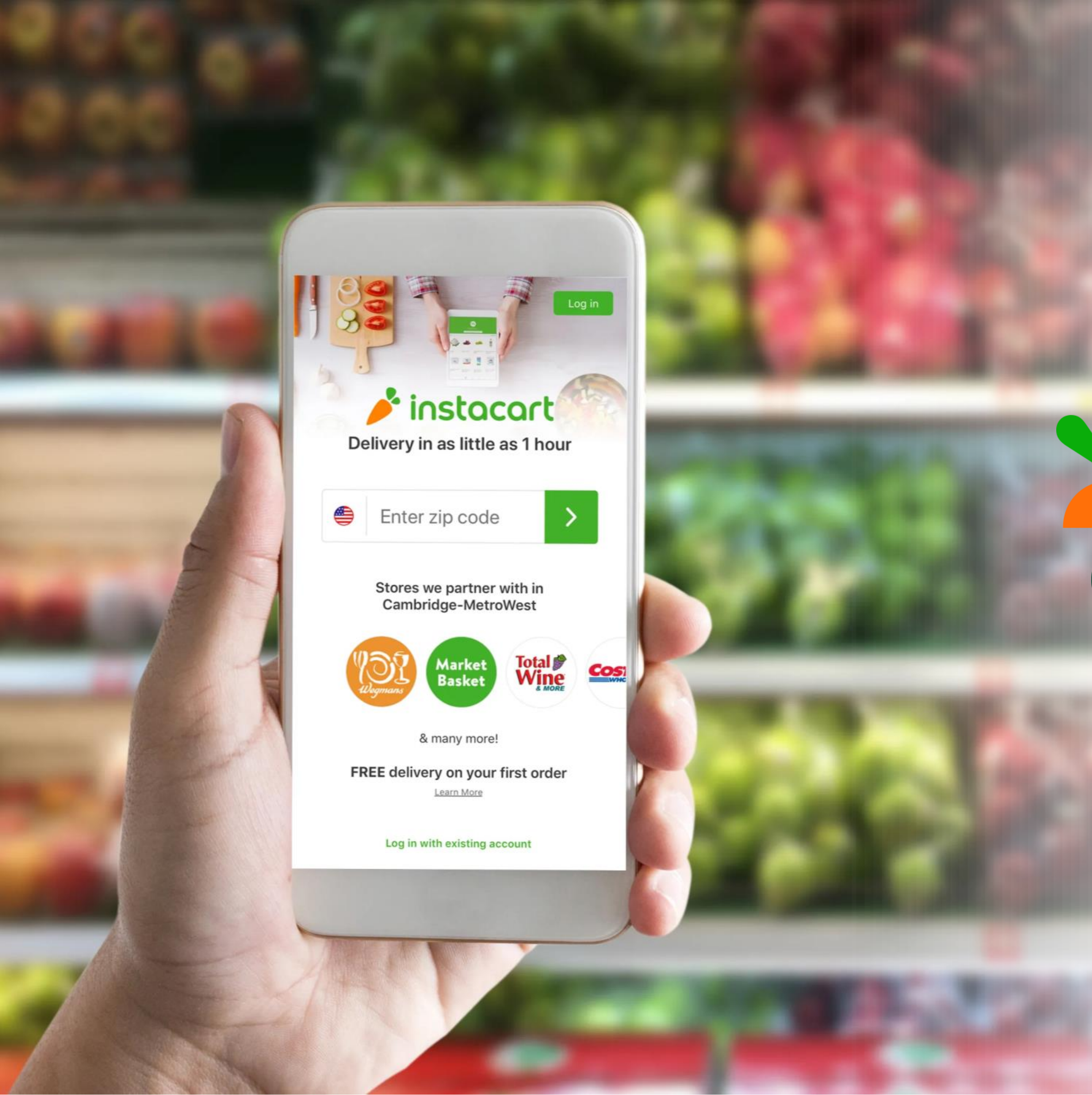



instacart

# Customer Segmentation & Market Basket Analysis

Widi Destrianda

# Data

**“The Instacart Online Grocery Shopping Dataset 2017”**

Data originally comes from https://www.instacart.com/datasets/grocery-shopping-2017
My data: Instacart Market Basket Analysis | Kaggle

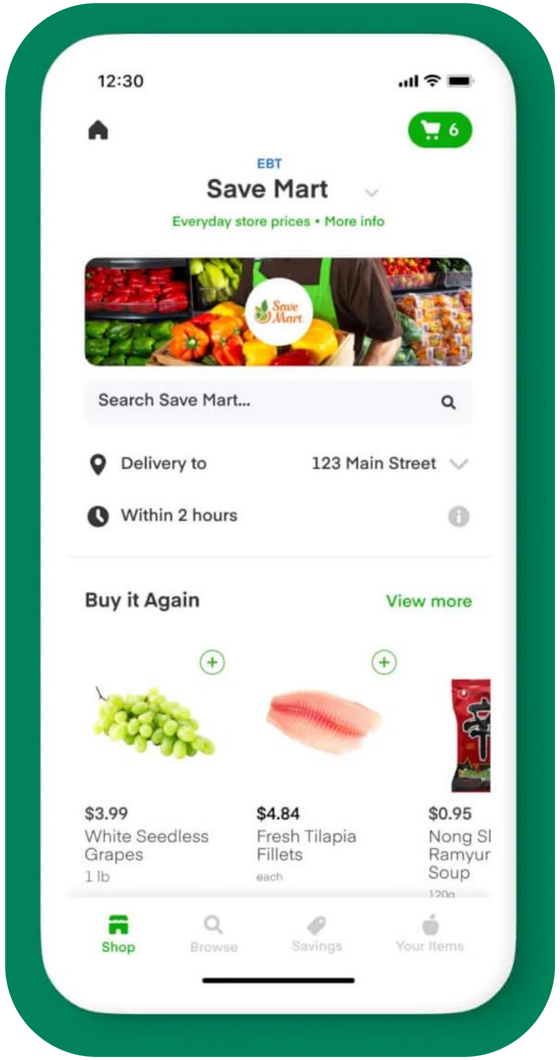
| Order.csv |
|---|
| Order_id |
| user_id |
| Order_number |
| Order_dow |
| Order_hour_of_day |
| Days_since_prior_order |

| Order_products.csv |
|---|
| Order_id |
| Product_id |
| Add_to_cart_order |
| reordered |

| Products.csv |
|---|
| Product_id |
| Product_name |
| Aisle_id |
| Department_id |

| Departments.csv |
|---|
| Department_id |
| Department |

| Aisle.csv |
|---|
| Aisle_id |
| Aisle |




(32.434.489 rows, 14 columns)
Memory usage: 1.6+ GB

instacart

# Background and Objectives


**Value Proposition** **simplifying grocery shopping**
**Objective** **increase profitability**

| | Business Model | Revenue Stream | Cost Structure |
|---|---|---|---|
|  **User** | Order using Web/App | **Subscription**<br>Monthly $9.99<br>Annual $99<br>**Pickup Fees**: $1.99 | Marketing |
|  **Web/App** | Assign User’s order to Shopper | **Advertising** Retailer Product | Web/App Maintenance |
|  **Shopper** | Shop the order from selected Store<br>**Full-Service** Shopper or **In-Store** Shopper | **Delivery Fee** $3.99 - $9.99<br>**Service Fee** 5 – 10% of Total | Salary/Shopper Payment |
|  **Store/Retailer** | Provide the item for the User | **Commission**<br>% of the price<br>**Mark-up Price** | Administration & Operations |

# Problem Statement

**Goal: Increase profitability by leveraging large customer base transactions behavior and purchasing history**
**Research Question: How to find lever from customer transactions behavior and purchasing history to improve profitability?**

Customer Transactions

Purchase Order History

## Customer Segmentation

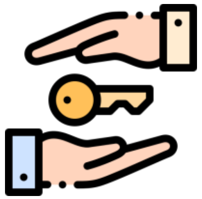
**Based on Spending Profile** — **RFM – K-means**

