University of CINCINNATI | CARL H. LINDNER COLLEGE OF BUSINESS

Market Basket Analysis

Statistical Computing Project (BANA 6043)

Group #4

Contributors: Ashish Saxena, Lohit Borah, Tanmay Shrivastava, Pratibha Prakash Tiwari

Index

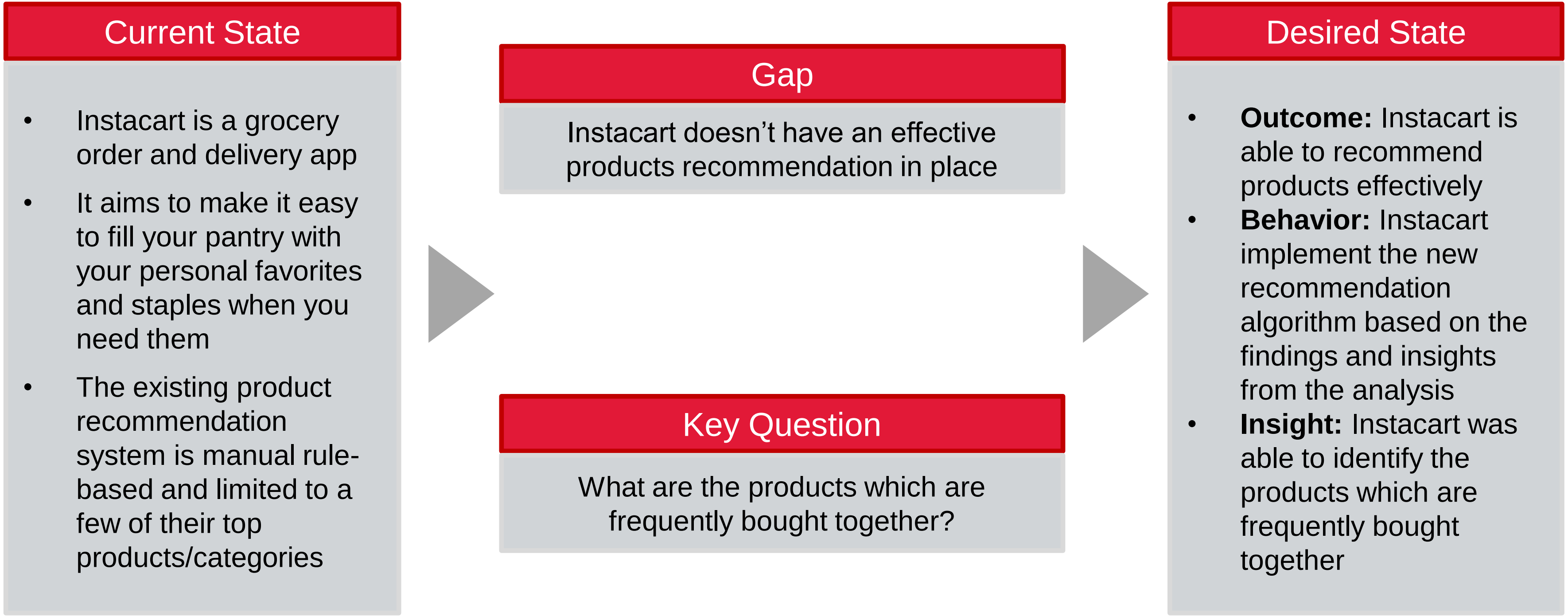
- Problem Statement
- Problem Definition
- Executive Summary
- Supporting Outcomes
  - Exploratory Data Analysis
  - Basket Analysis
- New Functions Used & Context
- R Code & Dataset (Embedded)
  - Attributes of the datasets used
- Appendix

Problem Statement

- Instacart is a grocery order and delivery app, which accepts grocery orders for various products and delivers them from various partnered stores to respective customers

- The company wants to analyze its past transactional data to develop a recommendation system for its products so that they can effectively suggest associated items

Problem Definition



Current State

- Instacart is a grocery order and delivery app
- It aims to make it easy to fill your pantry with your personal favorites and staples when you need them
- The existing product recommendation system is manual rule-based and limited to a few of their top products/categories

Gap

Instacart doesn’t have an effective products recommendation in place

Key Question

What are the products which are frequently bought together?

Desired State

- **Outcome:** Instacart is able to recommend products effectively
- **Behavior:** Instacart implement the new recommendation algorithm based on the findings and insights from the analysis
- **Insight:** Instacart was able to identify the products which are frequently bought together

Executive Summary

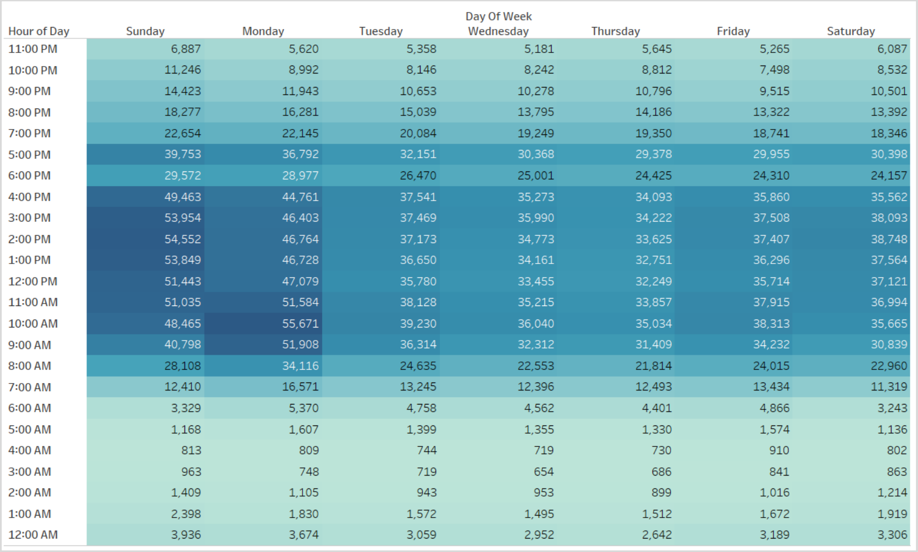
- Customers buying meat products, fresh fruits or meat alternatives should be recommended best selling products from fresh vegetables

- Customers buying Yogurt should be recommended top 5 best selling flavors in the category

