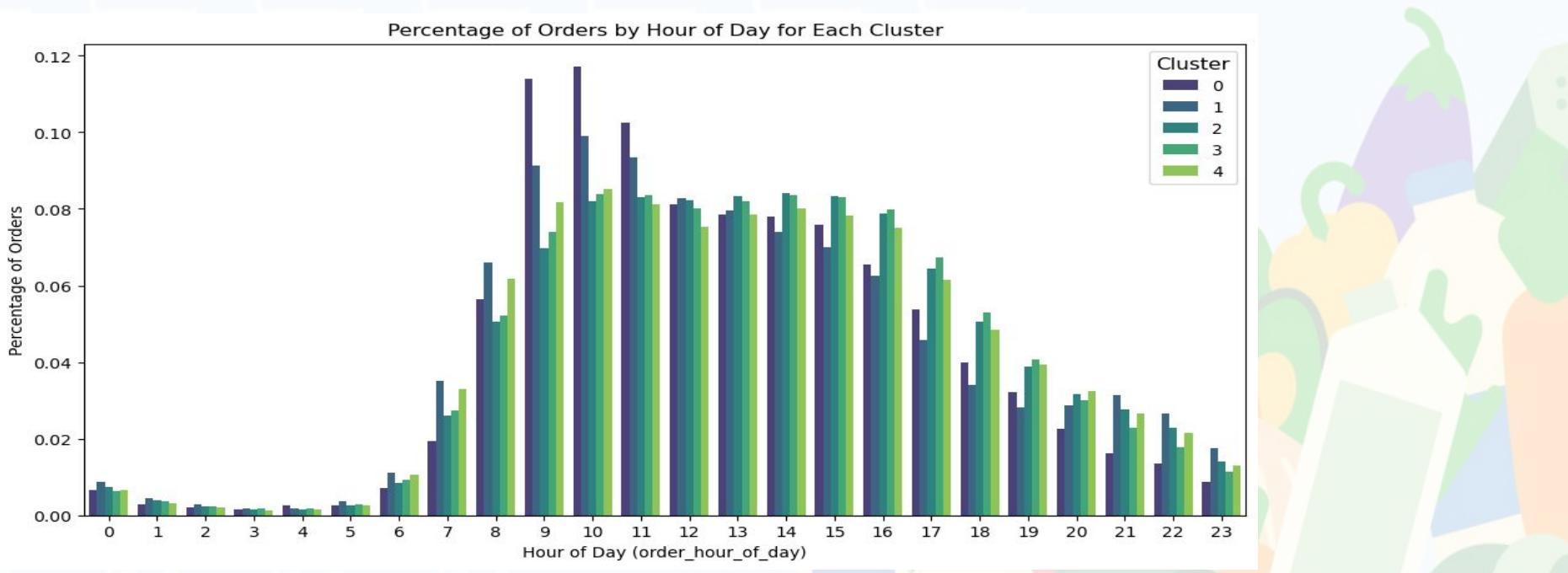
Cluster Statistics



Orders by Aisle Distribution for each cluster



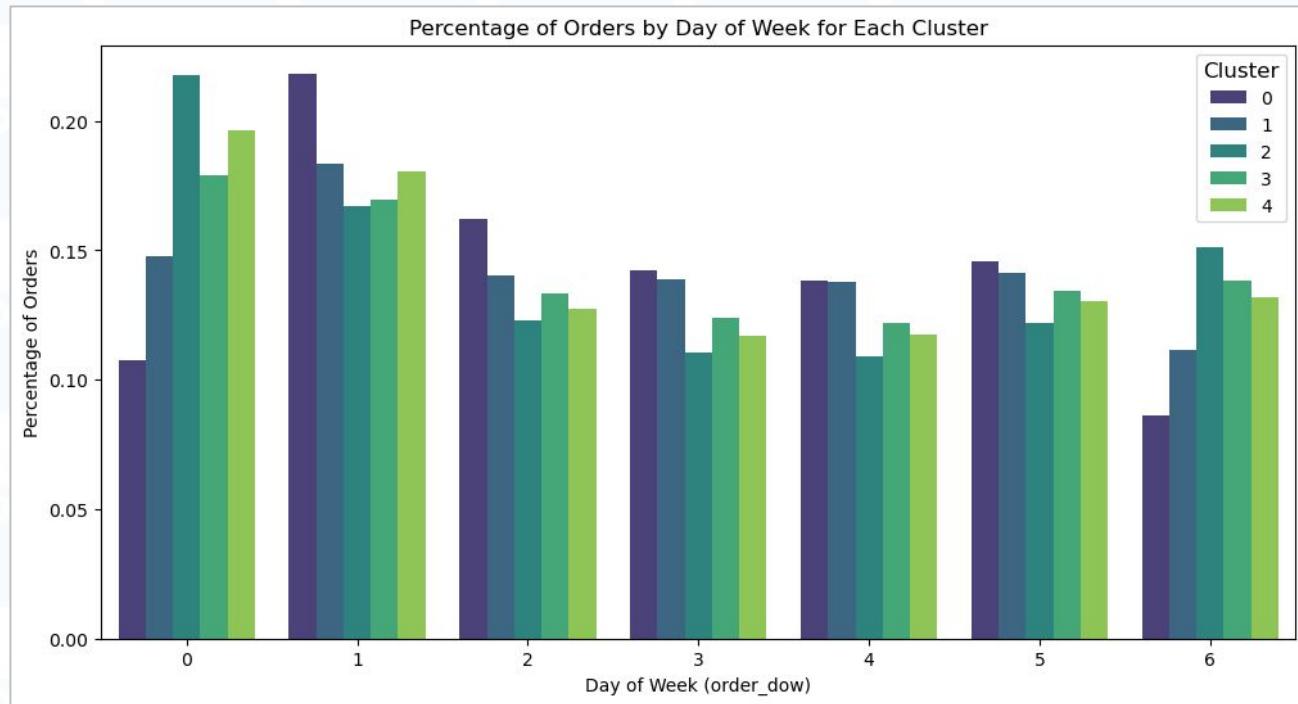
Cluster Statistics



A larger % of Health Enthusiasts order in