**Based on Buying Pattern** — **TSNE or PCA – K-means[1]**

**Association Rule – Apriori/FP-Growth/ECLAT**

## Market Basket Analysis

Tailored marketing approach

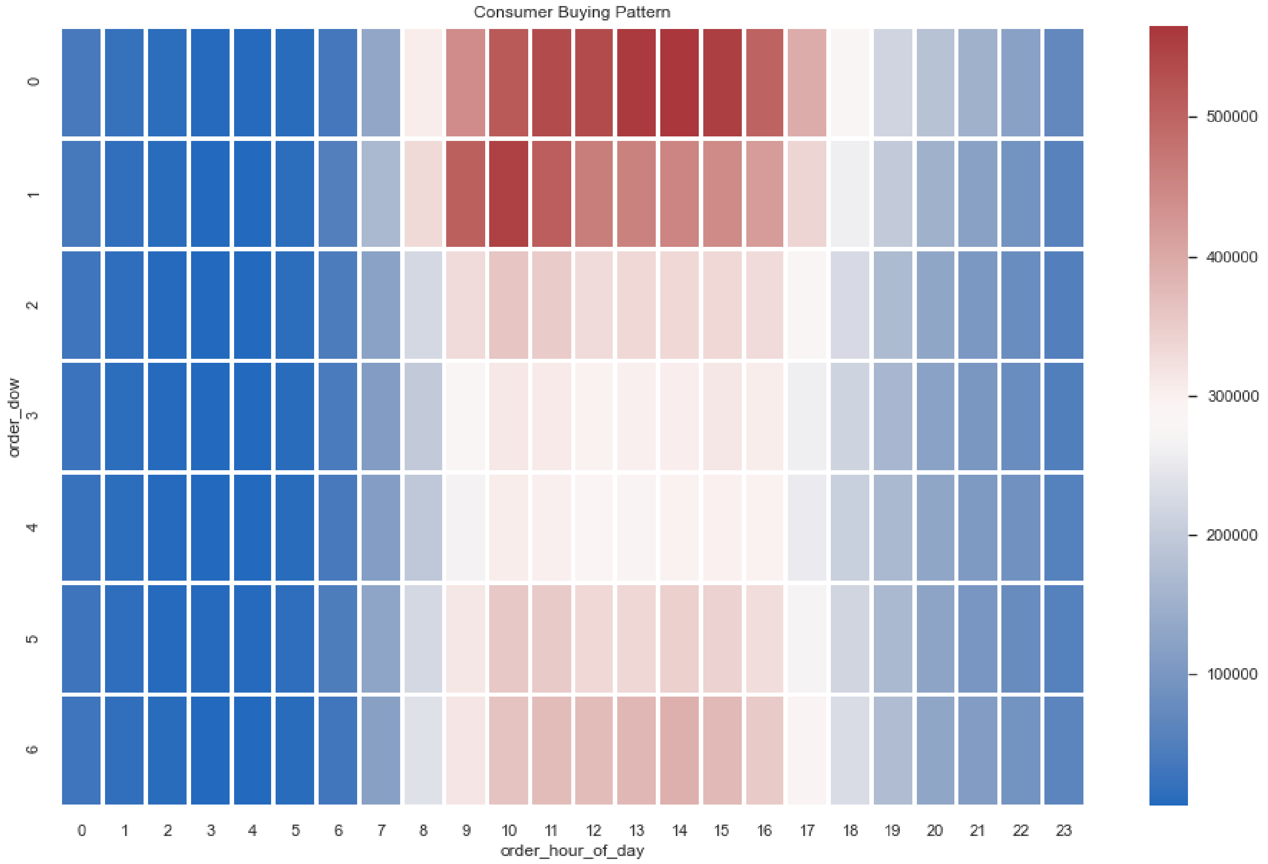


Determine key driver to increase profit

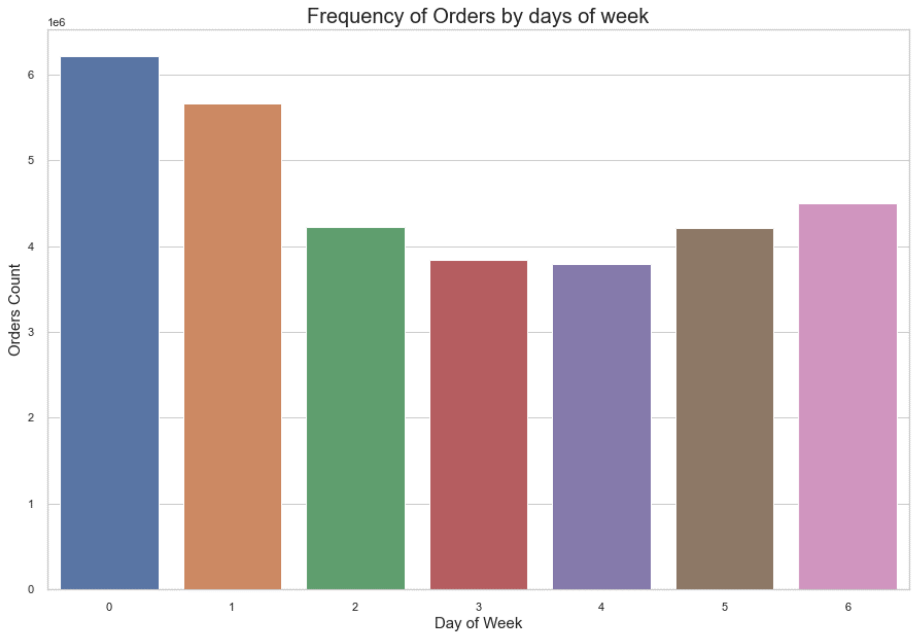
[1]Analysis of Accuracy K-Means and Apriori Algorithms for Patient Data Clusters
N P Dharshinni *et al* 2019 *J. Phys.: Conf. Ser.* **1230** 012020

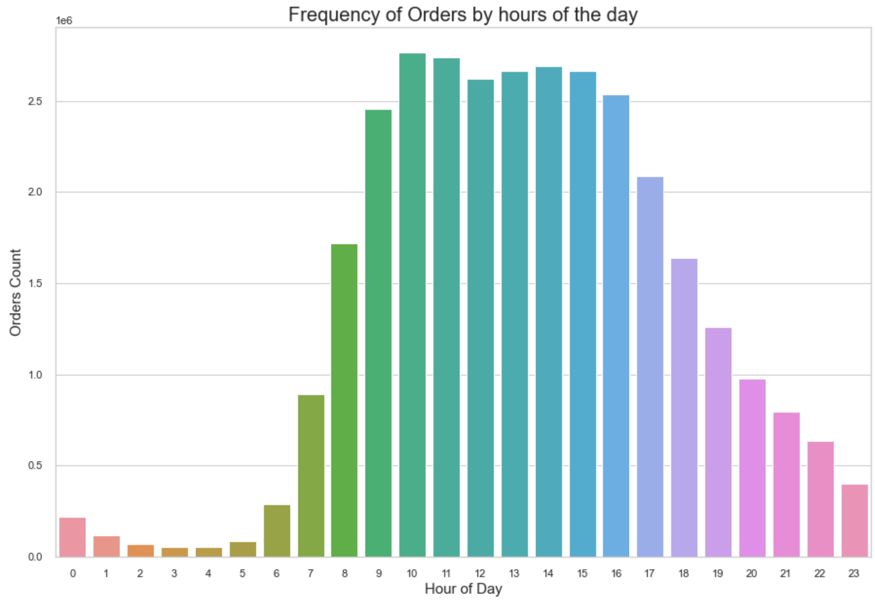
# Exploratory Data Analysis

**Total Order : 3.214.874**
**Total Product : 49.677**
**Total User : 206.209**

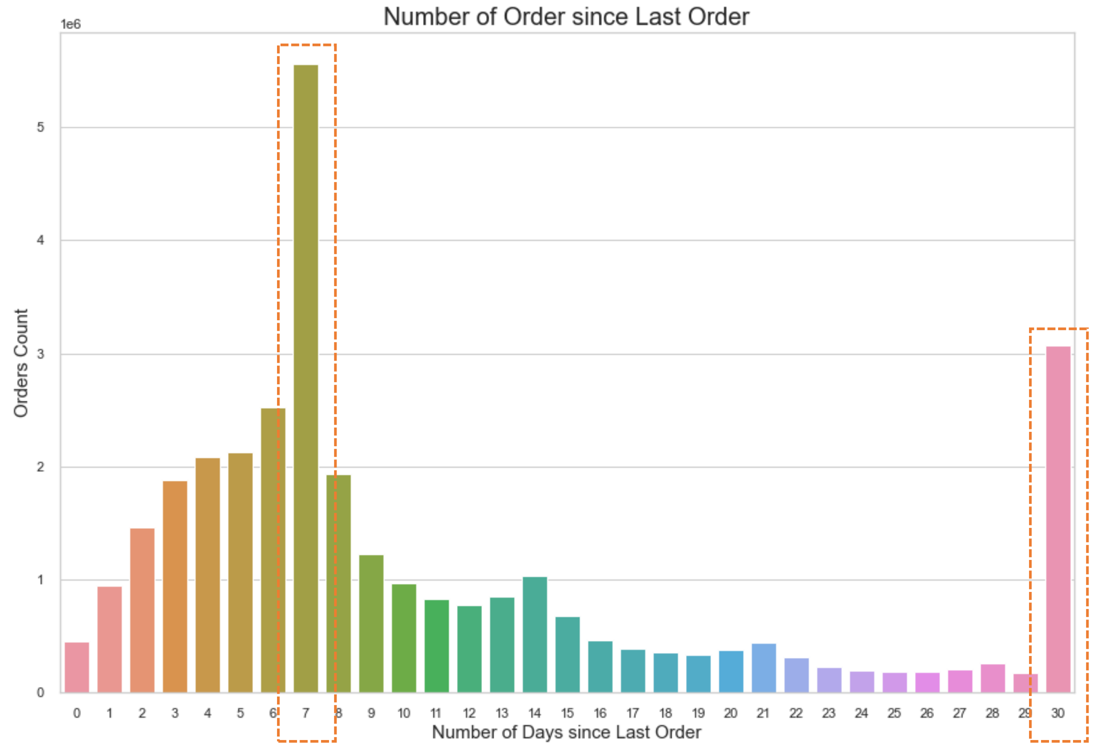



Most orders happens in the **weekend** and the **middle of the day** (9 AM – 16 PM)
The busiest hour is between **10 AM – 15 PM on Saturday**, and **9 AM – 11 AM on Sunday**

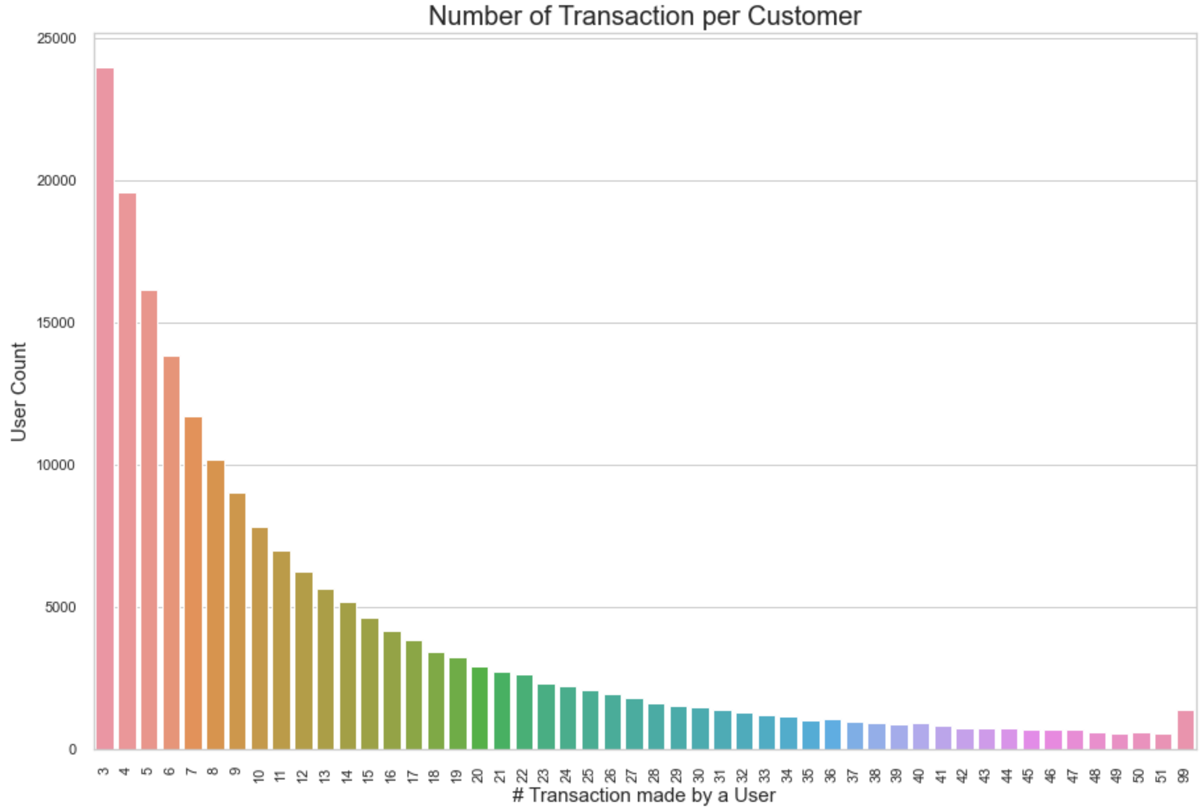


# Exploratory Data Analysis


Most reorder happens after 7 days and 30 days after prior purchase.
It means most of Instacart customers are doing weekly and monthly shopping.




Most Instacart customers only doing transactions between 3 – 5 times.
Its indicated Instacart has a problem in retaining new customer.