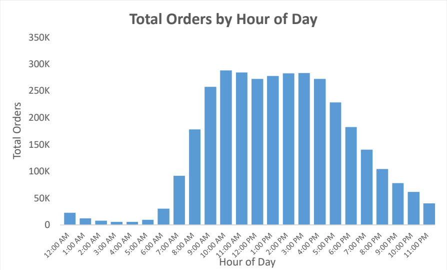
- Customers buying dairy, eggs, frozen and canned foods should be recommended fresh vegetables

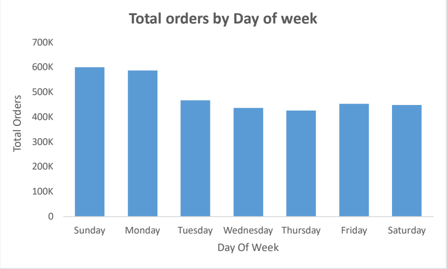
- All customers buying protein bars should also be recommended other health products like Peanut Butter, Greek Yogurt & Nutrition Blends

Mondays experienced highest orders between 9 AM – 11 AM



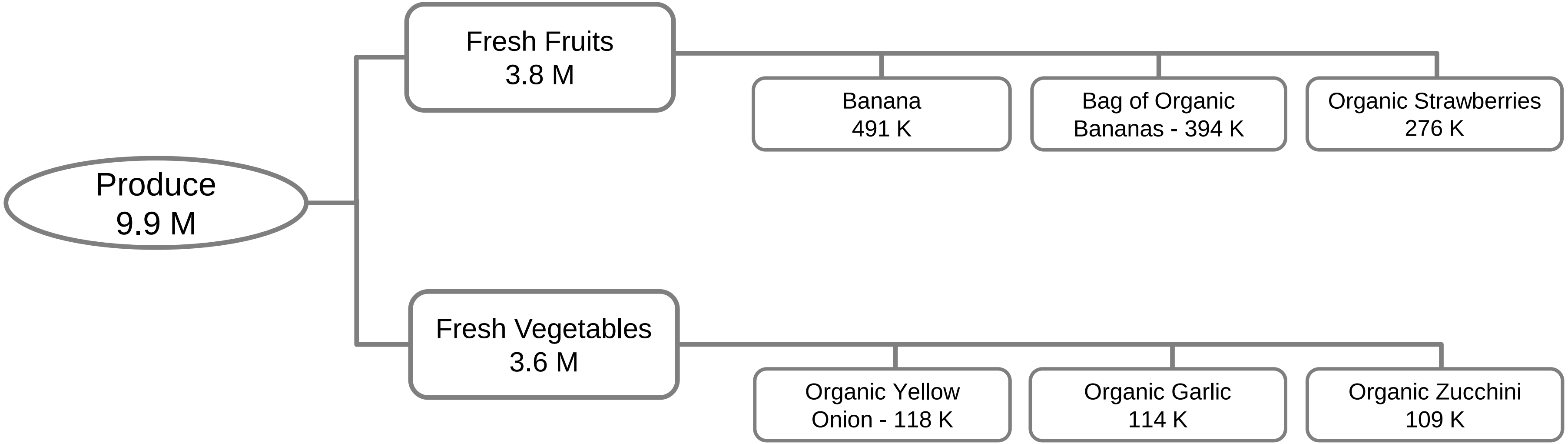
Hour of Day	Sunday	Monday	Tuesday	Day Of Week Wednesday	Thursday	Friday	Saturday
11:00 PM	6,887	5,620	5,358	5,181	5,645	5,265	6,087
10:00 PM	11,246	8,992	8,146	8,242	8,812	7,498	8,532
9:00 PM	14,423	11,943	10,653	10,278	10,796	9,515	10,501
8:00 PM	18,277	16,281	15,039	13,795	14,186	13,322	13,392
7:00 PM	22,654	22,145	20,084	19,249	19,350	18,741	18,346
5:00 PM	39,753	36,792	32,151	30,368	29,378	29,955	30,398
6:00 PM	29,572	28,977	26,470	25,001	24,425	24,310	24,157
4:00 PM	49,463	44,761	37,541	35,273	34,093	35,860	35,562
3:00 PM	53,954	46,403	37,469	35,990	34,222	37,508	38,093
2:00 PM	54,552	46,764	37,173	34,773	33,625	37,407	38,748
1:00 PM	53,849	46,728	36,650	34,161	32,751	36,296	37,564
12:00 PM	51,443	47,079	35,780	33,455	32,249	35,714	37,121
11:00 AM	51,035	51,584	38,128	35,215	33,857	37,915	36,994
10:00 AM	48,465	55,671	39,230	36,040	35,034	38,313	35,665
9:00 AM	40,798	51,908	36,314	32,312	31,409	34,232	30,839
8:00 AM	28,108	34,116	24,635	22,553	21,814	24,015	22,960
7:00 AM	12,410	16,571	13,245	12,396	12,493	13,434	11,319
6:00 AM	3,329	5,370	4,758	4,562	4,401	4,866	3,243
5:00 AM	1,168	1,607	1,399	1,355	1,330	1,574	1,136
4:00 AM	813	809	744	719	730	910	802
3:00 AM	963	748	719	654	686	841	863
2:00 AM	1,409	1,105	943	953	899	1,016	1,214
1:00 AM	2,398	1,830	1,572	1,495	1,512	1,672	1,919
12:00 AM	3,936	3,674	3,059	2,952	2,642	3,189	3,306





- The highest number of orders were observed during 9 AM – 11 AM on Monday
- The highest number of orders were recorded on Sundays and Mondays

Produce was the department with highest orders having ‘Fresh Fruits’ & ‘Fresh Vegetables’ as top categories



