Analyzing Customer Purchase Patterns and Recommendations

Presented by: Abdelrahman Hatem

Motivation

- Ecommerce is bigger than ever. Experts predict that the industry will see 39% growth over the coming years, with shoppers expecting to spend a collective $8 trillion online by 2027.

- With such explosive growth comes the opportunity to reach new shoppers and generate more revenue. That's only possible if you know where the industry is heading, what customers expect, and how their purchasing habits are evolving

- Keeping your finger on the pulse of industry trends is important when it comes to staying ahead of competition. But identifying valuable insights within your own business is just as crucial for growing your online sales.
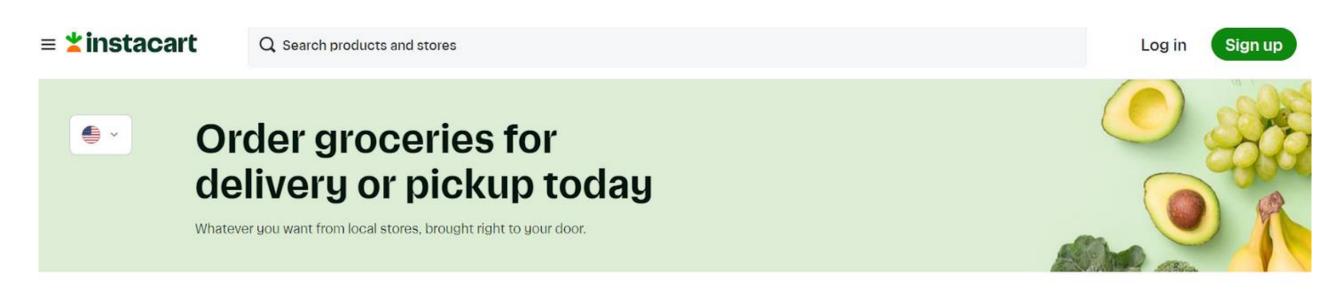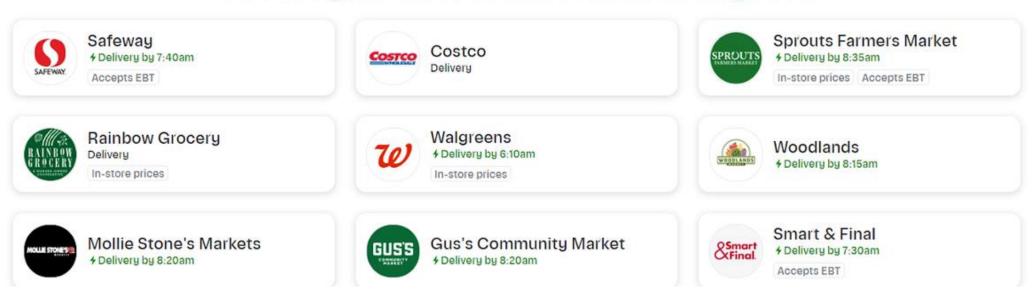
Problem Definition

What's instacart?

Instacart is an online delivery service that lets you shop at nearby grocery stores, wine shops, bulk warehouses, convenience and pet stores, and other retailers. Sign up for a free account or a paid membership (Instacart+) with your email address to shop stores in your area. Costs will vary depending on retailer, order size, and membership status

Problem Definition

Introduction

- In today's rapidly evolving world, driven by advancements in the Internet of Things, the need for innovation is increasingly important. According to Statista, a leading statistics portal, 1.8 billion people globally engaged in online shopping in 2018, with this number expected to rise in the coming years. There is ongoing research and development efforts aimed at meeting the growing demands of consumers and ensuring their satisfaction. Analyzing transaction details is a key technique used to understand customer purchasing behavior.

- Understanding transactions is essential for any business, as it directly impacts sales. This is particularly true in retail, where understanding customer purchasing patterns and the relationships between products sold together can lead to increased sales through impulse buying and a deeper understanding of customer preferences.

- To gain insights into market basket analysis, I have chosen to analyze the transactional data from Instacart. Instacart is an e-commerce platform that enables users to purchase groceries from local stores online, with personal shoppers fulfilling and delivering orders on the same day. While this process allows retailers to analyze user purchase behaviors, understanding customer purchasing patterns and behaviors can be challenging and time-consuming.

Objective:

- The objective is to anticipate the products that will feature in a user's forthcoming order.

- This goal aligns with the predictive modeling aspect mentioned earlier, where machine learning models can be trained to predict user behavior based on past orders and other relevant data.

- By predicting future purchases, Instacart can optimize its recommendations and provide a more personalized shopping experience for users.

Data Overview

Data is divided across 6 files:

orders [~3.42M orders]

	order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
0	2539329	1	prior	1	2	8	NaN
1	2398795	1	prior	2	3	7	15.0
2	473747	1	prior	3	3	12	21.0
3	2254736	1	prior	4	4	7	29.0
4	431534	1	prior	5	4	15	28.0
5	3367565	1	prior	6	2	7	19.0
6	550135	1	prior	7	1	9	20.0
7	3108588	1	prior	8	1	14	14.0
8	2295261	1	prior	9	1	16	0.0
9	2550362	1	prior	10	4	8	30.0

Products [~49.6k Products]

	product_id	product_name	aisle_id	department_id
0	1	Chocolate Sandwich Cookies	61	19
1	2	All-Seasons Salt	104	13
2	3	Robust Golden Unsweetened Oolong Tea	94	7
3	4	Smart Ones Classic Favorites Mini Rigatoni Wit...	38	1
4	5	Green Chile Anytime Sauce	5	13
5	6	Dry Nose Oil	11	11
6	7	Pure Coconut Water With Orange	98	7
7	8	Cut Russet Potatoes Steam N' Mash	116	1
8	9	Light Strawberry Blueberry Yogurt	120	16
9	10	Sparkling Orange Juice & Prickly Pear Beverage	115	7

order_products__train [~1.3M rows]

	order_id	product_id	add_to_cart_order	reordered
0	1	49302	1	1
1	1	11109	2	1
2	1	10246	3	0
3	1	49683	4	0
4	1	43633	5	1
5	1	13176	6	0
6	1	47209	7	0
7	1	22035	8	1
8	36	39612	1	0
9	36	19660	2	1

order_products__Prior [~32.4M rows]

	order_id	product_id	add_to_cart_order	reordered
0	2	33120	1	1
1	2	28985	2	1
2	2	9327	3	0
3	2	45918	4	1
4	2	30035	5	0
5	2	17794	6	1
6	2	40141	7	1
7	2	1819	8	1
8	2	43668	9	0

Departments [21department]

	department_id	department
0	1	frozen
1	2	other
2	3	bakery
3	4	produce
4	5	alcohol
5	6	international
6	7	beverages
7	8	pets
8	9	dry goods pasta
9	10	bulk

Aisles [134 sub-Department]

	aisle_id	aisle
0	1	prepared soups salads
1	2	specialty cheeses
2	3	energy granola bars
3	4	instant foods
4	5	marinades meat preparation
5	6	other
6	7	packaged meat
7	8	bakery desserts
8	9	pasta sauce
9	10	kitchen supplies

Data Preprocessing

Converting Hour of Day to 12-Hour Format

- We have a function convert_to_12_hour(hour) that converts the hour of the day into a 12-hour format.

order_hour_of_day	order_hour_of_day_12
8	8 AM
7	7 AM
12	12 PM
7	7 AM
15	3 PM
...	...
18	6 PM
10	10 AM
12	12 PM
12	12 PM
14	2 PM

Data Preprocessing

Converting the number of the day to its corresponding name can contribute to clearer communication, easier interpretation, and more intuitive analysis of temporal data, particularly when dealing with days of the week.

```python
# Dictionary mapping numbers to day names
day_of_week_mapping = {
    0: 'Sunday',
    1: 'Monday',
    2: 'Tuesday',
    3: 'Wednesday',
    4: 'Thursday',
    5: 'Friday',
    6: 'Saturday'
}

orders['order_dow_name'] = orders['order_dow'].apply(lambda x: day_of_week_mapping[x])
```