# Methodology

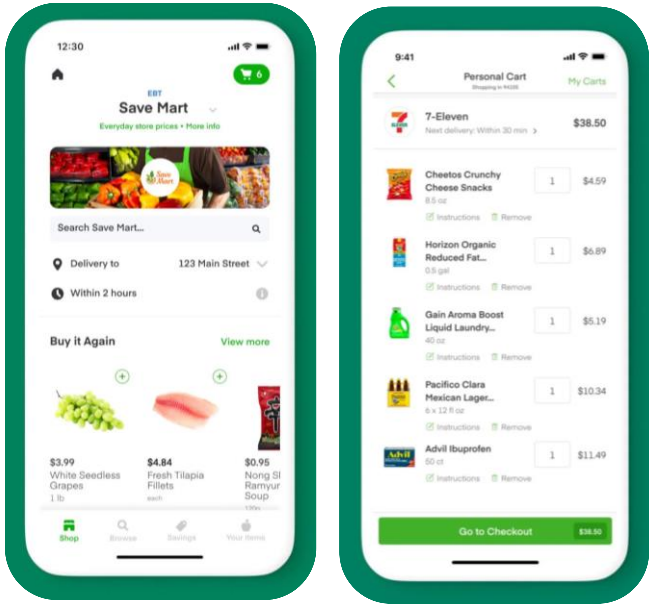
## Customer Segmentation

**Based on Spending Profile**

Feature Engineering Create RFM Metrics → K-Means Clustering → Choose Optimum K-Value → Analyze the Resulting Cluster using Visualization

**Based on Buying Pattern**

Feature Engineering User_id vs Aisle → Dimensionality Reduction TSNE/PCA → K-Means Clustering → Choose Optimum K-Value → Analyze the Resulting Cluster



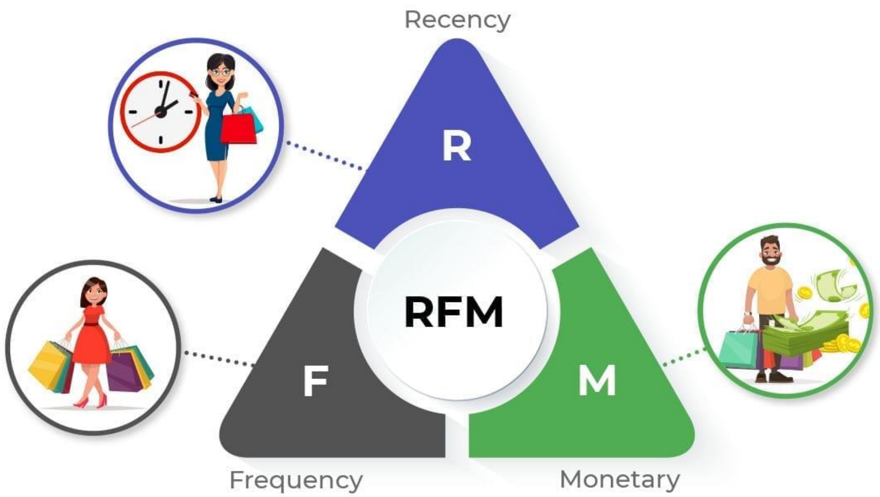

instacart

# Modelling

## Customer Segmentation – Based on Spending Profile

In order **to achieve better customer retention and profitability**, the company needs to **customize marketing strategies** and fulfill different customers' needs by allocating resources effectively and efficiently (Huang et al., 2009; Chang et al., 2010)[2].

RFM (recency, frequency and monetary) model is a behavior-based model used to **analyze the behavior of a customer** by measuring when people buy, how often they buy and how much they buy. (Hughes, 1996; Yeh et al., 2009).




**Why using RFM?**

1. RFM is **cost-effective** in acquiring important customer behavior analysis and is **easy to quantify customer behavior**.
2. It is **very effective** as the purchase behavior can be summarized by **using a small number of variables**.
3. RFM variables are gathered via an **internal database** regarding the transaction history and are not obtained through the aggregate level information in the demographic databases. Hence, RFM is more **meaningful for targeting particular customers** (Kaymak, 2001)[3].

## Feature Engineering

```
recency = df_final.groupby("user_id").aggregate(recency=("days_since_prior_order","mean")).reset_index()
recency
```

```
freq = df_final.groupby("user_id").aggregate(frequency=("order_number","max")).reset_index()
freq
```

```
monetary = df_final.groupby("user_id").aggregate(monetary = ('add_to_cart_order','mean')).reset_index()
monetary
```

| user_id | recency | frequency | monetary |
|---|---|---|---|
| 1 | 19.135593 | 10 | 3.627119 |
| 2 | 15.369231 | 14 | 8.553846 |
| 3 | 10.977273 | 12 | 4.443182 |
| 4 | 13.500000 | 5 | 2.777778 |
| 5 | 12.270270 | 4 | 5.513514 |

[2]Huang SC, Chang EC, Wu HH (2009). A case study of applying data mining techniques in an outfitter's customer value analysis. Expert Syst. Appl., 36: 5909-5915.

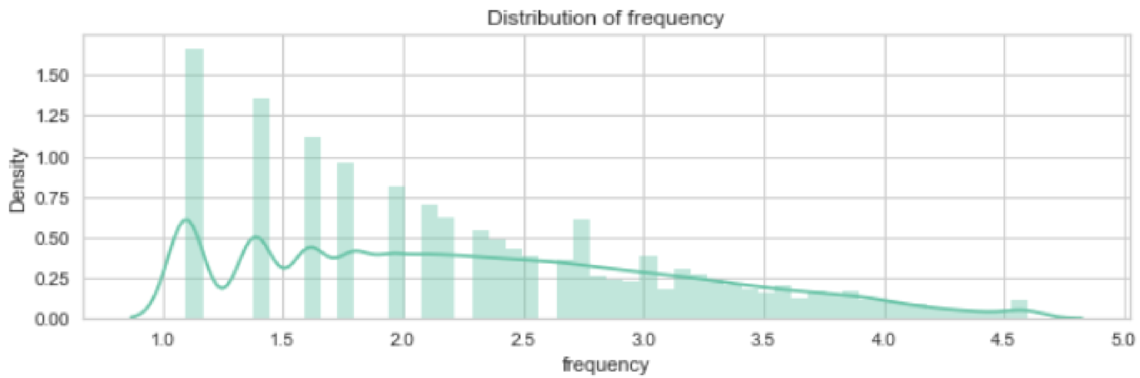
[2]Chang EC, Huang SC, Wu HH (2010). Using K-means method and spectral clustering technique in an outfitter's value analysis. Qual Quant., 44(4): 807-815.

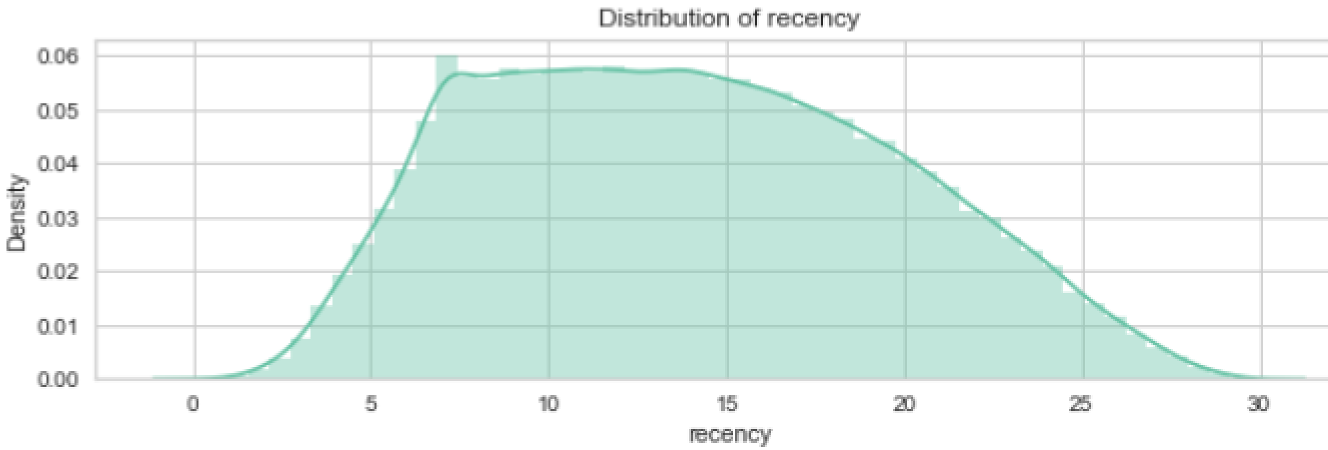
[3]Kaymak U (2001). Fuzzy target seleciton using RFM variables. IFSA World Congress 20th NAFIPS Int. Conf., 2: 1038-1043.