Produce
9.9 M

Fresh Fruits
3.8 M

Banana
491 K

Bag of Organic
Bananas - 394 K

Organic Strawberries
276 K

Fresh Vegetables
3.6 M

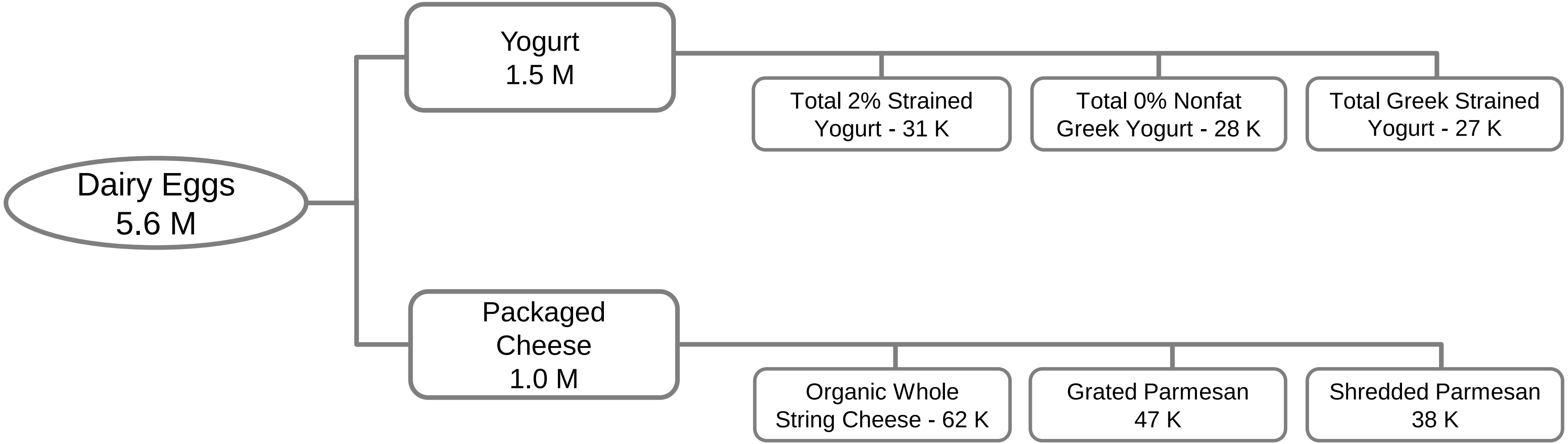
Organic Yellow
Onion - 118 K

Organic Garlic
114 K

Organic Zucchini
109 K

- The department with highest number of orders is Produce with a total order of 9.9 M
  - Bananas were the most ordered Fresh Fruits followed by Strawberries
  - Organic Yellow Onion and Organic Garlic were the most ordered Fresh Vegetables

Diary Eggs was the department with second highest orders having ‘Yogurt’ & ‘Packaged Cheese’ as top categories



Dairy Eggs
5.6 M

Yogurt
1.5 M

Total 2% Strained Yogurt - 31 K

Total 0% Nonfat Greek Yogurt - 28 K

Total Greek Strained Yogurt - 27 K

Packaged Cheese
1.0 M

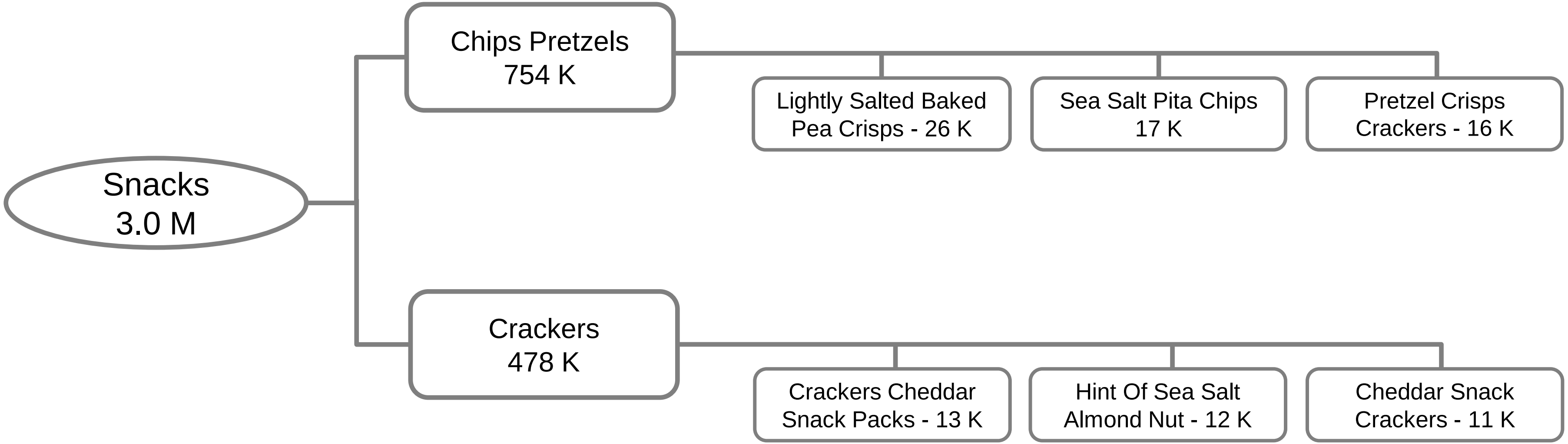
Organic Whole String Cheese - 62 K

Grated Parmesan
47 K

Shredded Parmesan
38 K

- The department with second highest number of orders is Dairy Eggs with a total order of 5.6 M
  - Total 2% Strained Yogurt was the most ordered Yogurt followed by Total 0% Nonfat Greek Yogurt
  - Organic Whole String Cheese was the most ordered Packaged Cheese followed by Grated Parmesan

Snacks was the department with third highest orders having ‘Chips Pretzels’ & ‘Crackers’ as top categories