EDA

Orders over the entire dataset

- There are 3,214,874 orders in total. Out of which, the last purchase of 131,209 orders are given as train set and we need to predict for the rest 75,000 orders.
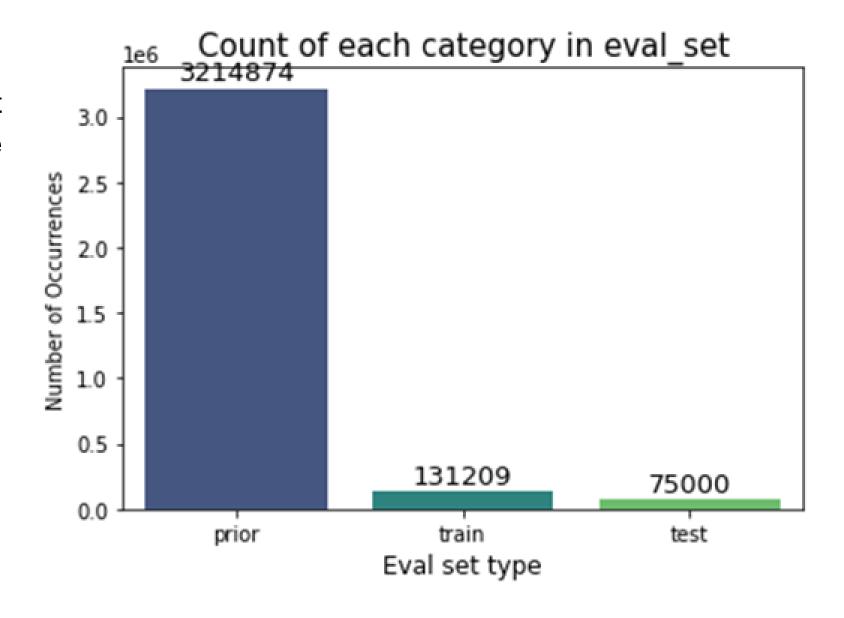
Clients over the entire dataset

```
eval_set
prior    206209
test      75000
train    131209
Name: user_id, dtype: int64
```


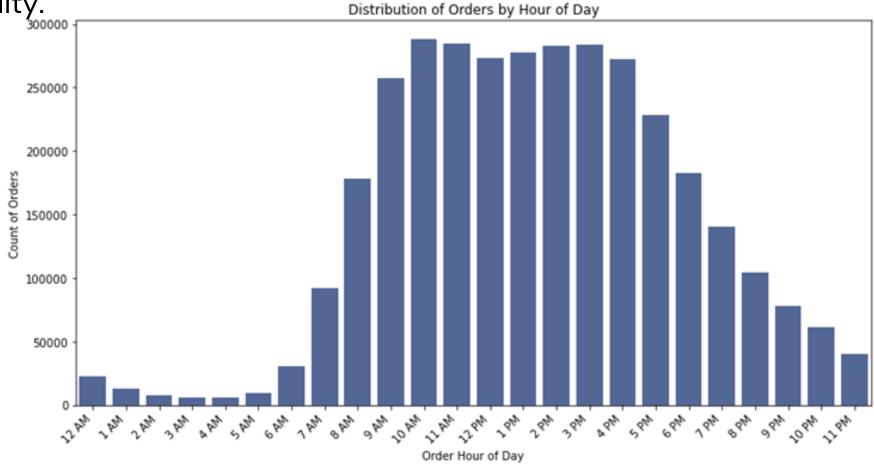Count of each category in eval_set

EDA

Time of orders

Distribution of Orders by Hour of Day:
Analyzing the distribution of orders by hour of the day provides valuable insights that can help businesses optimize their operations, improve customer experience, and drive growth and profitability.

Typically, orders are predominantly placed during daytime hours, typically between 8:00 AM and 5:00 PM.
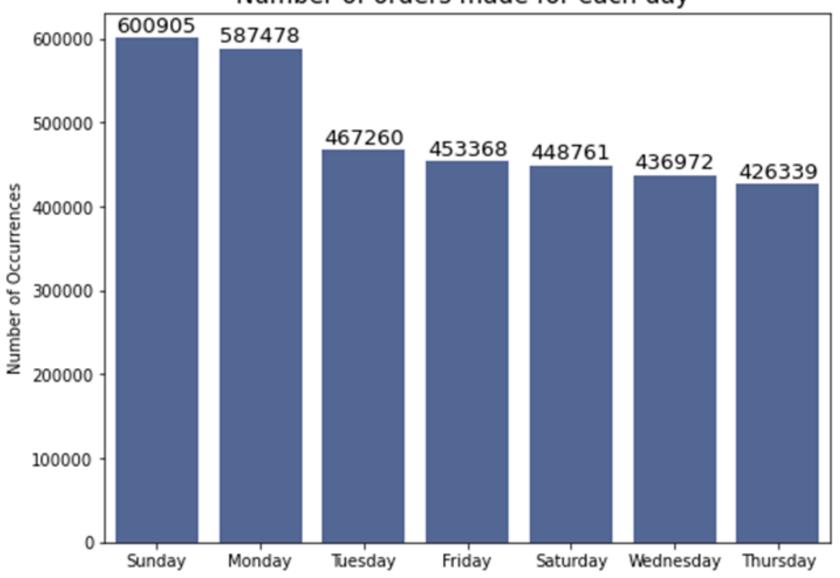

Distribution of Orders by Hour of Day

EDA

Days of Orders in a week

The data on the number of orders for each day of the week reveals several key insights. Weekends, particularly Sunday and Saturday, experience the highest order volumes, indicating peak demand. In contrast, there is a midweek slump, with Wednesday having the lowest number of orders among weekdays.
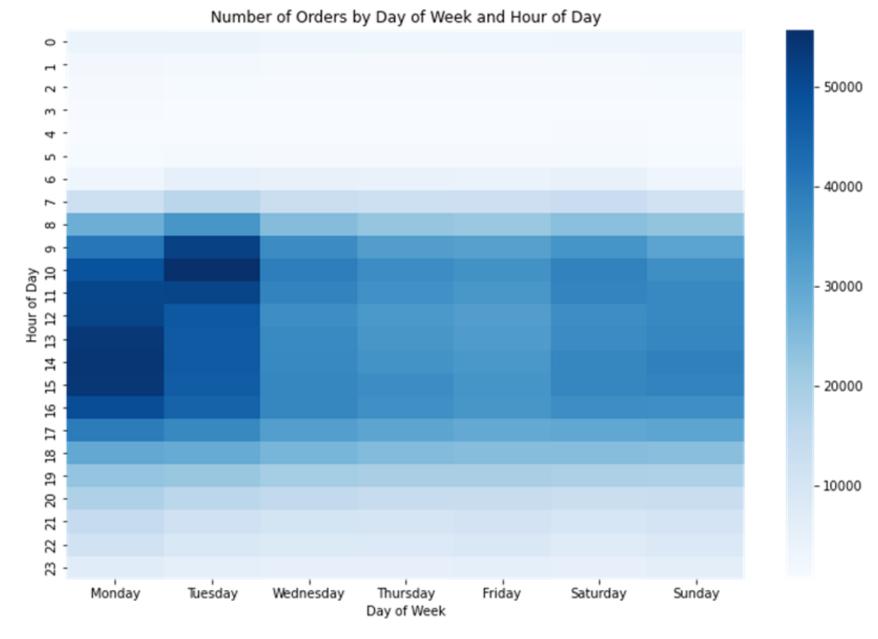

Number of orders made for each day

EDA

Number of Orders by Day of Week and Hour of Day

Combining the day of the week and the hour of the day to see the distribution of orders can provide a more granular understanding of when orders are placed throughout the week.

Number of Orders by Day of Week and Hour of Day

EDA

instacart

Analyzing products

Bestsellers Let's have look which products are sold most often (top10)

Banana: 491,291 units sold

Bag of Organic Bananas: 394,930 units sold.

Organic Strawberries: 275,577 units sold.

Organic Baby Spinach: 251,705 units sold.

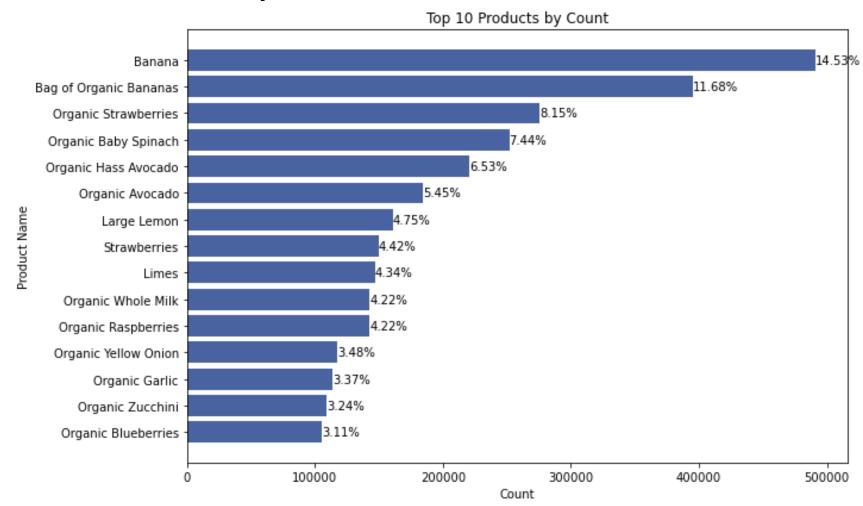Organic Hass Avocado: 220,877 units sold.