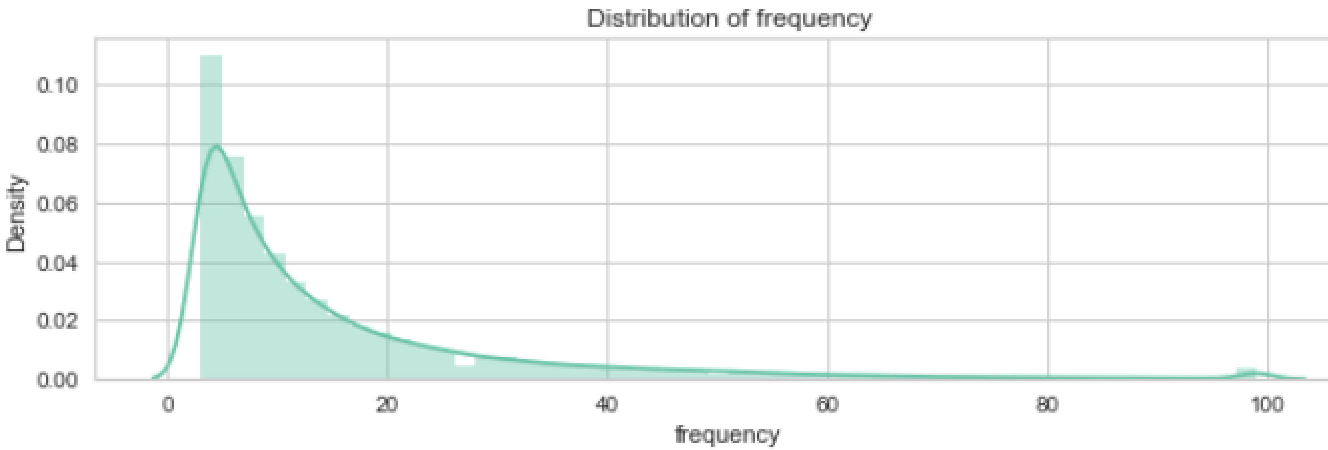
# Data Pre-processing

**Customer Segmentation – Based on Spending Profile**

Log Transformed



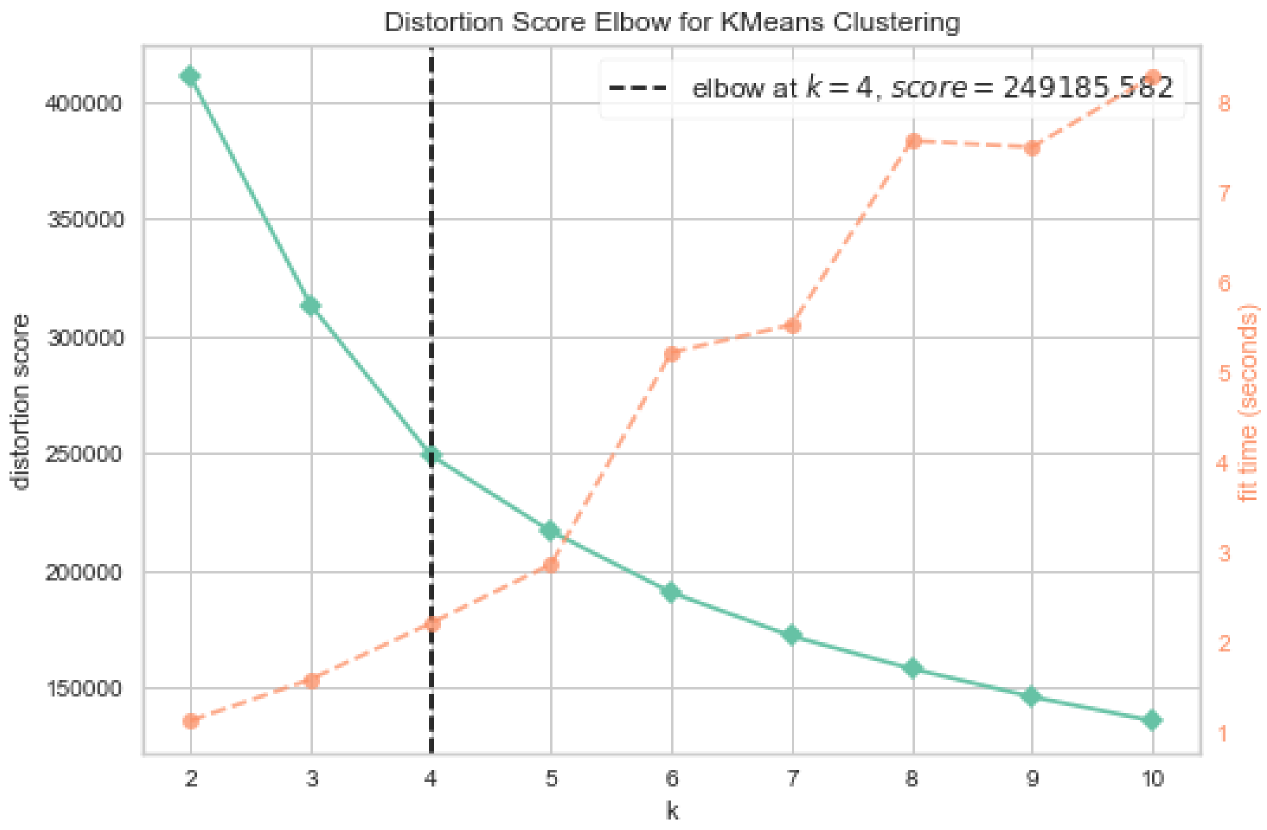




# Model Evaluation

**Customer Segmentation – Based on Spending Profile**

## Elbow Method

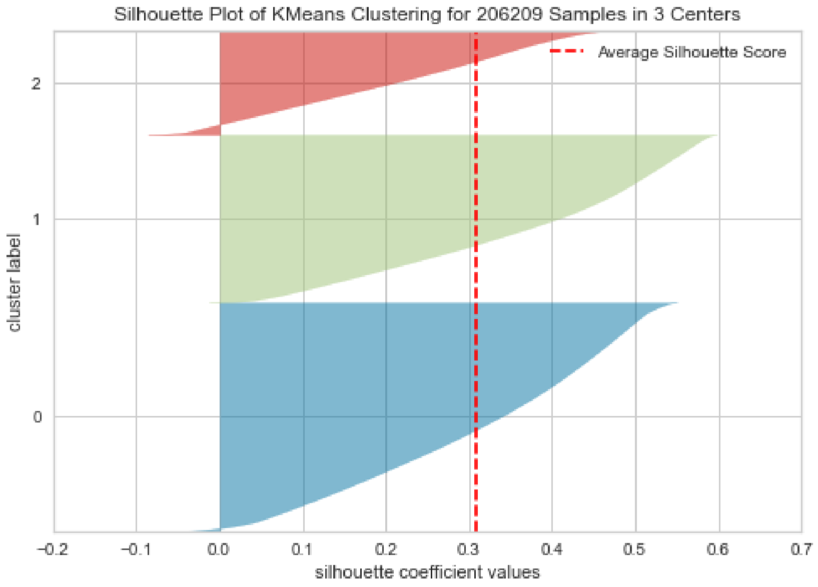
**Baseline**



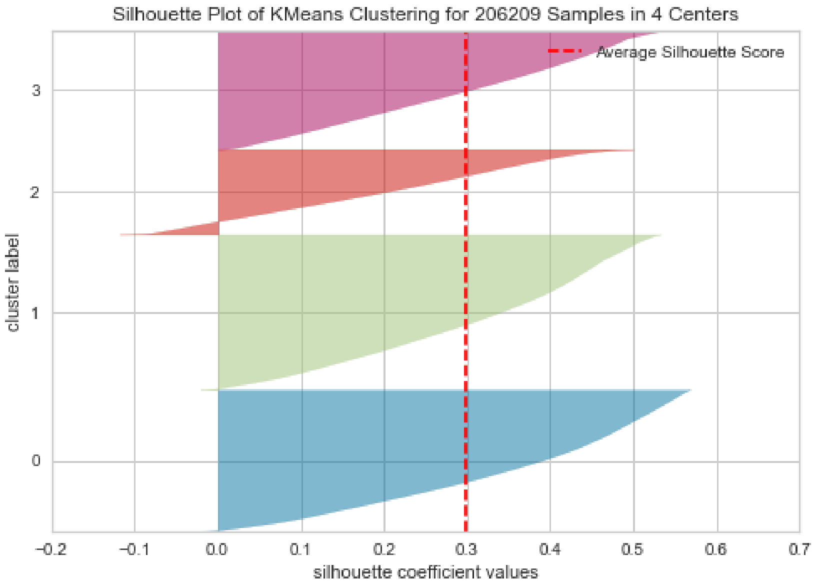

# Model Evaluation

## Customer Segmentation – Based on Spending Profile

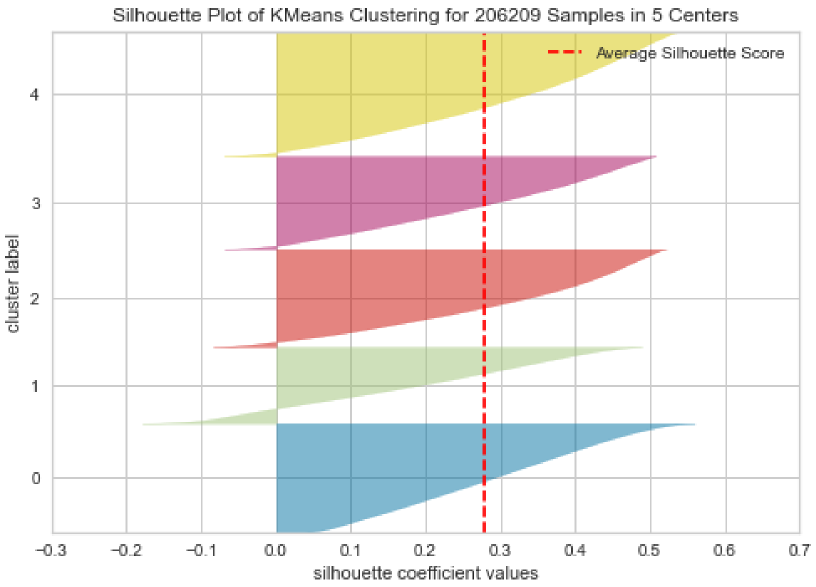
### Silhouette Score




K = 3



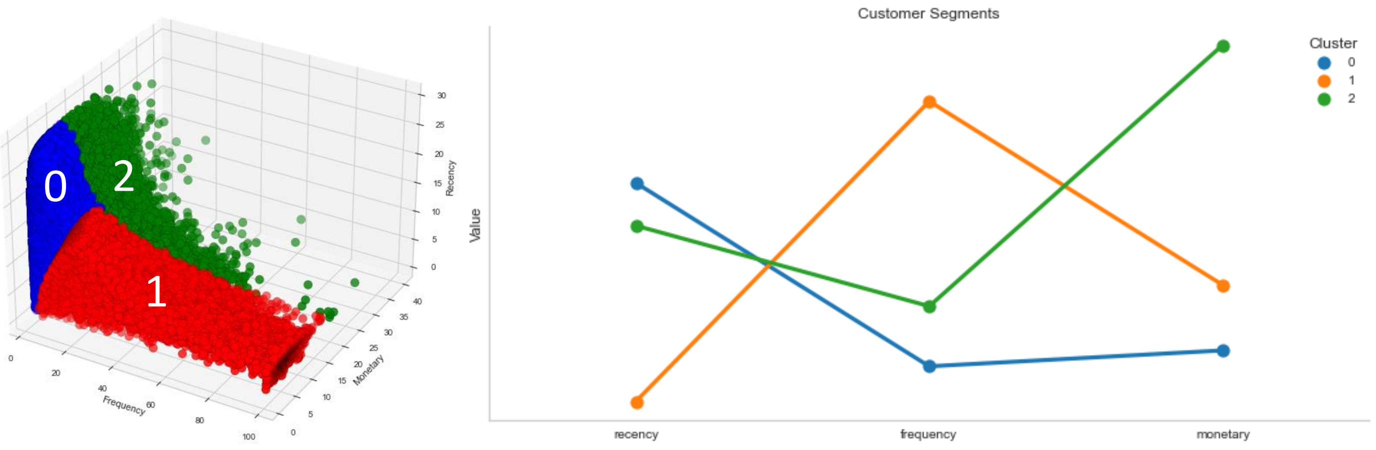

K = 4



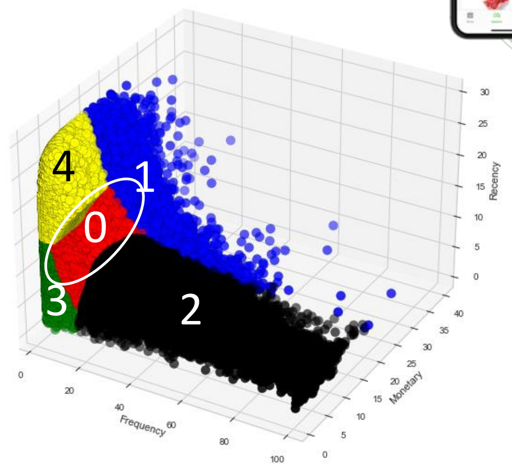

K = 5

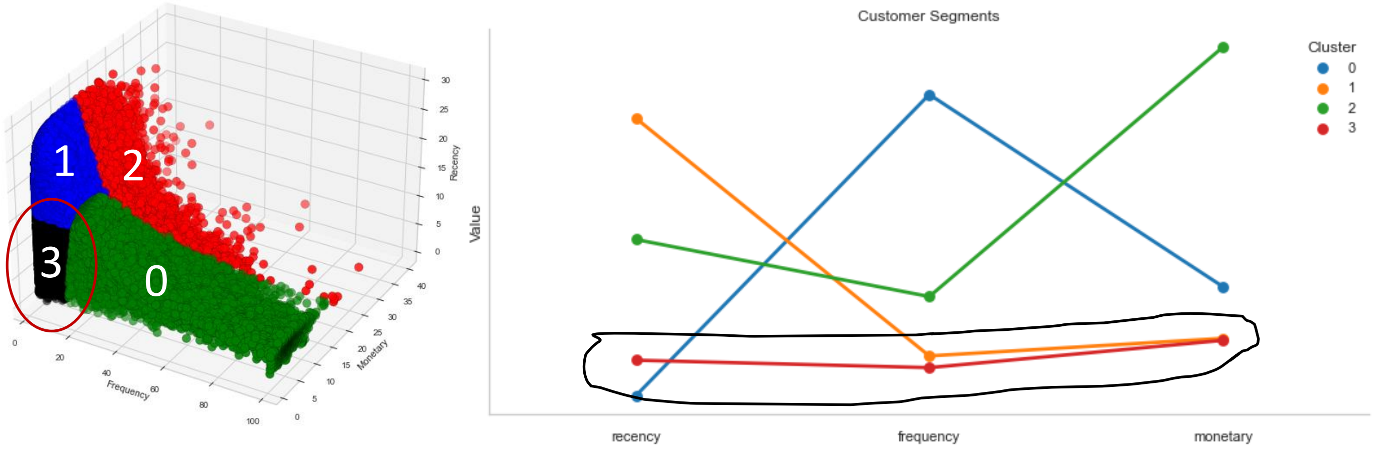
# Model Evaluation

## Customer Segmentation – Based on Spending Profile

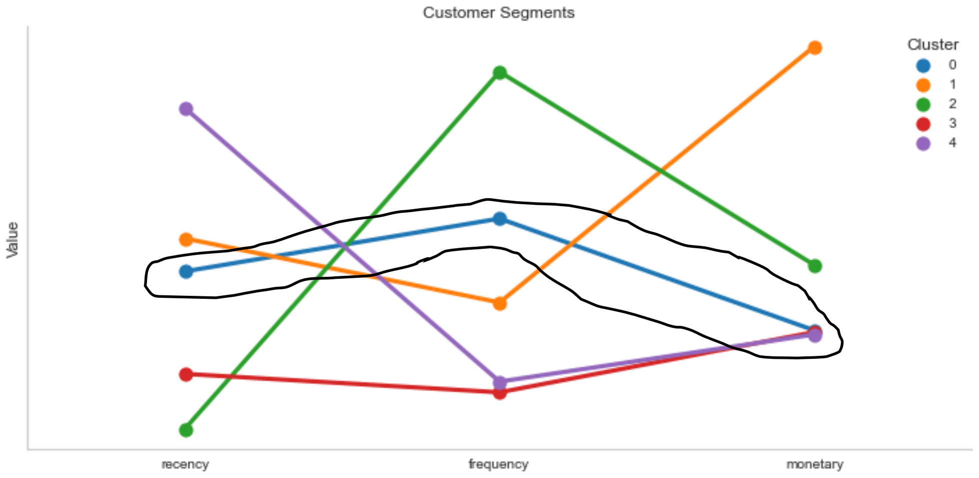




**K = 3**



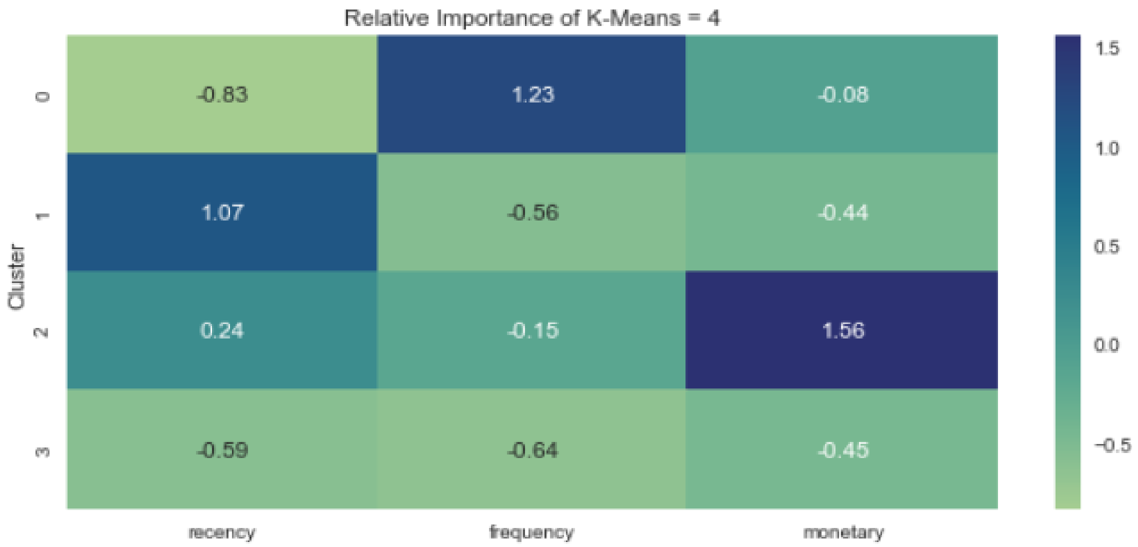

**K = 4**

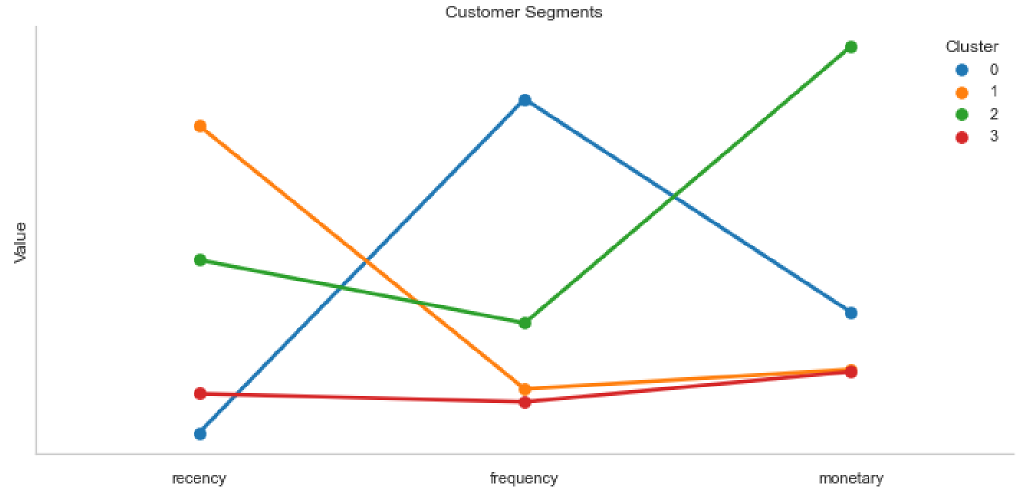



**K = 5**

# Analysis

## Customer Segmentation – Based on Spending Profile




Cluster 0: High Recency, High Frequency, Medium Monetary (**Loyal Customer**) 30.90%
Cluster 1: Low Recency, Low Frequency, Low Monetary (**Hibernating**) 17.05%
Cluster 2: Medium Recency, Medium Frequency, High Monetary (**Big Spender**) 28.24%
Cluster 3: High Recency, Low Frequency, Low Monetary (**Promising**) 23.82%