early? Majority of the orders come in between 9 am – 4 pm



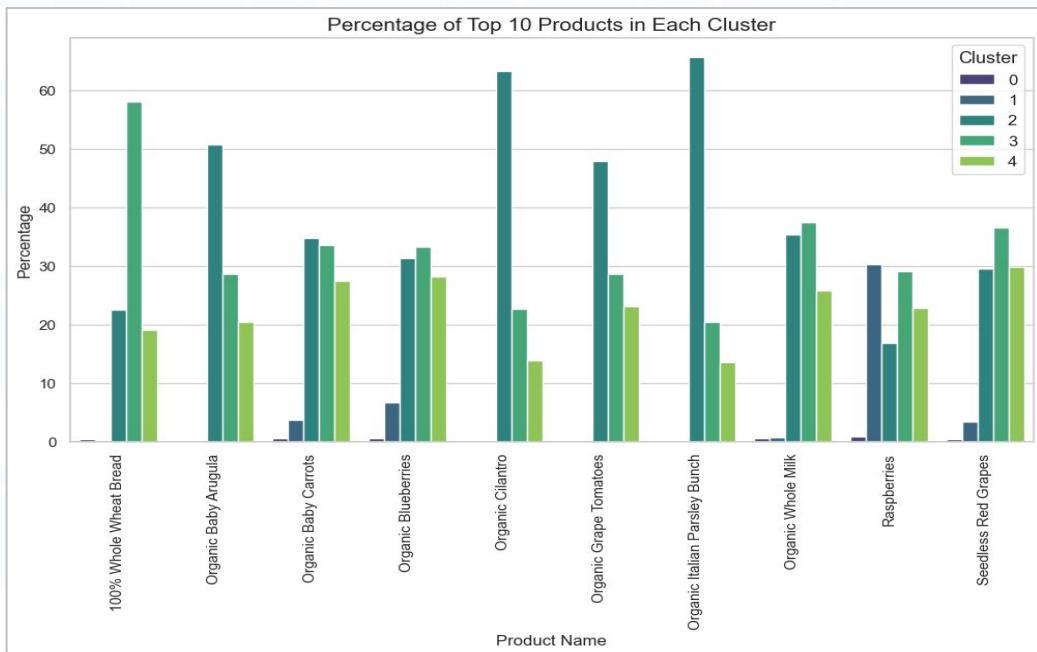
Cluster Statistics



Most orders at the start of the week particularly by veggie lovers and health enthusiasts