Organic Avocado: 184,224 units sold.

Large Lemon: 160,792 units sold.

Strawberries: 149,445 units sold

Limes: 146,660 units sold

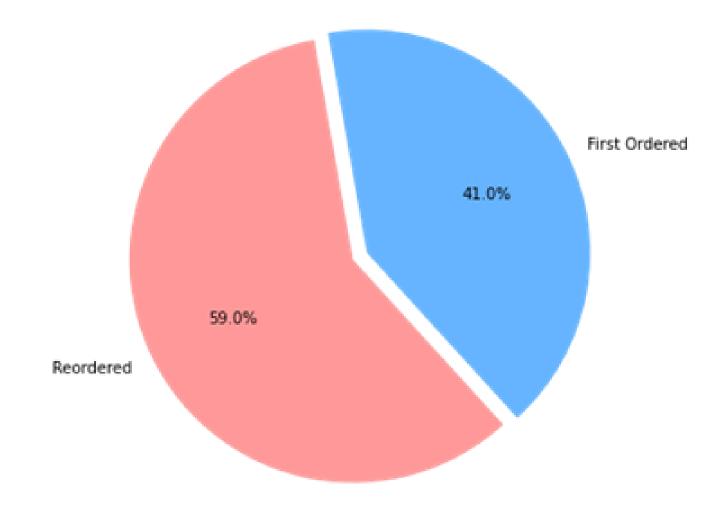Organic Whole Milk: 142,813 units sold.

Top 10 Products by Count

EDA | instacart

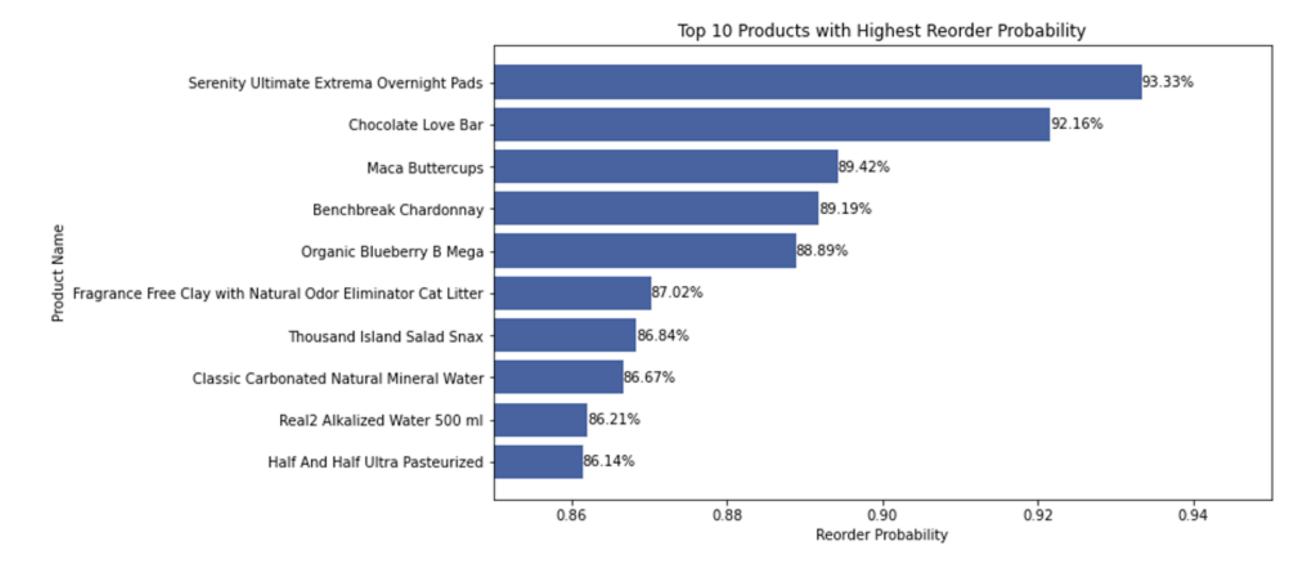**Percentage of Reordered and First Ordered Products in Prior Dataset**

In prior orders, 58.9697% of products were re-ordered.

In train orders, 59.8594% of products were re-ordered.

Percentage of orders without any reordered product: 11.868056%

Percentage of orders with all products re-ordered: 21.493998%

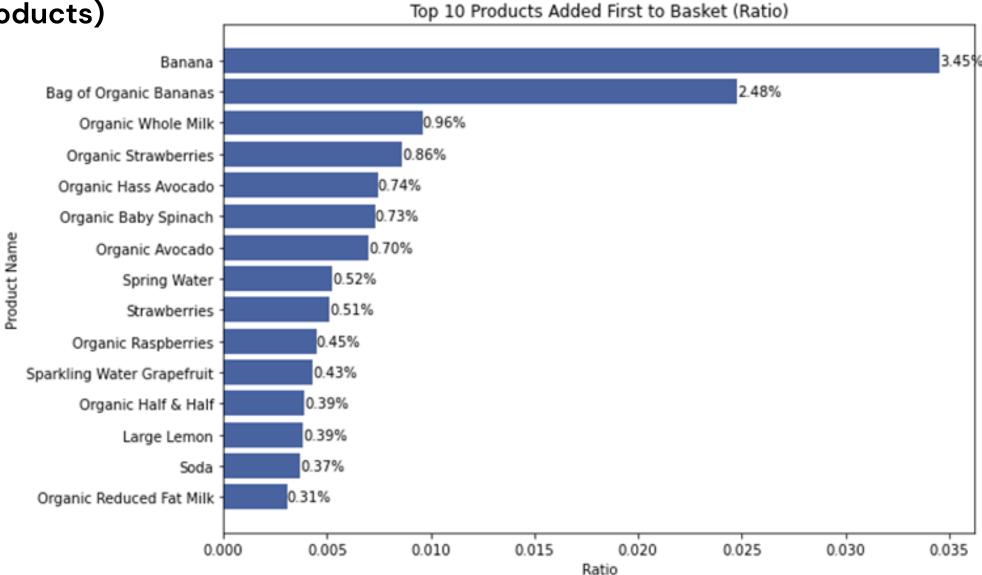Percentage of Reordered and First Ordered Products in Prior Dataset

First Ordered 41.0%

Reordered 59.0%

EDA | instacart

Most Reordered Products (Analyzing products)

Top 10 Products with Highest Reorder Probability

Product Name	Reorder Probability
Serenity Ultimate Extrema Overnight Pads	93.33%
Chocolate Love Bar	92.16%
Maca Buttercups	89.42%
Benchbreak Chardonnay	89.19%
Organic Blueberry B Mega	88.89%
Fragrance Free Clay with Natural Odor Eliminator Cat Litter	87.02%
Thousand Island Salad Snax	86.84%
Classic Carbonated Natural Mineral Water	86.67%
Real2 Alkalized Water 500 ml	86.21%
Half And Half Ultra Pasteurized	86.14%

EDA

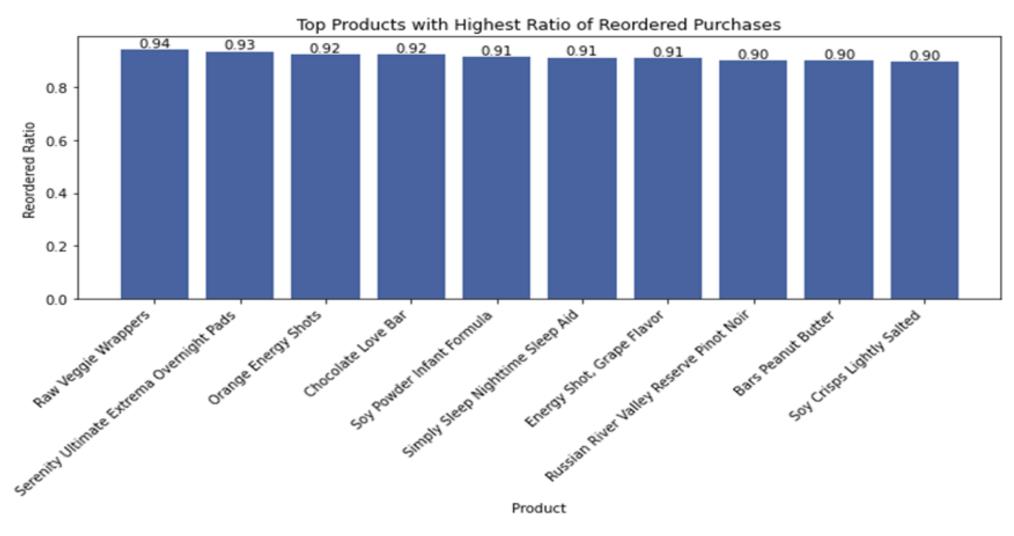**Top 10 Products Added First to Basket (Analyzing products)**

• The data reveals that certain products, such as Bananas, Bag of Organic Bananas, and Organic Whole Milk, are consistently among the top items added first to the basket. This suggests that these products hold significant importance to customers and are likely considered essential or preferred choices.