# Data Pre-processing

**Customer Segmentation – Based on Customer Buying Pattern**

| user_id | air fresheners candles | asian foods | baby accessories | baby bath body care | baby food formula | bakery desserts | baking ingredients | baking supplies decor | beauty | beers coolers | ... | spreads | tea | tofu meat alternatives | tortillas flat bread | trail mix snack mix | trash bags liners | vitamins supplements | water seltzer sparkling water | white wines | yogurt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **2** | 0 | 3 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | ... | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 42 |
| **3** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| **4** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| **5** | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **206205** | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| **206206** | 0 | 4 | 0 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | ... | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| **206207** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | ... | 3 | 4 | 0 | 2 | 1 | 0 | 0 | 11 | 0 | 15 |
| **206208** | 0 | 3 | 0 | 0 | 3 | 0 | 4 | 0 | 0 | 0 | ... | 5 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 33 |
| **206209** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 3 |

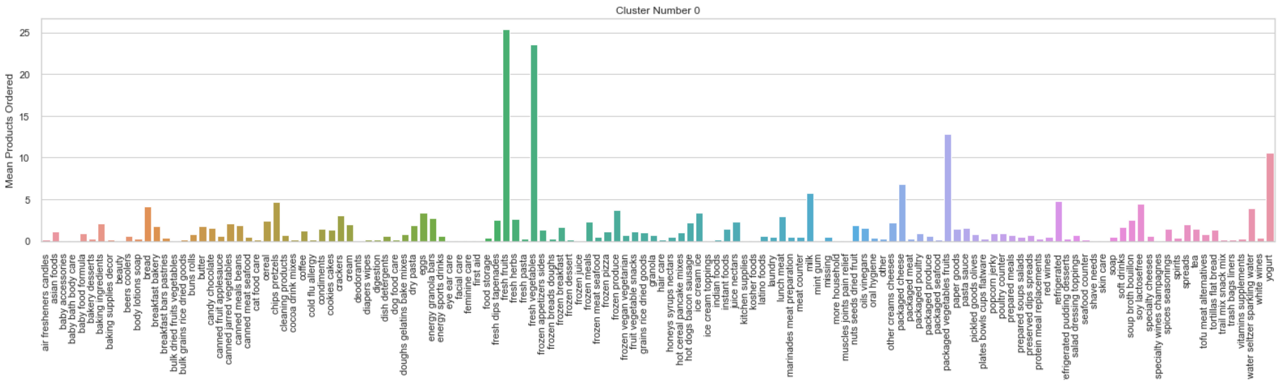
206209 rows × 134 columns

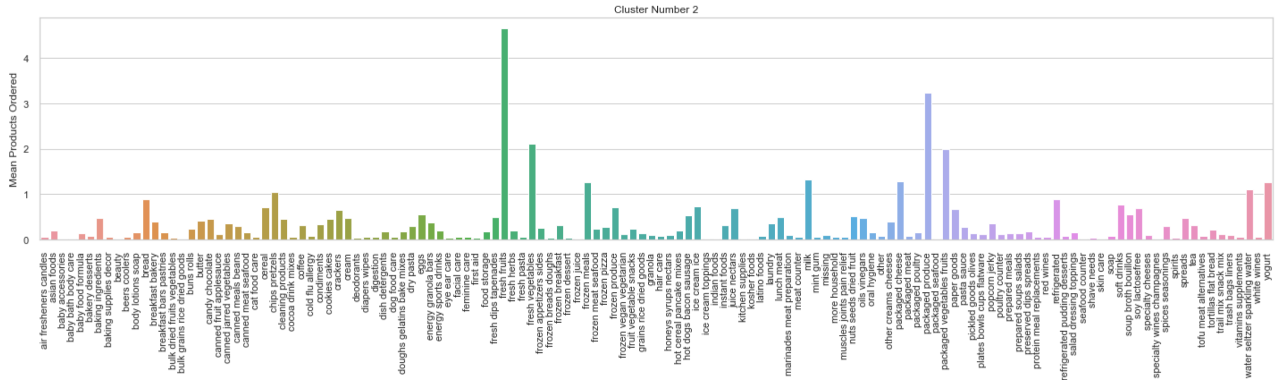
Because the dataset contain too many products (49.677) and too little departments (21), **aisle (134)** is used in the analysis to segment customer buying pattern.