Snacks
3.0 M

Chips Pretzels
754 K

Lightly Salted Baked Pea Crisps - 26 K

Sea Salt Pita Chips
17 K

Pretzel Crisps Crackers - 16 K

Crackers
478 K

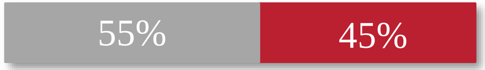
Crackers Cheddar Snack Packs - 13 K

Hint Of Sea Salt Almond Nut - 12 K

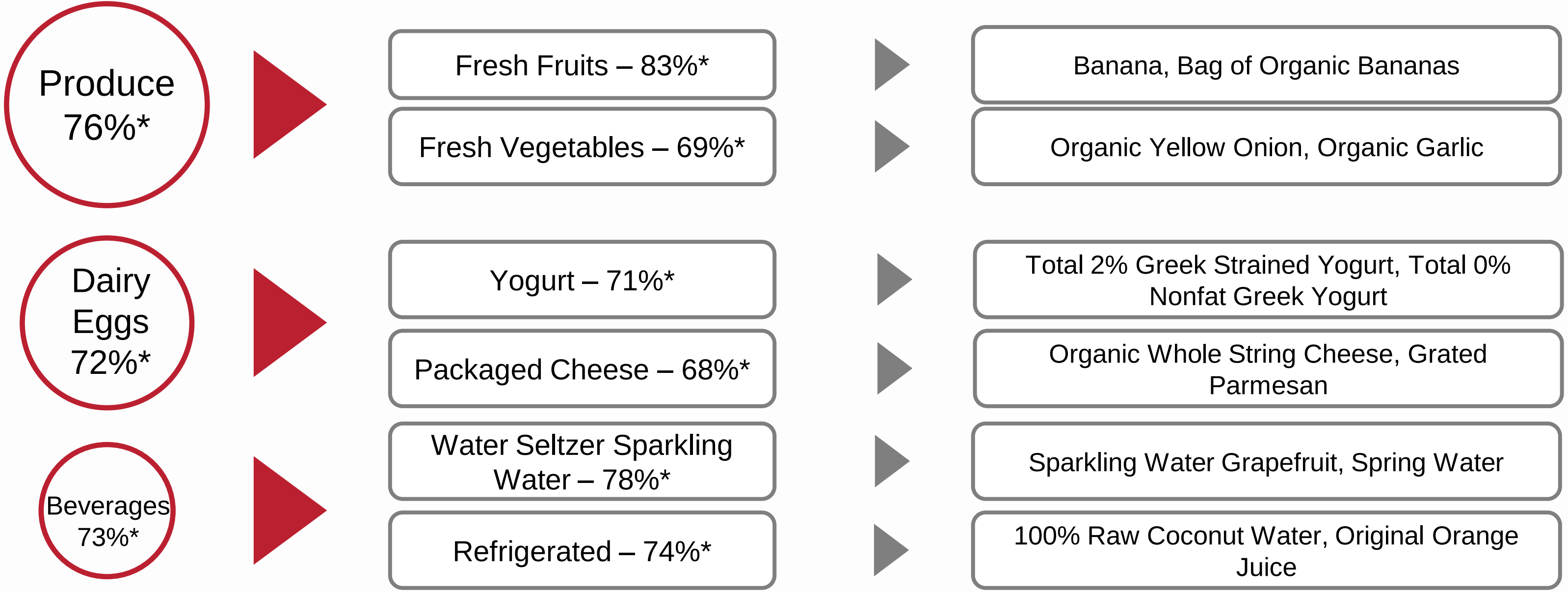
Cheddar Snack Crackers - 11 K

- The department with third highest number of orders is Snacks with a total order of 3.0 M
  - Lightly Salted Baked Pea Crisps was the most ordered Chips Pretzels followed by Sea Salt Pita Chips
  - Crackers Cheddar Snack Packs and Hint Of Sea Salt Almond Nut were the most ordered Crackers

Produce, Dairy Eggs & Beverages experienced highest reorder percentages



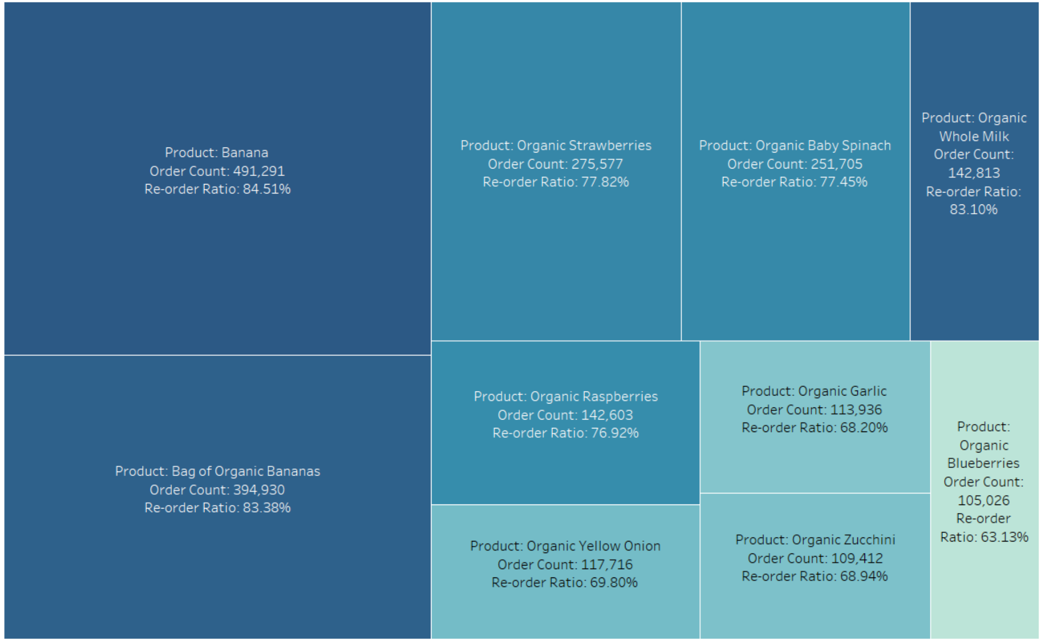
45% of total products were reordered



Department	Category	Products
Produce 76%*	Fresh Fruits – 83%*	Banana, Bag of Organic Bananas
	Fresh Vegetables – 69%*	Organic Yellow Onion, Organic Garlic
Dairy Eggs 72%*	Yogurt – 71%*	Total 2% Greek Strained Yogurt, Total 0% Nonfat Greek Yogurt
	Packaged Cheese – 68%*	Organic Whole String Cheese, Grated Parmesan
Beverages 73%*	Water Seltzer Sparkling Water – 78%*	Sparkling Water Grapefruit, Spring Water
	Refrigerated – 74%*	100% Raw Coconut Water, Original Orange Juice

* Re-order percentage

While Organic Whole Milk was ordered lesser times, it was the third most re-ordered product



- Banana (84.51%) was the most re-ordered product followed by Organic Bananas (83.38%) & Organic Whole Milk (83.10%)

- While Organic Strawberries was ordered significantly (over 275K orders), it was re-ordered 77.8%