Top 10 Products Added First to Basket (Ratio)

Product Name	Ratio
Banana	3.45%
Bag of Organic Bananas	2.48%
Organic Whole Milk	0.96%
Organic Strawberries	0.86%
Organic Hass Avocado	0.74%
Organic Baby Spinach	0.73%
Organic Avocado	0.70%
Spring Water	0.52%
Strawberries	0.51%
Organic Raspberries	0.45%
Sparkling Water Grapefruit	0.43%
Organic Half & Half	0.39%
Large Lemon	0.39%
Soda	0.37%
Organic Reduced Fat Milk	0.31%

EDA

Analyzing the top products with the highest ratio of reordered purchases.

Products like "Raw Veggie Wrappers, "Serenity Ultimate Extrema Overnight Pads," and "Orange Energy Shots"   exhibit very high reordered ratios (ranging from 90% to 94%). This suggests a high level of customer loyalty, as a significant proportion of customers who purchase these items tend to repurchase them.
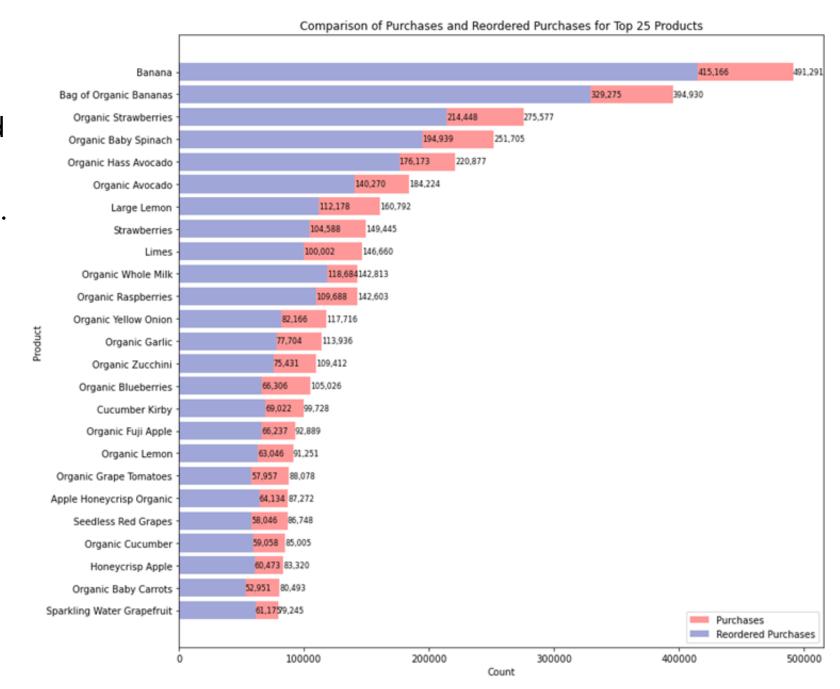

Top Products with Highest Ratio of Reordered Purchases

Reordered Ratio		

0.94 0.93 0.92 0.92 0.91 0.91 0.91 0.90 0.90 0.90

Raw Veggie Wrappers · Serenity Ultimate Extrema Overnight Pads · Orange Energy Shots · Chocolate Love Bar · Soy Powder Infant Formula · Simply Sleep Nighttime Sleep Aid · Energy Shot, Grape Flavor · Russian River Valley Reserve Pinot Noir · Bars Peanut Butter · Soy Crisps Lightly Salted

Product

EDA

Defining and analyzing the comparison between purchases and reordered purchases.

• Top Purchased Items: "Banana," "Bag of Organic Bananas," and "Organic Strawberries" are among the most frequently purchased items, indicating their popularity among customers.

• Organic Preference: The presence of organic items indicates a preference for organic products among customers.

Healthy Choices: Products such as "Organic Whole Milk," "Organic Baby Carrots," and "Organic Cucumber" reflect a trend towards healthier food choices.
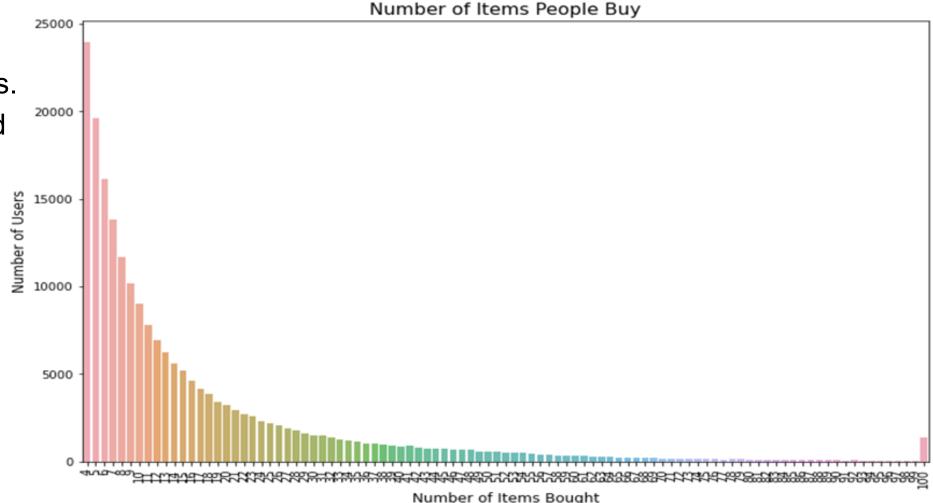

Comparison of Purchases and Reordered Purchases for Top 25 Products

Product	Purchases	Reordered Purchases
Banana	491,291	415,166
Bag of Organic Bananas	394,930	329,275
Organic Strawberries	275,577	214,448
Organic Baby Spinach	251,705	194,939
Organic Hass Avocado	220,877	176,173
Organic Avocado	184,224	140,270
Large Lemon	160,792	112,178
Strawberries	149,445	104,588
Limes	146,660	100,002
Organic Whole Milk	142,813	118,684
Organic Raspberries	142,603	109,688
Organic Yellow Onion	117,716	82,166
Organic Garlic	113,936	77,704
Organic Zucchini	109,412	75,431
Organic Blueberries	105,026	66,306
Cucumber Kirby	99,728	69,022
Organic Fuji Apple	92,889	66,237
Organic Lemon	91,251	63,046
Organic Grape Tomatoes	88,078	57,957
Apple Honeycrisp Organic	87,272	64,134
Seedless Red Grapes	86,748	58,046
Organic Cucumber	85,005	59,058
Honeycrisp Apple	83,320	60,473
Organic Baby Carrots	80,493	52,951
Sparkling Water Grapefruit	59,245	61,175

EDA

instacart

Knowing the number of items people buy:

According to what appears people usually order around 4 items. The distributions of item numbers are similar between train and pre-order sets.
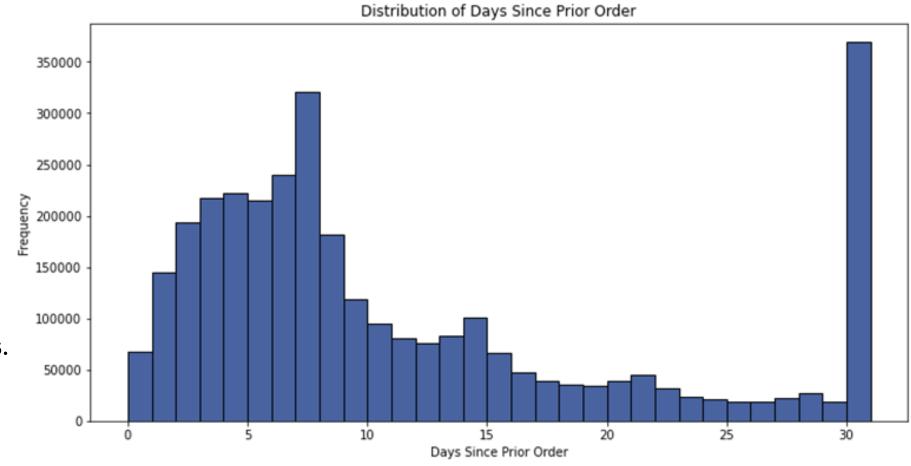- Minimum number of items bought per user: 4.
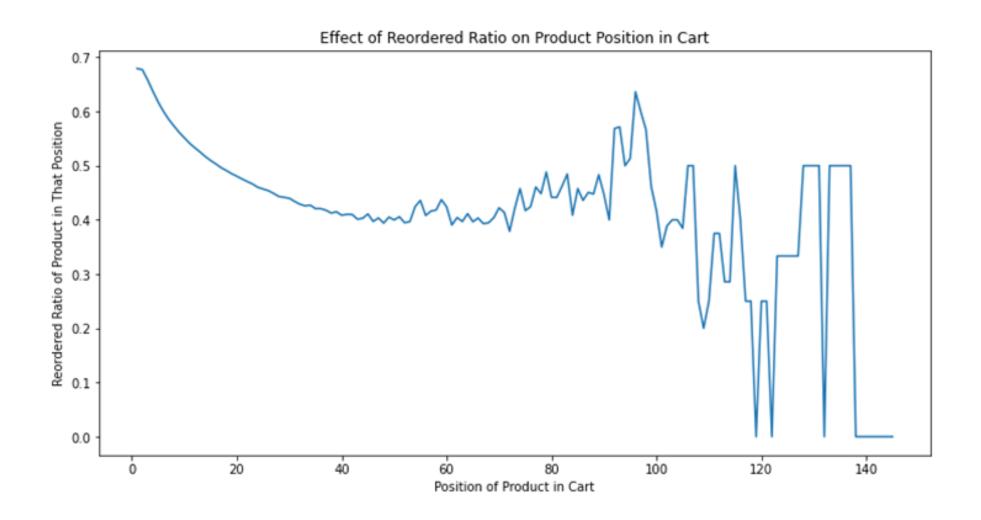- Maximum number of items bought per user: 100.

Number of Items People Buy

EDA

instacart

studying the count of days since the last order

Customers tend to place orders more frequently within the first week since their prior order, with a peak observed between 1 to 7 days. This suggests a significant portion of customers tend to reorder within a week's time. Conversely, order frequency declines notably after 14 days since the prior order, with the lowest frequency observed beyond two weeks. However, there is an increase in orders observed after 30 days.

Distribution of Days Since Prior Order

EDA

It appears that products added to the cart initially have a higher probability of being reordered compared to those added later. This observation aligns with the common behavior of first adding frequently purchased items to the cart and then exploring new products. This trend suggests that customers prioritize replenishing familiar items before considering new ones during their shopping sessions.

EDA

instacart
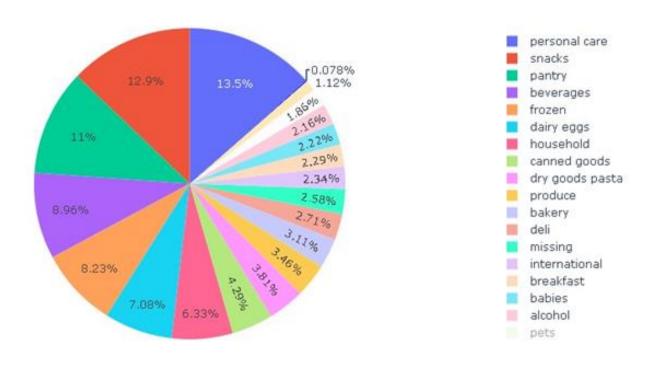
Analyzing the distribution of products across departments. Analyzing departments

Distribution of Products Across Departments (Treemap)

Distribution of Products Across Departments (Pie Chart)

EDA

Analyzing the distribution of purchases across departments provides valuable insights into customer behavior.
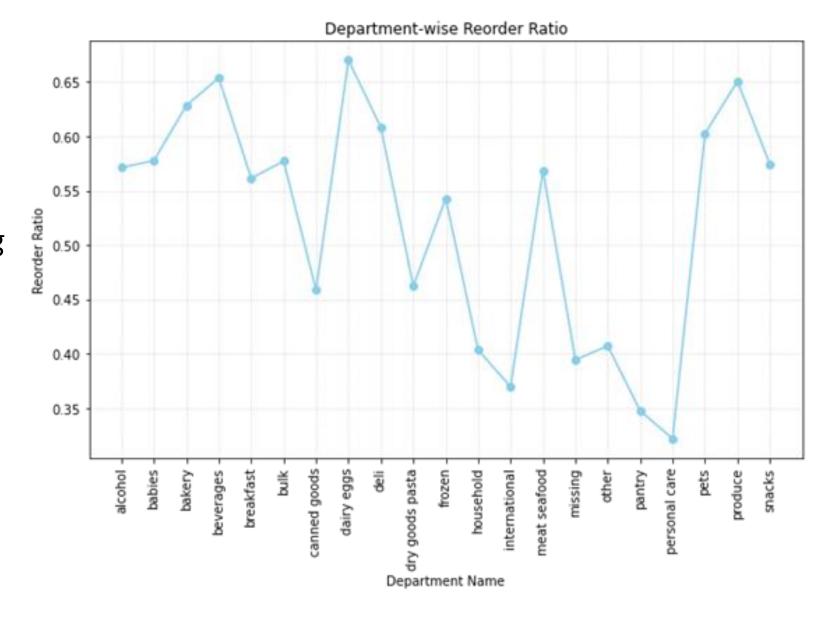
It appears that while the "Produce" department comprises only a small fraction of the total number of products available (2%), it accounts for a significant portion (29%) of the total purchases made.

EDA
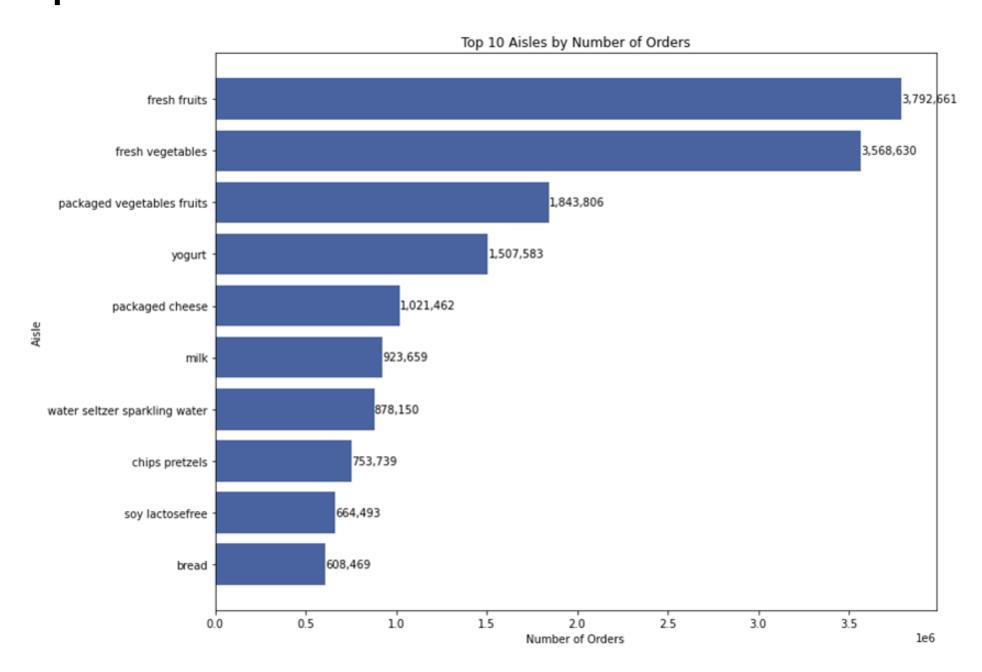
Analyzing departments
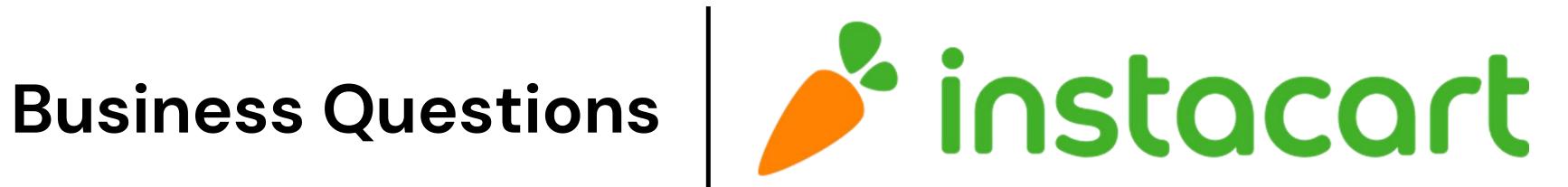
Customers who purchase items from the Dairy & Eggs and Produce departments are likely to exhibit higher reordering tendencies compared to those who buy from the Personal Care category.