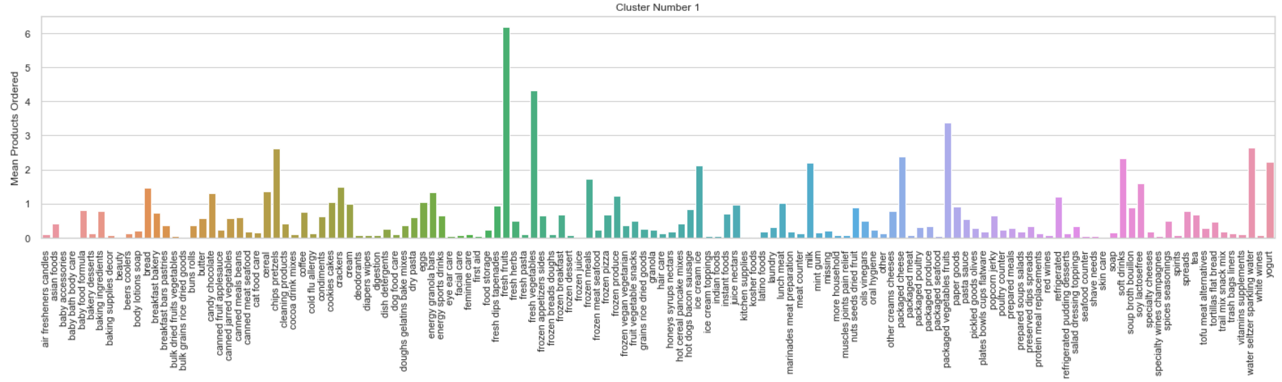
# Model Evaluation

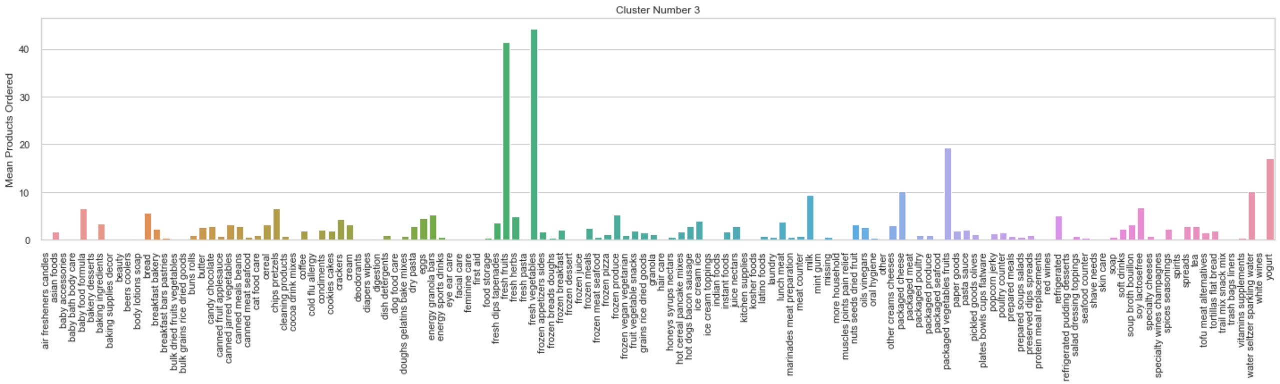
## Customer Segmentation – Based on Customer Buying Pattern


### T-SNE




# Model Evaluation

## Customer Segmentation – Based on Customer Buying Pattern


### T-SNE

### value_counts()/unique

**Cluster 0** — 23.97%

| | |
|---|---|
| fresh fruits | 25.36 |
| fresh vegetables | 23.51 |
| packaged vegetables fruits | 12.80 |
| yogurt | 10.58 |
| packaged cheese | 6.78 |
| milk | 5.74 |
| refrigerated | 4.82 |
| chips pretzels | 4.66 |
| soy lactosefree | 4.44 |
| bread | 4.16 |
| water seltzer sparkling water | 3.93 |
| frozen produce | 3.75 |
| ice cream ice | 3.39 |
| eggs | 3.37 |
| crackers | 3.05 |
| lunch meat | 2.91 |
| energy granola bars | 2.72 |
| fresh herbs | 2.58 |
| soup broth bouillon | 2.49 |
| fresh dips tapenades | 2.47 |

Name: aisle, dtype: float64

**Cluster 1** — 27.46%

| | |
|---|---|
| fresh fruits | 6.20 |
| fresh vegetables | 4.31 |
| packaged vegetables fruits | 3.37 |
| water seltzer sparkling water | 2.64 |
| chips pretzels | 2.61 |
| packaged cheese | 2.38 |
| soft drinks | 2.33 |
| yogurt | 2.22 |
| milk | 2.21 |
| ice cream ice | 2.12 |
| frozen meals | 1.74 |
| soy lactosefree | 1.59 |
| crackers | 1.50 |
| bread | 1.45 |
| cereal | 1.35 |
| energy granola bars | 1.33 |
| candy chocolate | 1.31 |
| frozen produce | 1.24 |
| refrigerated | 1.20 |
| eggs | 1.04 |

Name: aisle, dtype: float64

**Cluster 2** — 27.82%

| | |
|---|---|
| fresh fruits | 4.66 |
| packaged produce | 3.23 |
| fresh vegetables | 2.11 |
| packaged vegetables fruits | 2.00 |
| milk | 1.32 |
| packaged cheese | 1.28 |
| yogurt | 1.27 |
| frozen meals | 1.26 |
| water seltzer sparkling water | 1.11 |
| chips pretzels | 1.04 |
| refrigerated | 0.90 |
| bread | 0.88 |
| soft drinks | 0.76 |
| ice cream ice | 0.73 |
| cereal | 0.72 |
| frozen produce | 0.70 |
| juice nectars | 0.69 |
| soy lactosefree | 0.68 |
| paper goods | 0.67 |
| crackers | 0.66 |

Name: aisle, dtype: float64

**Cluster 3** — 20.74%

| | |
|---|---|
| fresh vegetables | 44.19 |
| fresh fruits | 41.39 |
| packaged vegetables fruits | 19.32 |
| yogurt | 17.08 |
| packaged cheese | 10.19 |
| water seltzer sparkling water | 10.15 |
| milk | 9.50 |
| soy lactosefree | 6.76 |
| baby food formula | 6.68 |
| chips pretzels | 6.65 |
| bread | 5.76 |
| frozen produce | 5.30 |
| energy granola bars | 5.26 |
| refrigerated | 5.10 |
| fresh herbs | 4.93 |
| eggs | 4.56 |
| crackers | 4.34 |
| ice cream ice | 3.94 |
| lunch meat | 3.87 |
| fresh dips tapenades | 3.57 |

Name: aisle, dtype: float64

# Methodology

Separate dataset according to **Buying Pattern Cluster**

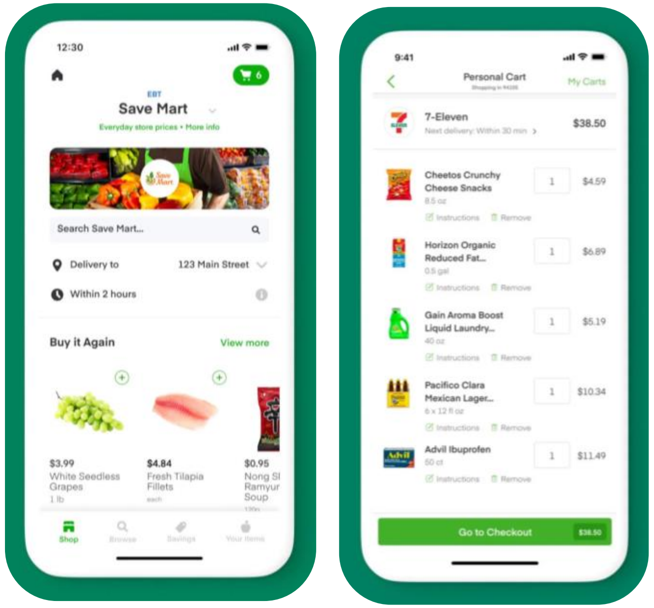
Feature Engineering
Order_id vs Product_id