Cluster Statistics



Veggie Lovers and Diary Enthusiasts order popular products. Cross-sell opportunities exist for Health Enthusiasts and Snack Lovers.



Identifying Clusters?



Health Enthusiasts



Snack Lovers



Veggie Lovers



Diary Enthusiasts



Balanced Shoppers





Cluster Statistics

Category	Health Enthusiasts	Snack Lovers	Veggie Lovers	Diary Enthusiasts	Balanced Shoppers
% Cust	2.2 %	3.5 %	27 %	50.5 %	16.6 %
% Orders	1.1 %	3 %	28.6 %	49.3 %	17.8 %
Median Orders	3	6	7	7	7
Order Gap	9	10	8	8	7
Basket Size	1	2	3	3	3
Basket Re-order	51.4 %	68.5 %	55.6%	57%	62.2 %



Let's Deep Dive!



MARKET BASKET ANALYSIS



Market Basket Analysis

Model:
Heuristic &
Apriori
Algorithm

Results &
Robustness



What will you buy next?





Market Basket Analysis

APRIORI ALGORITHM

TRANSACTION	ITEM1	ITEM2	ITEM3
1	MILK	SUGAR	Coffee
2	MILK	SUGAR	
3	MILK	SUGAR	
4	MILK	SUGAR	
5	MILK	SUGAR	

ARDUINO STARTUPS
Eduardo Latorre - University of Valencia

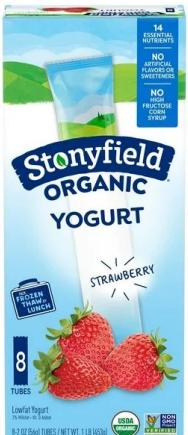
$$Support = \frac{frq(X, Y)}{N}$$

Rule: $X \Rightarrow Y$

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



Product Wise Lift



What should I recommend to the customer ?