EDA

- The top two aisles are fresh fruits and fresh vegetables.

Top 10 Aisles by Number of Orders

Aisle	Number of Orders
fresh fruits	3,792,661
fresh vegetables	3,568,630
packaged vegetables fruits	1,843,806
yogurt	1,507,583
packaged cheese	1,021,462
milk	923,659
water seltzer sparkling water	878,150
chips pretzels	753,739
soy lactosefree	664,493
bread	608,469

Business Questions | instacart
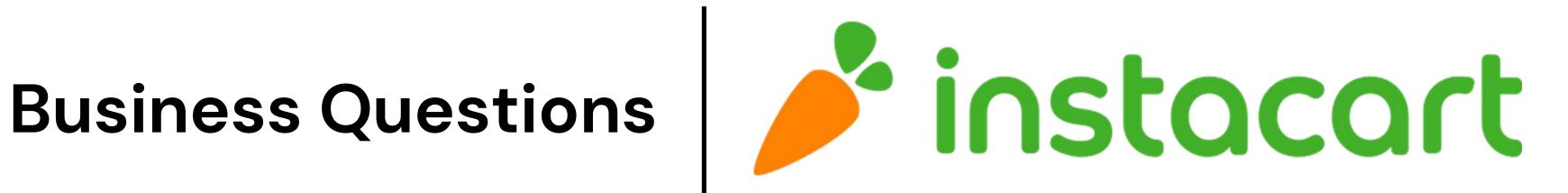
Customer Engagement and Retention:

- How can Instacart improve its user experience to make it more memorable and drive repeat purchases?
- What strategies can be implemented to increase customer loyalty and ensure that customers continue to use Instacart for their grocery shopping needs?

Product Offering Optimization:

- Which aisles or departments currently have a restricted product variety and may benefit from a wider range?

Business Questions

Product Recommendation Strategy:

- What are the best times to introduce new products to clients, taking into account order history, frequency of purchases, and seasonality?
- When should Instacart stop proposing new products to prevent overloading clients or interrupting their regular buying habits?

Personalization and Customization:
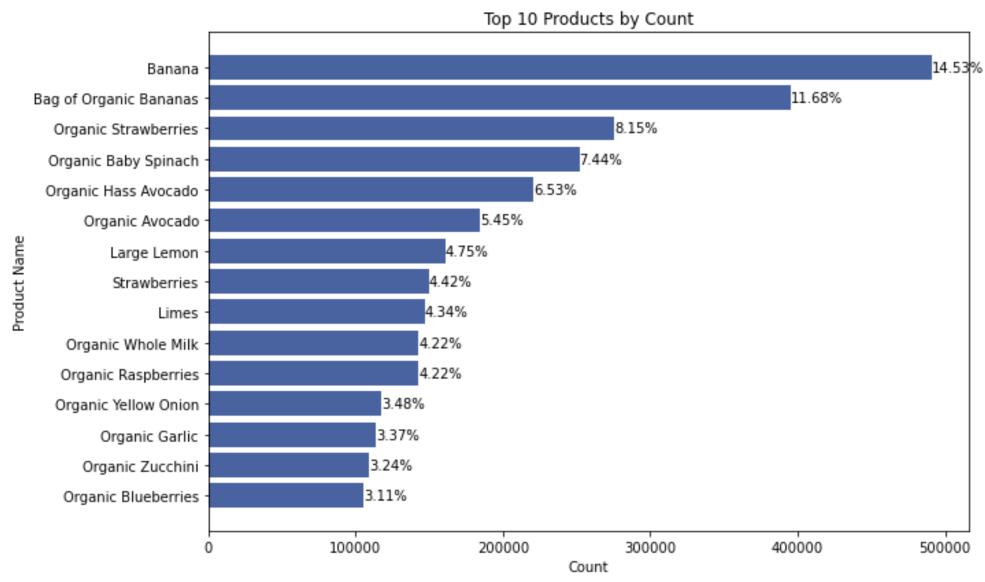
- How can Instacart use customer data and activity trends to make personalized recommendations that appeal to each user?

Business Questions

Can we modify in instacart's portfolio?

We found that some product names holds "organic" in them! So we decided to analyze them.

Top 10 Products by Count

Product Name	Count
Banana	14.53%
Bag of Organic Bananas	11.68%
Organic Strawberries	8.15%
Organic Baby Spinach	7.44%
Organic Hass Avocado	6.53%
Organic Avocado	5.45%
Large Lemon	4.75%
Strawberries	4.42%
Limes	4.34%
Organic Whole Milk	4.22%
Organic Raspberries	4.22%
Organic Yellow Onion	3.48%
Organic Garlic	3.37%
Organic Zucchini	3.24%
Organic Blueberries	3.11%

Business Questions
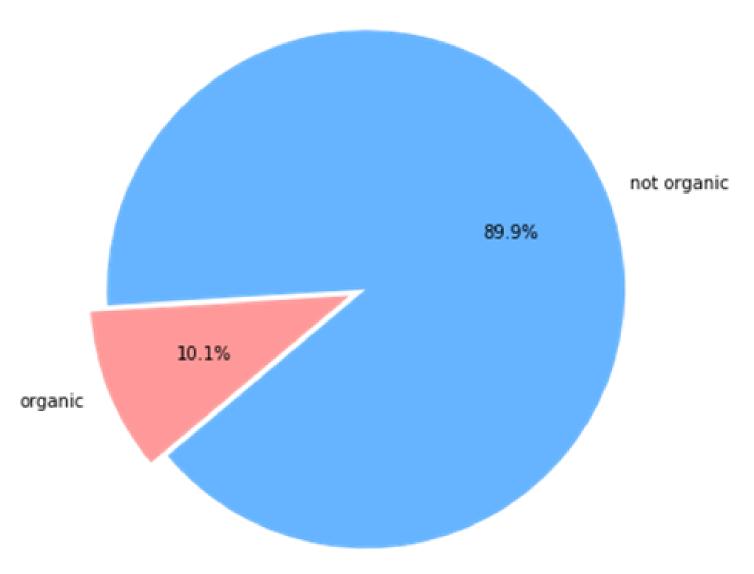
Analyzing organic products

- Percentage of organic products: ~10%
- Percentage of non-organic products: ~90%

They are few!

Ratio of Organic Products in the Supermarket

not organic

89.9%

10.1%

organic

Business Questions
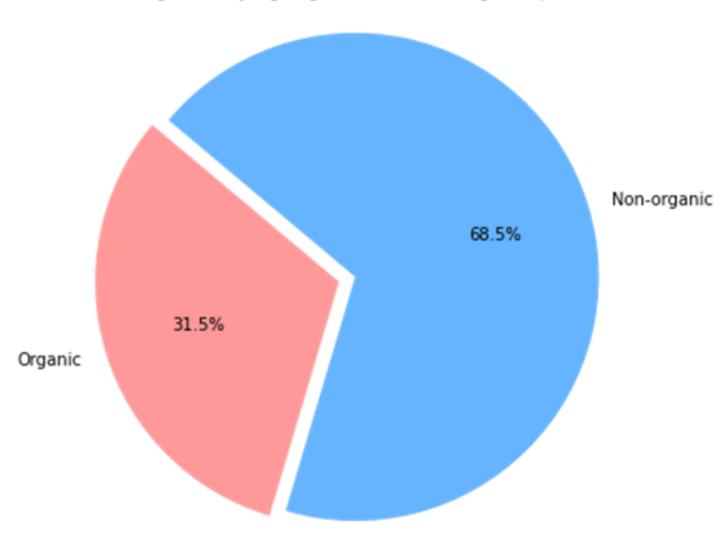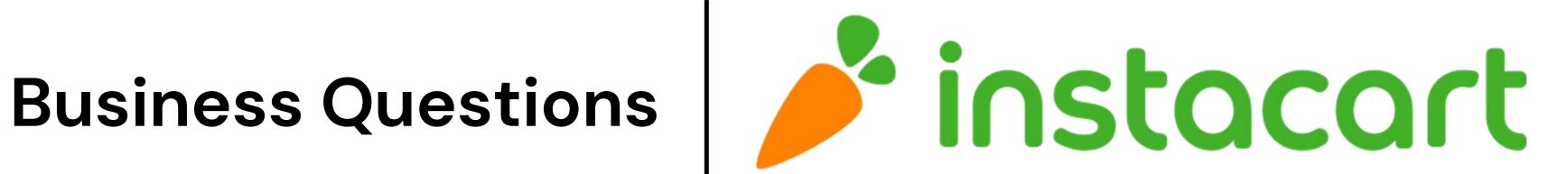
Analyzing organic products