Generate frequent itemset
Min support = 0.002

Association Rules Algorithm
Min lift = 1.5

Filter the rules
Confidence > Avg Confidence
Lift > Avg Lift

Combined All Rules from each Cluster

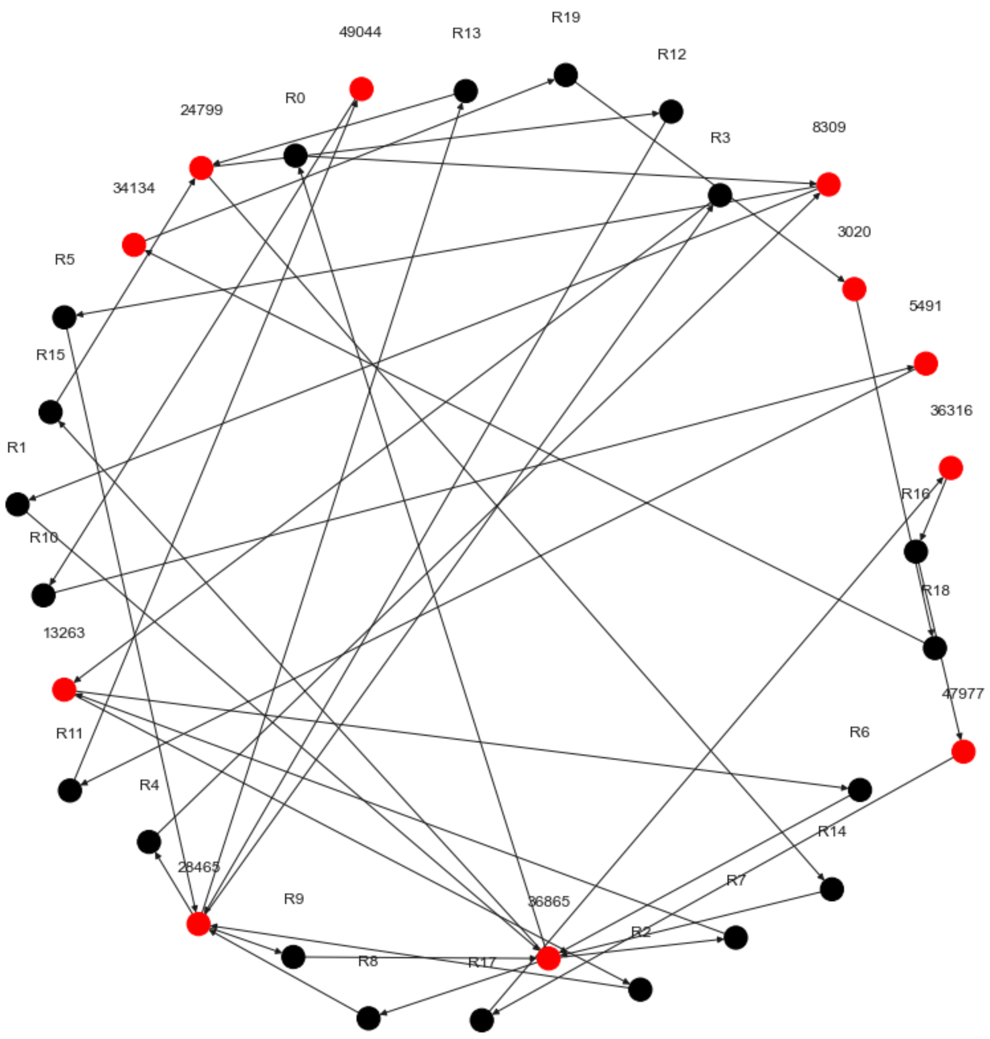



instacart

# Model Evaluation

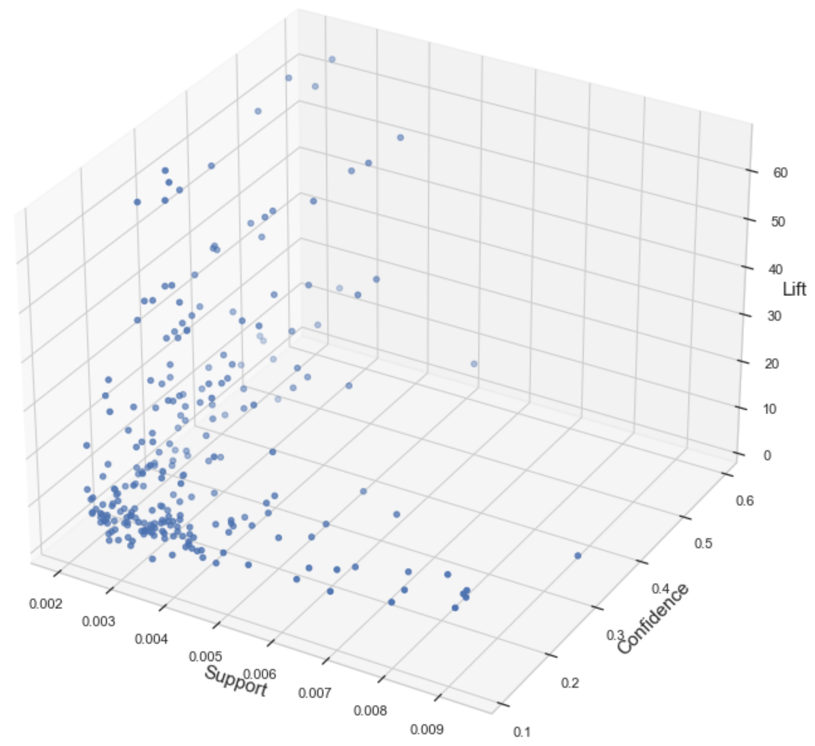
## Market Basket Analysis – Association Rules

Network Graph for Association Rules




Combined Cluster

Association Rules Distribution




| | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|---|---|---|---|---|---|---|---|
| **count** | 251.000000 | 251.000000 | 251.000000 | 251.000000 | 251.000000 | 251.000000 | 251.000000 |
| **mean** | 0.015933 | 0.029694 | 0.003195 | 0.234801 | 15.027893 | 0.002678 | 1.297013 |
| **std** | 0.008769 | 0.021169 | 0.001336 | 0.101569 | 15.680076 | 0.001054 | 0.225980 |
| **min** | 0.004132 | 0.004132 | 0.002006 | 0.115061 | 2.462832 | 0.001350 | 1.079016 |
| **25%** | 0.009170 | 0.011716 | 0.002326 | 0.151126 | 4.440272 | 0.002004 | 1.140980 |
| **50%** | 0.014313 | 0.025914 | 0.002774 | 0.206522 | 6.510077 | 0.002358 | 1.205295 |
| **75%** | 0.020856 | 0.040395 | 0.003490 | 0.282595 | 20.932092 | 0.003036 | 1.362823 |
| **max** | 0.047455 | 0.134664 | 0.009387 | 0.592233 | 64.910215 | 0.007268 | 2.429755 |

# Model Evaluation

## Market Basket Analysis – Association Rules

**Final Rules**

**251 rules**

| | antecedents_id | antecedents_name | ant_aisle_id | ant_aisle | ant_dept_id | ant_dept | consequents_id | consequents_name | con_aisle_id | con_aisle | con_dept_id | con_dept | confidence | lift |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 49044 | Organic Pears, Peas and Broccoli Puree Stage 1 | 92 | baby food formula | 18 | babies | 5491 | Stage 1 Apples Sweet Potatoes Pumpkin & Bluebe... | 92 | baby food formula | 18 | babies | 0.380435 | 57.376482 |
| 1 | 5491 | Stage 1 Apples Sweet Potatoes Pumpkin & Bluebe... | 92 | baby food formula | 18 | babies | 49044 | Organic Pears, Peas and Broccoli Puree Stage 1 | 92 | baby food formula | 18 | babies | 0.318182 | 57.376482 |
| 2 | 3020 | Broccoli & Apple Stage 2 Baby Food | 92 | baby food formula | 18 | babies | 34134 | Spinach Peas & Pear Stage 2 Baby Food | 92 | baby food formula | 18 | babies | 0.361905 | 42.281690 |
| 3 | 43875 | Baby Food Stage 2 Blueberry Pear & Purple Carrot | 92 | baby food formula | 18 | babies | 34134 | Spinach Peas & Pear Stage 2 Baby Food | 92 | baby food formula | 18 | babies | 0.271523 | 31.722321 |
| 4 | 34134 | Spinach Peas & Pear Stage 2 Baby Food | 92 | baby food formula | 18 | babies | 3020 | Broccoli & Apple Stage 2 Baby Food | 92 | baby food formula | 18 | babies | 0.267606 | 42.281690 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 93 | 39275 | Organic Blueberries | 123 | packaged vegetables fruits | 4 | produce | 21137 | Organic Strawberries | 24 | fresh fruits | 4 | produce | 0.227201 | 2.463706 |
| 94 | 22035 | Organic Whole String Cheese | 21 | packaged cheese | 16 | dairy eggs | 27845 | Organic Whole Milk | 84 | milk | 16 | dairy eggs | 0.152898 | 3.227555 |
| 96 | 30391 | Organic Cucumber | 83 | fresh vegetables | 4 | produce | 47209 | Organic Hass Avocado | 24 | fresh fruits | 4 | produce | 0.206924 | 2.595843 |
| 97 | 34243 | Organic Baby Broccoli | 83 | fresh vegetables | 4 | produce | 47209 | Organic Hass Avocado | 24 | fresh fruits | 4 | produce | 0.206831 | 2.594674 |
| 98 | 8174 | Organic Navel Orange | 24 | fresh fruits | 4 | produce | 13176 | Bag of Organic Bananas | 24 | fresh fruits | 4 | produce | 0.338583 | 2.514278 |

251 rows × 14 columns


# Business Implication

## Market Basket Analysis – Association Rules


### Top 25 Most Order /Reorder Product

```
'Banana',
'Bag of Organic Bananas',
'Organic Strawberries',
'Organic Baby Spinach',
'Organic Hass Avocado',
'Organic Avocado',
'Organic Whole Milk',
'Large Lemon',
'Organic Raspberries',
'Strawberries',
'Limes',
'Organic Yellow Onion',
'Organic Garlic',
'Organic Zucchini',
'Cucumber Kirby',
'Organic Fuji Apple',
'Organic Blueberries',
'Apple Honeycrisp Organic',
'Organic Lemon',
'Organic Half & Half',
'Sparkling Water Grapefruit',
'Honeycrisp Apple',
'Organic Large Extra Fancy Fuji Apple',
'Organic Cucumber',
'Organic Grape Tomatoes']
```

### Recommendation

| | antecedents_id | antecedents_name | ant_aisle_id | ant_aisle | ant_dept_id | ant_dept | consequents_id | consequents_name | con_aisle_id | con_aisle | con_dept_id | con_dept | confidence | lift |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 4605 | Yellow Onions | 83 | fresh vegetables | 4 | produce | 47626 | Large Lemon | 24 | fresh fruits | 4 | produce | 0.144725 | 2.462832 |
| **1** | 4605 | Yellow Onions | 83 | fresh vegetables | 4 | produce | 24964 | Organic Garlic | 83 | fresh vegetables | 4 | produce | 0.158537 | 4.584871 |
| **2** | 4920 | Seedless Red Grapes | 123 | packaged vegetables fruits | 4 | produce | 16797 | Strawberries | 24 | fresh fruits | 4 | produce | 0.116124 | 2.508991 |
| **3** | 5876 | Organic Lemon | 24 | fresh fruits | 4 | produce | 24964 | Organic Garlic | 83 | fresh vegetables | 4 | produce | 0.140473 | 4.062469 |
| **4** | 5876 | Organic Lemon | 24 | fresh fruits | 4 | produce | 22935 | Organic Yellow Onion | 83 | fresh vegetables | 4 | produce | 0.124277 | 4.482093 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **23** | 47626 | Large Lemon | 24 | fresh fruits | 4 | produce | 26209 | Limes | 24 | fresh fruits | 4 | produce | 0.200384 | 4.960636 |
| **24** | 47626 | Large Lemon | 24 | fresh fruits | 4 | produce | 47766 | Organic Avocado | 24 | fresh fruits | 4 | produce | 0.201342 | 4.278996 |
| **25** | 47766 | Organic Avocado | 24 | fresh fruits | 4 | produce | 26209 | Limes | 24 | fresh fruits | 4 | produce | 0.158568 | 3.925458 |
| **26** | 47766 | Organic Avocado | 24 | fresh fruits | 4 | produce | 47626 | Large Lemon | 24 | fresh fruits | 4 | produce | 0.179028 | 4.278996 |
| **27** | 49683 | Cucumber Kirby | 83 | fresh vegetables | 4 | produce | 47766 | Organic Avocado | 24 | fresh fruits | 4 | produce | 0.217949 | 4.631921 |

28 rows × 14 columns

```
antecedents_name
Cucumber Kirby                                            [Organic Avocado]
Large Lemon                                        [Limes, Organic Avocado]
Limes                     [Organic Garlic, Organic Cilantro, Organic Avo...
Organic Avocado                                       [Limes, Large Lemon]
Organic Blueberries           [Organic Strawberries, Organic Raspberries]
Organic Garlic              [Organic Yellow Onion, Organic Red Onion, Limes]
Organic Lemon             [Organic Garlic, Organic Yellow Onion, Organic...
Organic Yellow Onion           [Organic Small Bunch Celery, Organic Garlic]
Seedless Red Grapes                                          [Strawberries]
Sparkling Water Grapefruit  [Peach Pear Flavored Sparkling Water, Pure Spa...
Strawberries                                                  [Raspberries]
Yellow Onions                                 [Large Lemon, Organic Garlic]
Name: consequents_name, dtype: object
```

# Potential Impact

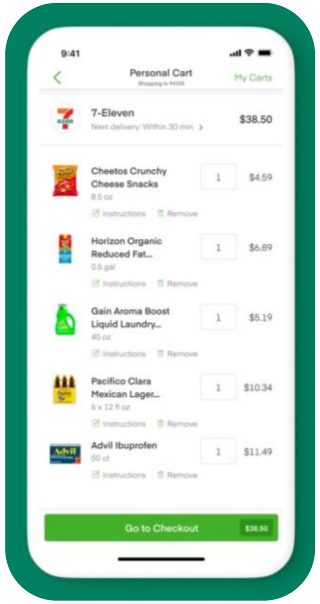
Amazon reportedly attributes as much as 35 percent of its sales to cross-selling through its “frequently bought together” options on every product page

It means, the sales **increase 54%** from normal sales.