Product Wise Lift





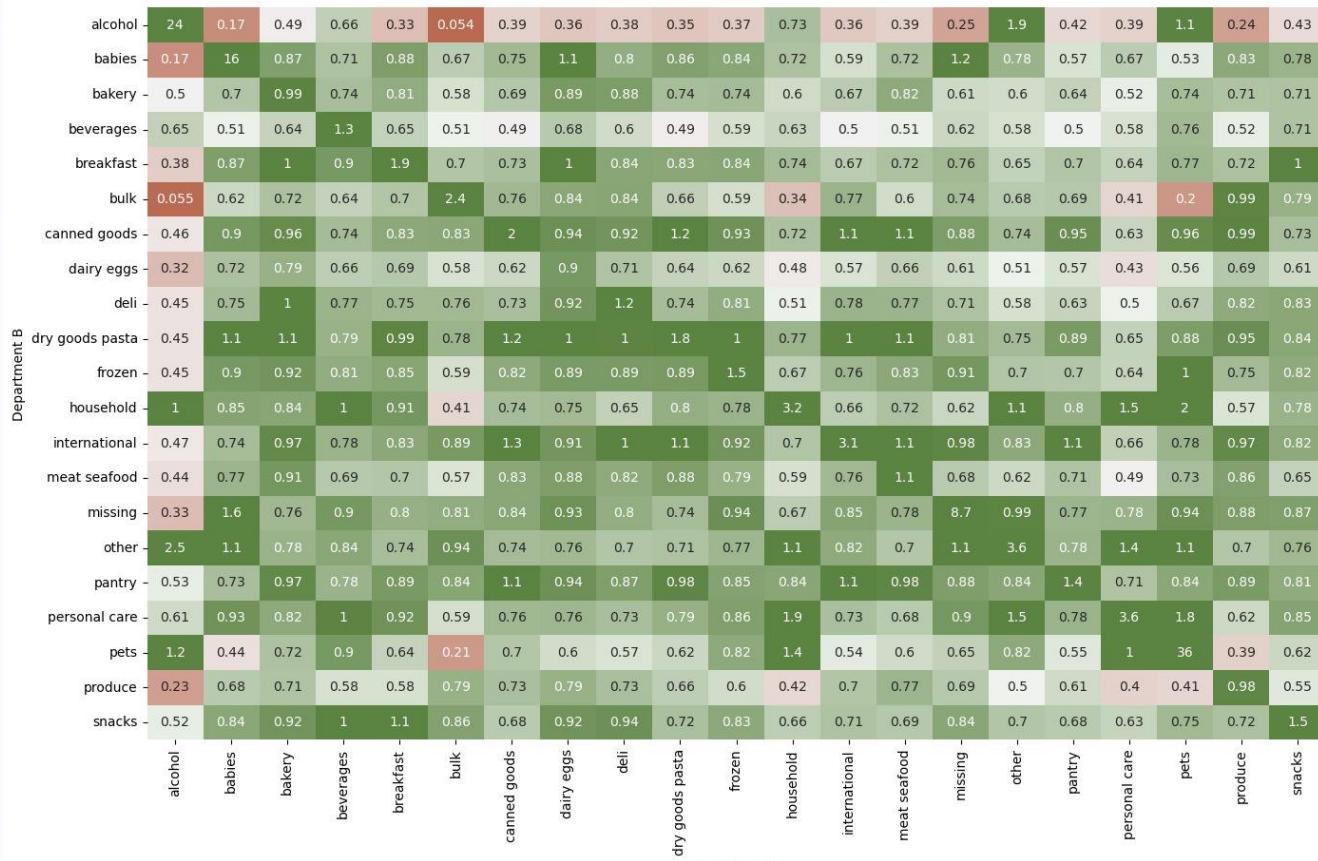
Product Wise Lift

Same Brand Same Product - Different Flavour



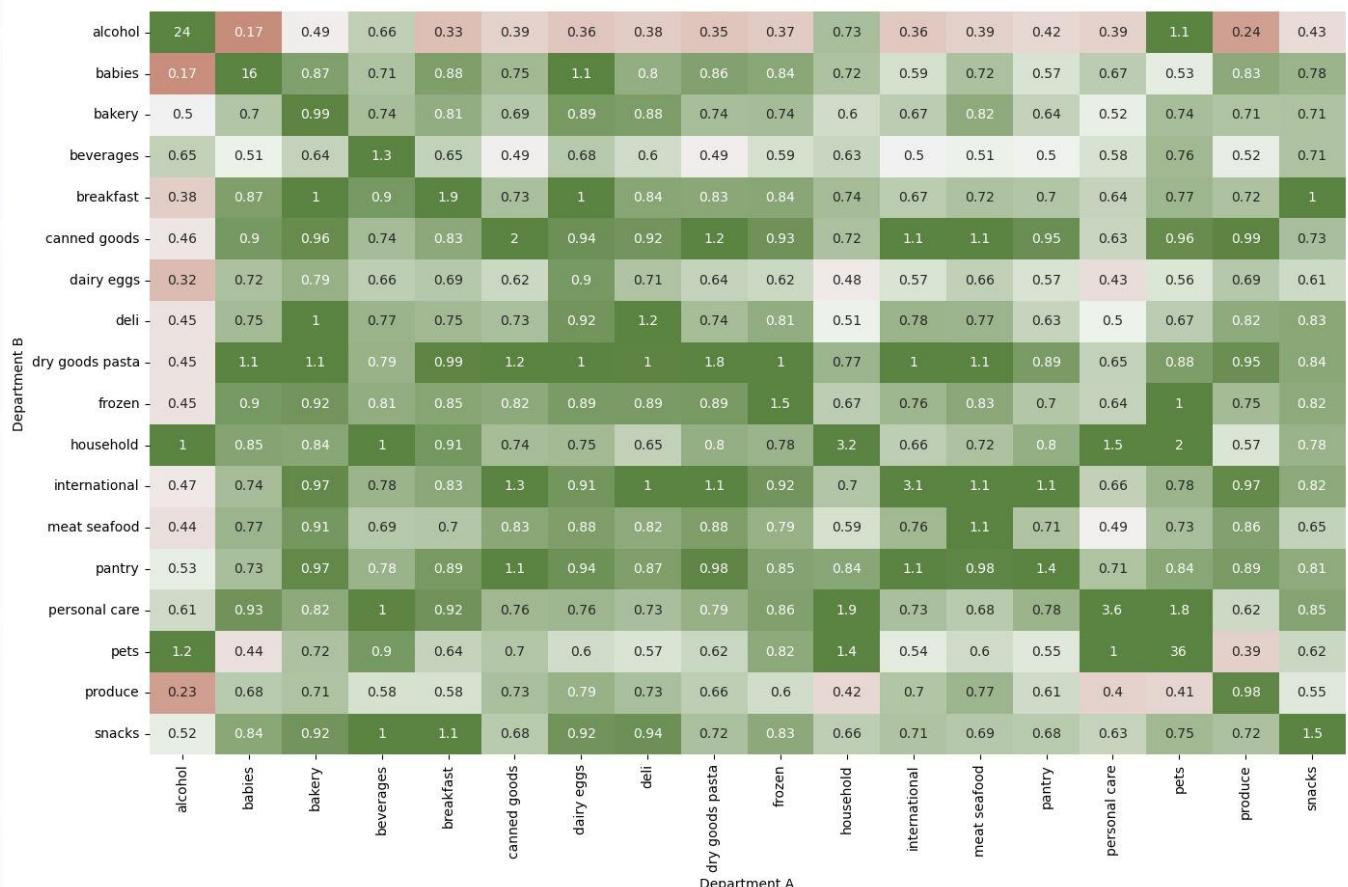


Department Wise Lift





Department Wise Lift





Cross-Selling Enhancement Strategy

1

User adds a product to the cart

2

Find the Cluster of the User
(as defined in our Customer Segment Analysis)

3

Locate the aisle of that product

4

Find the aisle having the highest lift with
that aisle for re-ordered products (using
Apriori Algorithm) for users in that cluster.

5

Recommend the best-selling and most
re-ordered product for that aisle



Cross-Selling Enhancement Strategy

11:32

X Bookmark Share



Many in stock

Best seller

Stonyfield Organic YoKids Squeezers Lowfat Yogurt Variety Pack
Organic • Gluten-Free • Kosher
2 oz

\$8.69 16 x 2 oz

Auto-order item >
Add to list >

Product information

Details Ingredients Nutrition

Picked For You Sponsored

- 1 + Add to Cart



11:33

X Bookmark Share

Best seller

Stonyfield Organic YoKids Squeezers Lowfat Yogurt Variety Pack
Organic • Gluten-Free • Kosher
2 oz

\$8.69 16 x 2 oz

Auto-order item >
Add to list >
Choose replacement >
Add a note for your shopper >

Product information

Details Ingredients Nutrition

Picked For You Sponsored


\$9.99


\$8.69


\$6.89

In Cart



Key Recommendations & Takeaways

To optimize Market Basket Analysis, we need to categorize customers into distinct segments based on the user's preferences.

We can improve the understanding of these preferences by including additional data such as consumer demographics

If we get Revenue information from Instacart, we can run more Customer Targeting analysis, like RFM Analysis, Customer Lifetime Value, etc.

We propose two levels of cross-selling product recommendations

- Cross-selling with Highest Lift Product
- Recommending products with the strongest lift with aisle

We can leverage A/B Testing to test our recommended approach.

We recommend getting updated data, because of the fast-changing customer dynamics, specially post Covid era



Thank you!



Questions?



Appendix