- Percentage of buying organic products: 31.51%
- Percentage of buying non–organic products: 68.49%

They are 31% of purchases!
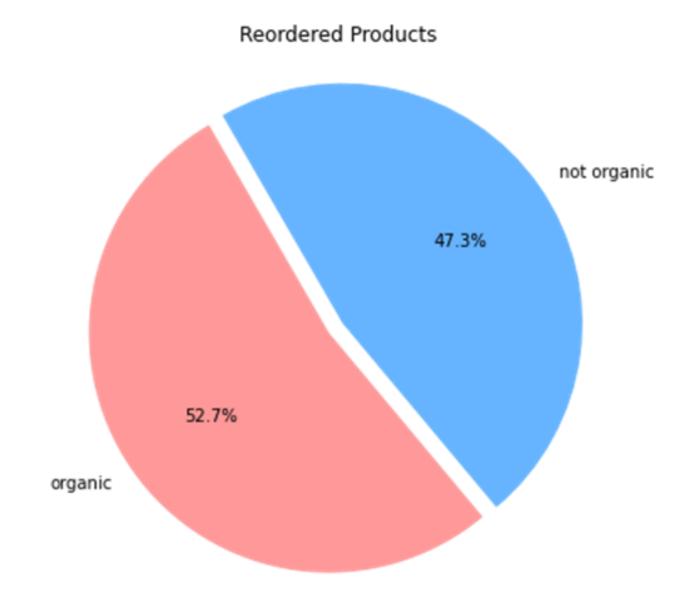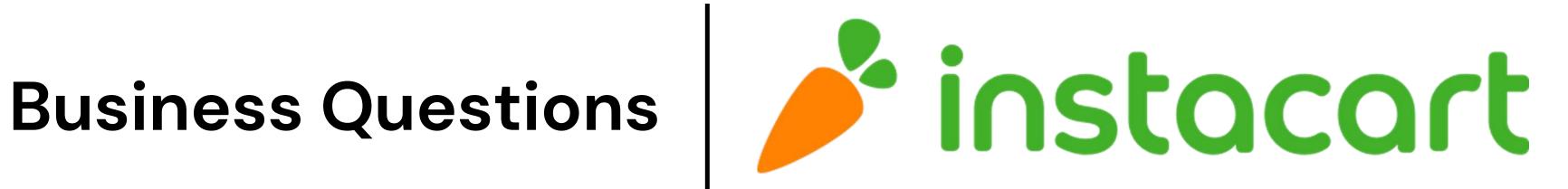

Percentage of buying organic and non-organic products

Non-organic 68.5%

Organic 31.5%

Business Questions

instacart

Analyzing organic products

organic	52.7%
not organic	47.3%

52% of reordered purchases is organic are users who previously experienced and bought these products.

People are loving them!

Reordered Products

not organic — 47.3%

organic — 52.7%

Business Questions | instacart

Can we modify in instacart's portfolio?

- Introducing new organic products to instacart's retailers portfolio.

- Consider introducing new products in the departments that offer few product choices, but these products are highly being purchased.
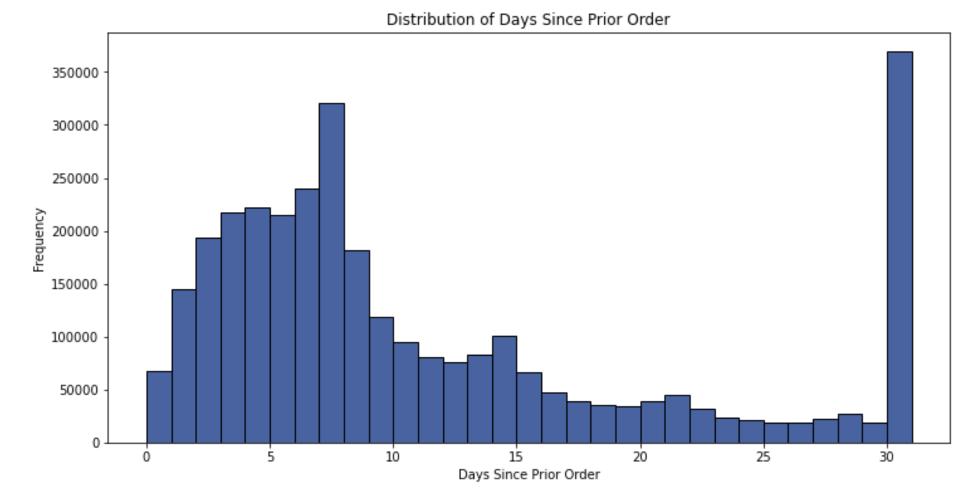
Business Questions

What strategies can be implemented to increase customer loyalty and ensure that customers continue to use Instacart for their grocery shopping needs?

Most users make orders after a week from their last order.

Send reminders to users who haven't ordered since 7 days from their last order.

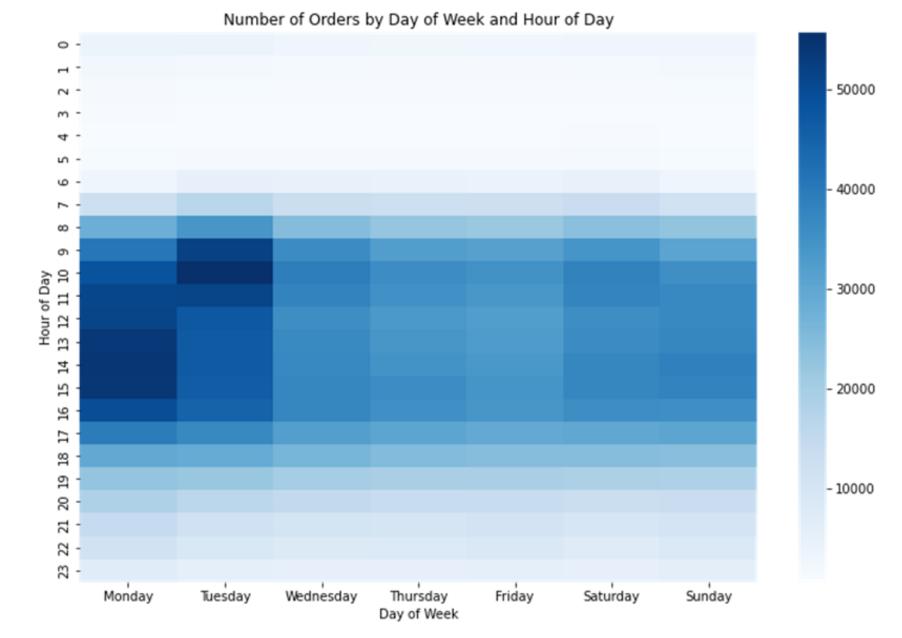Distribution of Days Since Prior Order

Business Questions

What are the best times to introduce new products to clients, taking into account order history, frequency of purchases, and seasonality?

Probability of a user buying during the afternoon of the weekends is high.

Thus can target the weekends to recommend users to try new products they haven't bought before.
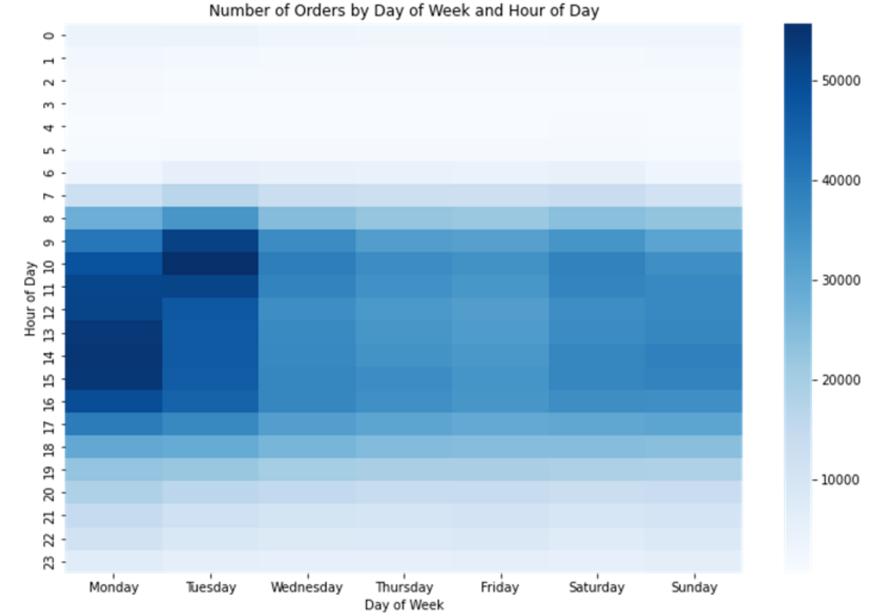
Number of Orders by Day of Week and Hour of Day

Business Questions

When should Instacart stop proposing new products to prevent overloading clients or interrupting their regular buying habits?