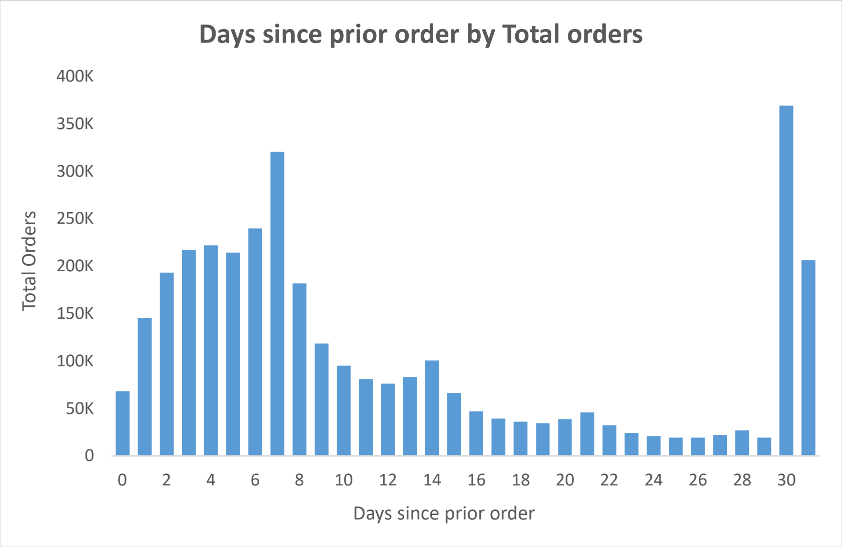
- Organic Blueberries was the least of top 10 re-ordered items (over 105K orders) with 63.13% re-orders

Associations across categories & within were observed for top selling products

Top Products	Products brought together **
Banana	{Organic Sweet Pea, Whole Hearts of Palm}, {Chicken Base, Organic Apple}, {Organic White Onions, Bing Cherries, Organic Granny Smith Apple, Organic Raspberries}
Bag of Organic Bananas	{Organic Sweet Vanilla Bean Nutrition Complete Protein Shake, Organic Strawberries},{Large Alfresco Eggs, Natural Chicken & Maple Breakfast Sausage Patty}
Organic Yellow Onion	{Small Hass Avocado, Organic Large Extra Fancy Fuji Apple, Gluten Free Steel Cut Oats}, {Banana, Multi-Seed Original Crackers, Just Green Unsweetened Tea}
Organic Garlic	{One French Vanilla Nutritional Shake, Whole Wheat Sourdough, Organic Whole Strawberries},{Organic Free Range Chicken Broth, Organic Baby Spinach}
Total 2% Greek Strained Yogurt	{Unsalted Pure Irish Butter, California Cauliflower, Original Instant Oatmeal, Boneless Skinless Chicken Thighs},{Original Real Vegetable Chips, Organic Fuji Apple}
Total 0% Nonfat Greek Yogurt	{Banana, Multi-Seed Original Crackers, Just Green Unsweetened Tea},{Granola Bar, Fig, Cranberry & Hazelnut, Vitamin Water Zero Rise Orange}
Organic Whole String Cheese	{Half & Half, Organic Mango Chunks, Bag of Organic Bananas},{Pure Goat Milk Cheese Log, Organic Navel Orange, Organic Tomato Cluster}
Grated Parmesan	{Lowfat Vanilla Yogurt, Organic Reduced Fat 2% Milk, Butterhead Lettuce, Honeycrisp Apple},{Nine Grain Sourdough Dough, Clementines, Bag}
Sparkling Water Grapefruit	{Bag of Organic Bananas, Large Alfresco Eggs, Natural Chicken & Maple Breakfast Sausage Patty, Organic Coconut Milk},
Spring Water	{Vitamin Water Zero Rise Orange, Original Pure Creamy Almond Milk},{Distilled Water, Organic Baby Spinach, Cole Slaw, Organic Mixed Baby Kale Salad}
100% Raw Coconut Water	{Sparkling Natural Mineral Water, Arancita Rossa, Organic Creamy Peanut Butter},{Sparkling Water Grapefruit, Bag of Organic Bananas, Large Alfresco Eggs}
Original Orange Juice	{Organic Apple Juice, Vanilla Bean Ice Cream, Organic Whole Milk}, {Pure Tart Cherry 100% Juice, Organic Cream Cheese Bar, Half & Half}

** Associations are not exhaustive

7th and 30th day observed highest re-orders with associations coherent with those on other days



Top Products	Products brought together **
Banana	{Organic Sweet Pea, Whole Hearts of Palm}, {Chicken Base, Organic Apple}, {Organic White Onions, Bing Cherries, Organic Granny Smith Apple, Organic Raspberries}
Bag of Organic Bananas	{Organic Sweet Vanilla Bean Nutrition Complete Protein Shake, Organic Strawberries},{Large Alfresco Eggs, Natural Chicken & Maple Breakfast Sausage Patty}
Organic Yellow Onion	{Small Hass Avocado, Organic Large Extra Fancy Fuji Apple, Gluten Free Steel Cut Oats}, {Banana, Multi-Seed Original Crackers, Just Green Unsweetened Tea}
Organic Garlic	{One French Vanilla Nutritional Shake, Whole Wheat Sourdough, Organic Whole Strawberries},{Organic Free Range Chicken Broth, Organic Baby Spinach}

- Most re-orders were observed on 7th and 30th day from the previous order
  - Association of products bought together in basket showed similar trends as those on other days

** Associations are not exhaustive

Amongst the top products ordered individually in cart, soda exhibited the highest standalone ratio however still occurring in groups

Top Products	**Standalone Ratio**	**Associations ***
Bag of Organic Bananas	*0.73%*	{Organic Sweet Vanilla Bean Nutrition Complete Protein Shake, Organic Strawberries}
Soda	*7.37%*	{Crunch Granola Bar Chocolate Chip, Crunch White Chocolate Macadamia Nut Granola Bars}
Spring Water	*3.72%*	{Vitamin Water Zero Rise Orange, Original Pure Creamy Almond Milk}
Banana	*0.43%*	{Organic White Onions, Bing Cherries, Organic Granny Smith Apple, Organic Raspberries}
Organic Baby Spinach	*0.48%*	{Organic Large Grade AA Omega-3 Eggs, 100% Raw Coconut Water}