### Top 25 Most Order /Reorder Product Recommendation

```
antecedents_name
Cucumber Kirby                                              [Organic Avocado]
Large Lemon                                          [Limes, Organic Avocado]
Limes                        [Organic Cilantro, Organic Avocado, Organic Ga...
Organic Avocado                                         [Large Lemon, Limes]
Organic Blueberries                [Organic Raspberries, Organic Strawberries]
Organic Garlic                [Organic Red Onion, Organic Yellow Onion, Limes]
Organic Lemon                [Organic Hass Avocado, Organic Cucumber, Organ...
Organic Yellow Onion           [Organic Garlic, Organic Small Bunch Celery]
Seedless Red Grapes                                           [Strawberries]
Sparkling Water Grapefruit   [Lime Sparkling Water, Peach Pear Flavored Spa...
Strawberries                                                  [Raspberries]
Yellow Onions                                   [Organic Garlic, Large Lemon]
Name: consequents_name, dtype: object
```

| Antecedent | Consequent | Forecasted Additional Sales Quantity (54%) | Price (USD) | Confidence | Expected Additional Revenue |
|---|---|---|---|---|---|
| Organic Blueberries | Organic Strawberries | 142.928 | 6.99 | 0.227201 | 226.990 |
| Limes | Large Lemon | 82.434 | 0.69 | 0.207547 | 11.805 |
| … | … | … | … | … | … |
| Cucumber Kirby | Organic Avocado | 95.480 | 3.49 | 0.217949 | 72.626 |




Total Revenue 2017 : **8.8086.587**

Expected Additional Revenue : **1.268.043**

**INCREASE 15.68%**

instacart

# Potential Impact

**Promotional Pricing** (discount to stimulate purchase for other products), leveraging association rules

Discount 20%
Sales Increase 20%

Lift : 6.51

Cross-sell
Sales increase 54%

$0.97 each (est.)
Organic Green Kiwi
$4.99 / lb

$2.89 /lb
Organic D'Anjou Pears

| | | Organic Green Kiwi | Organic D'Anjou Pears | Organic Green Kiwi | Organic D'Anjou Pears |
|---|---|---|---|---|---|
| Price | | 0.97 | 2.89 | 0.776 | 2.89 |
| #Order | | 50.141 | 48.915 | 60.169 | 90.394 |
| Markup Price | | 0% | 5% | 0% | 5% |
| Sales Revenue | | 48.636 | 51.360 | 46.691 | 86.811 |
| Commission | 10% | 4.863 | 5.136 | 4.669 | 8.681 |
| Service Fee | 5% | 2.431 | 2.568 | 2.334 | 4.340 |
| Revenue | | 7.295 | 7.704 | 7.003 | 13.021 |
| Total Revenue | | 14.999 | | 20.025 **(Increase 33%)** | |

# Business Implication

**Market Basket Analysis – Association Rules**


## Use Case

Most association rule that generated, antecedent and consequent have the same aisle. So, the placement already appropriate. Still, there are 83 out of 251 rules that have different aisle. So, layout arrangement is needed to **optimize shopping efficiency**.

| **Department** | Canned Goods | Dairy Eggs | Frozen | Meat Seafood | Produce |
|---|---|---|---|---|---|
| **Aisle** | Canned Meals Beans | | Frozen Produce | Packaged Poultry | Fresh Fruits |
| | | Eggs | | | Packaged Vegetables Fruit |
| | | Packaged Cheese | | | Fresh Herbs |
| | | Milk | | | Fresh Vegetables |
| | | Yogurt | | | Packaged Produce |

# Conclusion

- **Instacart** has a problem in retaining new customer.
- The busiest hour is between 10 AM – 15 PM on Saturday, and 9 AM – 11 AM on Sunday, lower activity on weekday. Most weekday activity is in the morning.

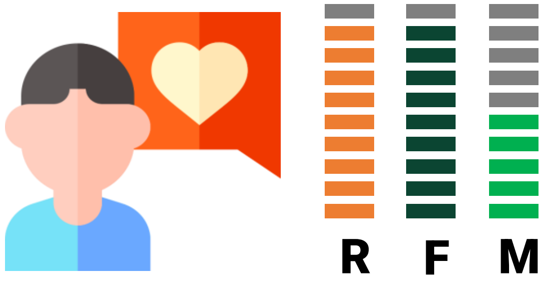
- There are **4 Clusters** of Customer based on Spending Profile
  Loyal Customer 30.90%
  Hibernating 17.05%
  Big Spender 28.24%
  Promising 23.82%

- There are 251 association rules generated from Customer's Cluster based on buying pattern.
- From these association rules, 28 items can be recommended to top 25 most order/reorder product.
- 83 out of 251 rules have different aisle. So, layout arrangement is needed to optimize shopping efficiency.

instacart

# Recommendation

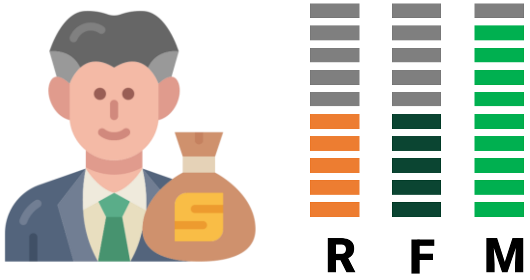
## Marketing Approach based on Priority




**Loyal Customer** 30.90%

Main objective: increase AOV (Average Order Value)

- Focus on **loyalty programs** and new product introductions. These customers have proven to enjoy Instacart services, so **don't use discount pricing** of services to generate incremental sales.
- **Reward** these customers. They can become early adopters for new products & services and will help promote Instacart brand.
- **Leverage Herd Mentality**: Its inborn people do what they see other people do. That’s the herd mentality concept smart marketers use to persuade shoppers to spend more: **product recommendations**, **frequently bought together**, best seller items to **cross-sell** the products.
- **Send birthday, anniversary, and milestone emails** (send gifts to customer in their birthday), anniversary of the first time they bought from company, or when they’ve purchased certain number of items. Birthday emails result in 481% more transactions than promotional emails.



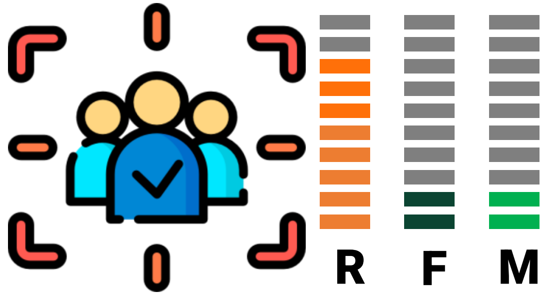

**Big Spender** 28.24%

Main objective: increase engagement

- These customers have demonstrated a high willingness to pay. **Don't waste margin on too many product discounts**.
- **Offer Subscription** to increase frequency and recency, and provide upgrade offers.
- **Offer membership or loyalty programs** or recommend related products to **cross-sell** them and help them become ideal Customer.
- **Cart abandonment email**
- **Post-purchase email** series that suggests similar items to the one originally purchased.

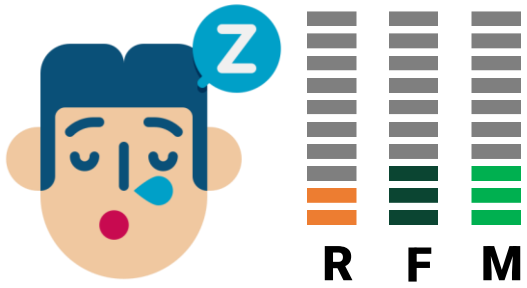
# Recommendation

## Marketing Approach based on Priority




**Promising** 23.82%

- Need to handle with care by **improving relationship** with them. Try to **enhance their purchasing experience** by providing good services and products, and customer care services.
- Provide Onboarding support, Free trials, special offers, and Brand awareness campaigns to **increase their visits.**
- **Increase Customer satisfaction**: resolve a complaint in your customers' favor (e.g., offering them a refund or accepting a return that's out of warranty).




**Hibernating** 17.05%

- For the sake of allocating efficient resources, **the effort** to bring back these customers **should be minimum**, because the hurdle to improve sales from this type is actually harder.
- **Discount may perform well**. Use a **'discount ladder',** a series of increasing discounts only available if the previous discount hasn't been redeemed, to incentivize purchase.
- Give **frequent** product updates, brand updates, and company newsletter email in **regular basis**
- Instacart should try to **understand why they left**, and make sure that it does not happen again, from Customer Survey.

# Recommendation


## Profit

### Revenue

#### Commissions Mark-up Price

**Promotional Pricing** (discount to stimulate purchase for other products), leveraging association rules

#### Delivery & Service Fee

**Price Discrimination** for **delivery fee** in busy hour to generate more profit

#### Subscription

Target Customer with transaction **frequency <= 2 times/month** to maximize profit
**Cluster – Big Spender**

**Upsell** other services

#### Advertising

Encourage Retailer to **advertise** their product which in **antecedent** in **association rules**

**Cross-sell** other products
**Bundling** products
*Frequently bought together*

**Highly associated Aisles and Department** were recommended to place close together (in store/web/app) to boost the sales of products.

**Customizing the store layout**, reduce time for Shopper, **increase order capacity**

Homepage/Product/Shopping Cart Recommendations

Help Retailer in better **Inventory Management**

### Cost

#### Variable Cost Full-Service Shopper

**Price Discrimination** for **Shopper's payment** in busy hour to attract more **Full-Service Shopper** to work at that time

#### Fixed Cost In-Store Shopper

**Customizing the store layout**, reduce time for Shopper, **reduce cost from In-Store Shopper** salary

**Allocate** more **In-Store Shopper** on the **weekends**, to maximize 29 working hour

| Day | Manhours | |
|---|---|---|
| Monday | 4 | |
| Tuesday | | 4 |
| Wednesday | 4 | |
| Thursday | | 4 |
| Friday | 5 | 5 |
| Saturday | 8 | 8 |
| Sunday | 8 | 8 |
| **Total** | **29** | **29** |


# instacart

## Customer Segmentation & Market Basket Analysis

## Thank You