By more than 65%, People usually buy previously ordered products from 6:00AM to 8:00AM

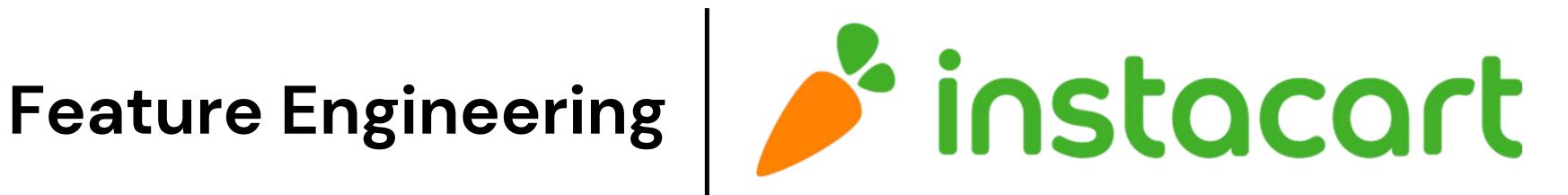Recommend previously ordered products at these hours, while avoiding recommending new products at these hours.

Number of Orders by Day of Week and Hour of Day

Predictive Analysis Model

What do we want? We want to predict the products that will be in user's next future order.

Forming the data: We take each user with his/her previously ordered products, and form each record to be a user-product pair.

Then predict a Boolean, whether this user will order or not this product in his/her future order.
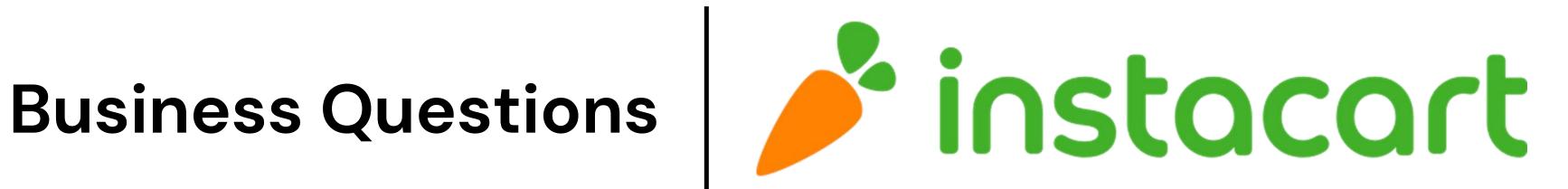
Feature Engineering | instacart

We wanted to extract features that strongly describe the relation between user and product

Create Predictor Variables

We at the moment are prepared to perceive and calculate predictor variables primarily based totally at the provided data.

* User predictors describing the conduct of a consumer e.g. overall variety of orders of a consumer.
* Product predictors describing traits of a product e.g. overall variety of instances a product has been purchased.
* User & product predictors describing the conduct of a consumer toward a particular product e.g. overall instances a consumer ordered a particular product.
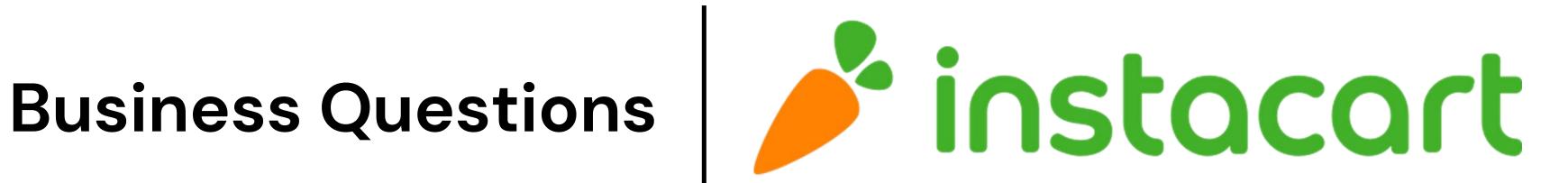
Business Questions | instacart

User predictors

- u_total_orders: Total number of orders made by a specific user.

- u_reordered_ratio: The ratio of how many times a user has reordered products compared to their total number of orders.

- user_distinct_products: Number of distinct products purchased by a specific user.

- user_total_products: Total number of products purchased by a specific user.

Business Questions | instacart

User predictors

- **days_since_prior_order: Number of days since the last order for a specific user.**

- **order_hour_of_day: The hour of the day when an order was placed.**

- **order_dow: The day of the week when an order was placed (0 = Sunday, 1 = Monday, ..., 6 = Saturday).**

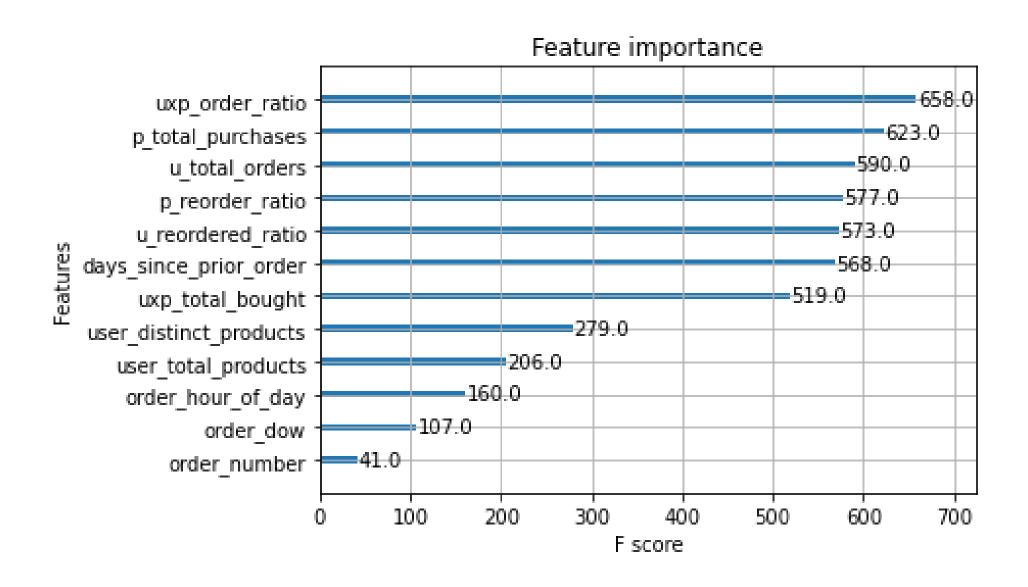- **order_number: The sequence number of the order for a specific user.**

Business Questions | instacart

product predictors

- `p_total_purchases`: Total number of purchases made for a specific product across all users.

- `p_reorder_ratio`: The ratio of how many times a product has been reordered compared to the total number of purchases for that product.

- `uxp_total_bought`: Total number of times a user has bought a specific product.

- `uxp_order_ratio`: The ratio of how many times a product has been ordered by a specific user compared to the total number of orders made by that user.
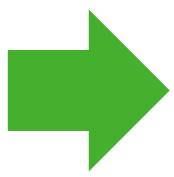
Business Questions

The features listed at the top have the highest importance scores, indicating they are most influential in predicting the target variable. In this case, features like uxp_order_ratio, p_total_purchases, and u_total_orders have the highest importance scores.

Feature importance

Feature	F score
uxp_order_ratio	658.0
p_total_purchases	623.0
u_total_orders	590.0
p_reorder_ratio	577.0
u_reordered_ratio	573.0
days_since_prior_order	568.0
uxp_total_bought	519.0
user_distinct_products	279.0
user_total_products	206.0
order_hour_of_day	160.0
order_dow	107.0
order_number	41.0

Deployment | instacart

Power Bi Dashboard ➡ **Insights deployment**

Deployment

Deployment

Thank you