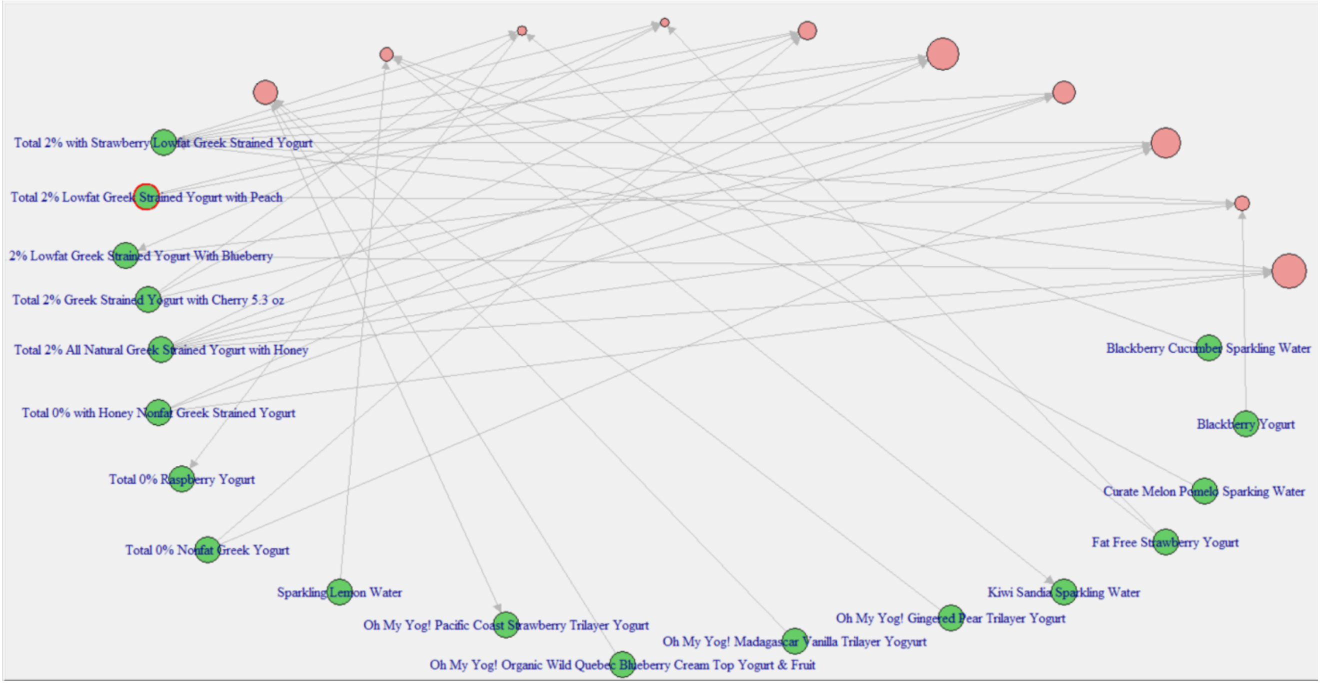
- Cases of standalone products observed were very few
- Such products were more observed more grouped with other products
- Soda & Spring Water exhibited the highest standalone ratio amongst the top products

$$\textit{Standalone Ratio} = \frac{Orders\ with\ Only\ Product\ in\ cart}{Orders\ with\ Product\ in\ Cart}$$

*Associations are not exhaustive

Market Basket Analysis & Results

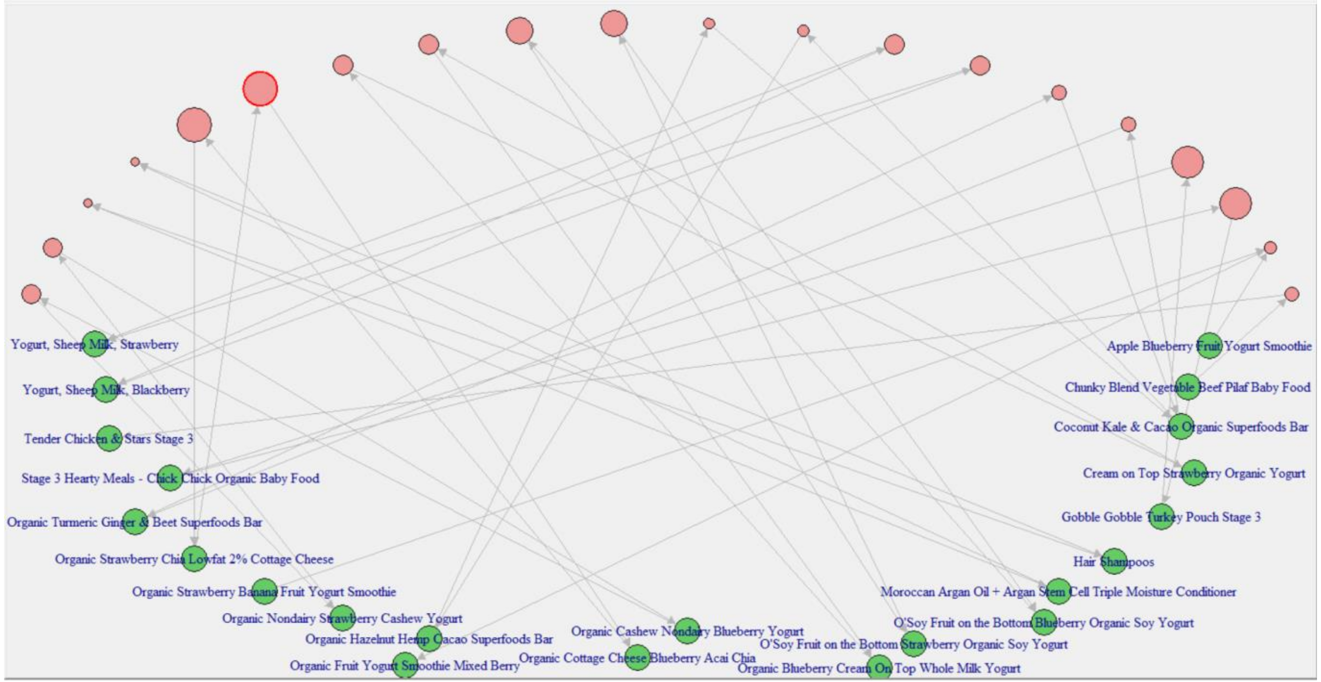
Associations with category ‘Fresh Vegetables’ were observed to show highest affinity


Total 2% with Strawberry Lowfat Greek Strained Yogurt
Total 2% Lowfat Greek Strained Yogurt with Peach
2% Lowfat Greek Strained Yogurt With Blueberry
Total 2% Greek Strained Yogurt with Cherry 5.3 oz
Total 2% All Natural Greek Strained Yogurt with Honey
Total 0% with Honey Nonfat Greek Strained Yogurt
Total 0% Raspberry Yogurt
Total 0% Nonfat Greek Yogurt
Sparkling Lemon Water
Oh My Yog! Pacific Coast Strawberry Trilayer Yogurt
Oh My Yog! Organic Wild Quebec Blueberry Cream Top Yogurt & Fruit
Oh My Yog! Madagascar Vanilla Trilayer Yogyurt
Oh My Yog! Gingered Pear Trilayer Yogurt
Kiwi Sandia Sparkling Water
Fat Free Strawberry Yogurt
Curate Melon Pomelo Sparking Water
Blackberry Yogurt
Blackberry Cucumber Sparkling Water

- A support threshold of ‘0.0001’ with confidence of ‘0.8’ returned 59,430 rules
- ‘Fresh Vegetables’ showed the highest lift of 2.2
- The most associated club was ‘{fresh herbs, meat counter, tofu meat alternatives}’ with 437 such associations

Note: Top 20 associations are showcased

Strong associations were observed for products of type ‘Yogurt’


Yogurt, Sheep Milk, Strawberry
Yogurt, Sheep Milk, Blackberry
Tender Chicken & Stars Stage 3
Stage 3 Hearty Meals - Chick Chick Organic Baby Food
Organic Turmeric Ginger & Beet Superfoods Bar
Organic Strawberry Chia Lowfat 2% Cottage Cheese
Organic Strawberry Banana Fruit Yogurt Smoothie
Organic Nondairy Strawberry Cashew Yogurt
Organic Hazelnut Hemp Cacao Superfoods Bar
Organic Fruit Yogurt Smoothie Mixed Berry
Organic Cottage Cheese Blueberry Acai Chia
Organic Cashew Nondairy Blueberry Yogurt
Organic Blueberry Cream On Top Whole Milk Yogurt
O'Soy Fruit on the Bottom Strawberry Organic Soy Yogurt
O'Soy Fruit on the Bottom Blueberry Organic Soy Yogurt
Moroccan Argan Oil + Argan Stem Cell Triple Moisture Conditioner
Hair Shampoos
Gobble Gobble Turkey Pouch Stage 3
Cream on Top Strawberry Organic Yogurt
Coconut Kale & Cacao Organic Superfoods Bar
Chunky Blend Vegetable Beef Pilaf Baby Food
Apple Blueberry Fruit Yogurt Smoothie

Note: Top 20 associations are showcased

- A support threshold of ‘0.0001’ with confidence of ‘0.1’ returned 70,437 rules
- ‘Organic Nondairy Strawberry Cashew Yogurt’ showed the highest lift of 2050
- The most associated club was ‘{Organic Nondairy Blueberry Cashew Yogurt}’ with 416 such associations

New Functions Used & Context

Function Name	Why was it used? / What does it do?
Apriori	To generate association rules across transaction with corresponding support, confidence and lift values
rbind	Used to combine data from two datasets : order_products__train, order_products__prior
left_join	To perform left join on the provided pair of datasets
inner_join	To perform inner join on the provided pair of datasets
colSums	To calculate columnar summation of Nas in dataset
unique.data.frame	To obtain unique values of specified columns in a dataset
coord_flip	To transpose a bar chart for better representation
top_n	To select top ‘n’ entities from a dataset
colnames	To enumerate the column names of the respective dataset
as.data.frame	To convert the desired output to a dataframe

R Code & Dataset link for the Project

R Code: Final Project - R Code

Dataset : Instacart Data

Attributes of the datasets used

- Aisles (Category):
  - Aisle Id: A unique Id to represent each aisle
  - Aisle: Contains the name of aisle based on products on the aisle
- Departments:
  - Department Id : Unique integer to represent each department
  - Department : String which tells the name of department depending upon products in department
- Orders:
  - Order Id : Unique integer to represent order
  - User Id : Unique integer to represent different users
  - Eval Set : Tells whether order is from Prior or Train
  - Order Number : Order number for the order made by customer
  - Order Dow : Ranges from 0-6 where 0 = Sunday and 6 = Saturday
  - Order Hour Of Day : Ranges from 0-23 where 0 = 12 AM and 23 = 11 PM
  - Day Since Prior Order : Number of days since last order is placed

Attributes of the datasets used

- Products:
  - Product Id : Unique Id for each product since there are huge number of products so will have large range
  - Product Name : Name of the product
  - Aisle Id : Id of the aisle where the product is present
  - Department Id : Unique Id for each department
- Order Products Prior & Order Products Train:
  - Order Id : Unique integer to represent order
  - Product Id : Unique integer for each product
  - Add to cart order : Order in which product is added in the cart
  - Reordered : Binary variable (0 = Not reordered /1 = Reordered)

Appendix

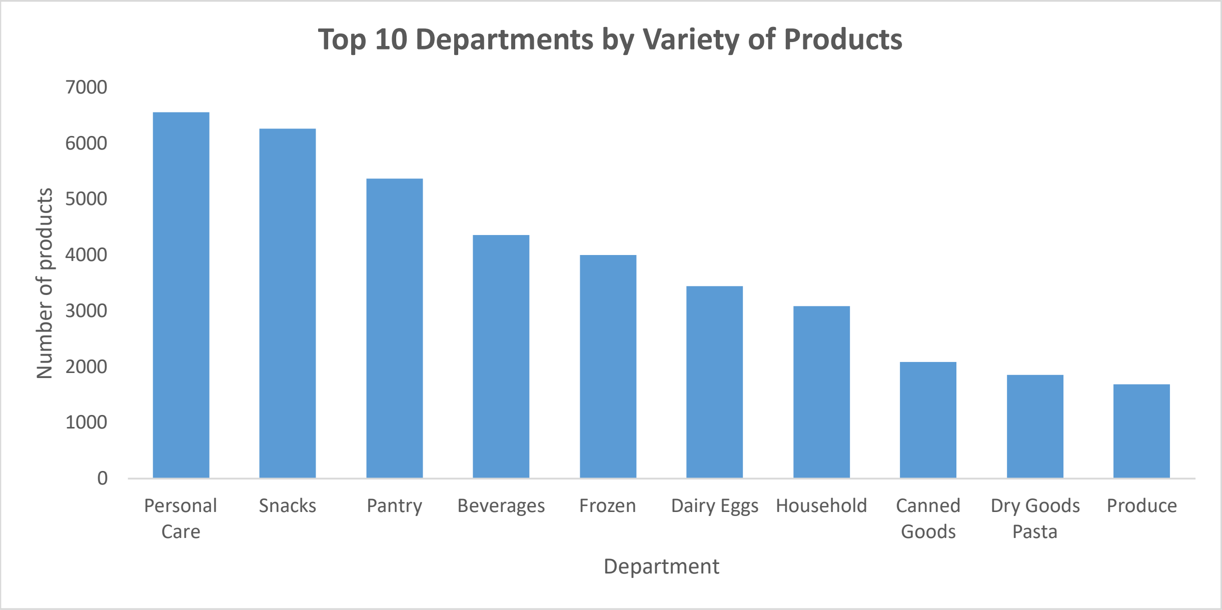
Market Basket Analysis Overview

- Market Basket Analysis is one of the key techniques used by retailers and e-commerce sites to understand which products are bought together by customers

- This understanding of associated products help companies to make effective product recommendation by identifying relationships between the items that people buy

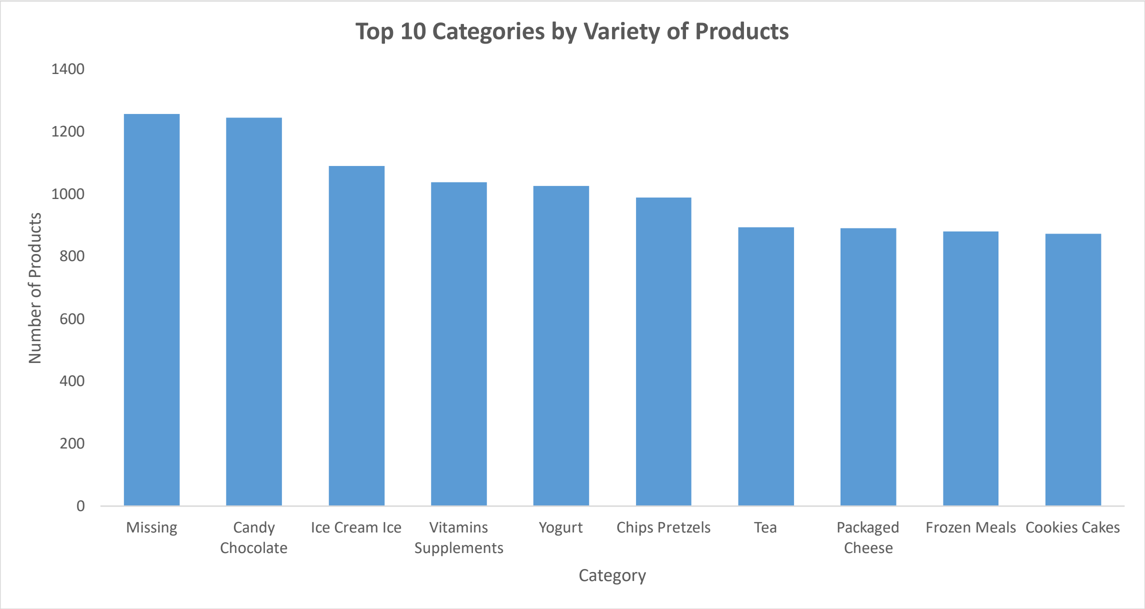
- Association Rules are widely used to analyze retail basket or transaction data and are intended to identify strong rules discovered in transaction data using measures of interestingness, based on the concept of strong rules

Personal Care and Snacks were the top departments by variety of products



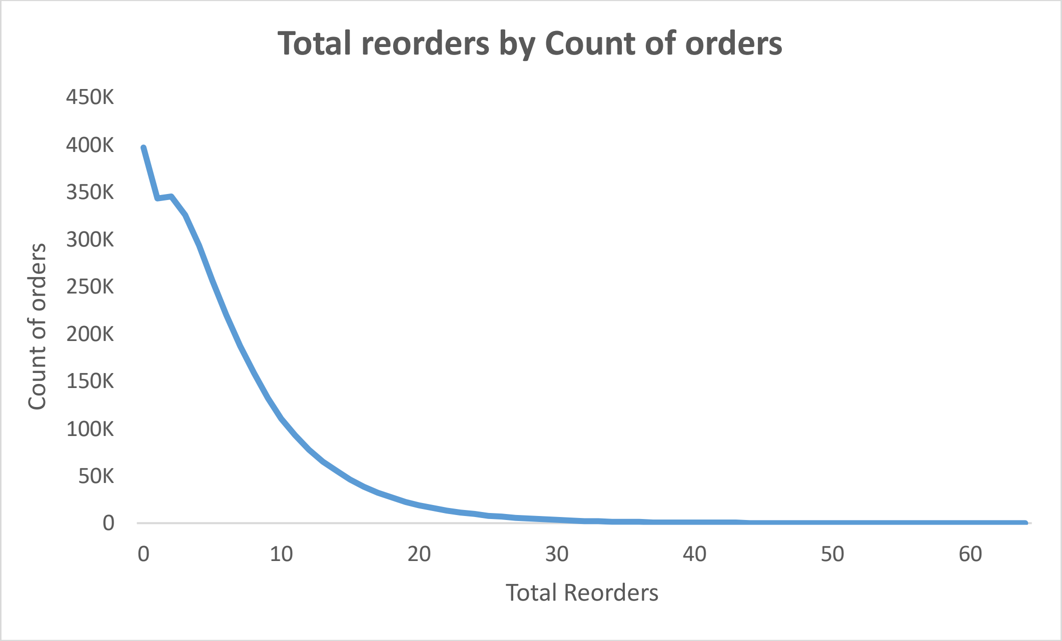
- Personal Care and Snacks are the departments with the highest variety of products
- Pantry and Breakfast are the 3rd and 4th highest department

Candy Chocolate was the top category by variety of products



- Candy Chocolate was the top category with highest variety of products
- Ice Cream, Vitamins Supplements, Yogurt and Chips Pretzels were other top categories by variety

Most of the reorder have less than 10 reordered items



- Most of the orders have less than 10 reordered items
- Number of orders begin to decrease as re-ordered products in cart increase

Thank You!