Association rule analysis and Recommendation system using Instacart's transactions data

Brightics Ambassador
Final Project
Hasong Cho
2021.11.2

Agenda

1. Project Overview & Goals

2. Data Used & Analysis Process

3. Association Rules

4. ALS Recommend

5. Conclusion

1. Project Overview & Goals

Analysis of Grocery Transactions Data

- Instacart sells groceries and other goods from ~500 brick-and-mortar retail stores in the US. through an online platform

- The project aims to implement a association rule analysis and recommendation system using Instacart's sales data.

- This analysis allows the brick-and-mortar retail stores to closely position products that are purchased together

- Instacart can personalize recommendations to users using the recommendation algorithm.

Overview & Goals

Association Rules Analysis

Analytical Technique:

- Find out the **relevance** between products purchased in the same shopping cart and predict which product a customer will buy.

Effects:

- Product placement and configuration **optimization**
- **Increase sales** by finding relevant products, optimizing product placement in offline stores / apps, and organizing free gifts or package products.
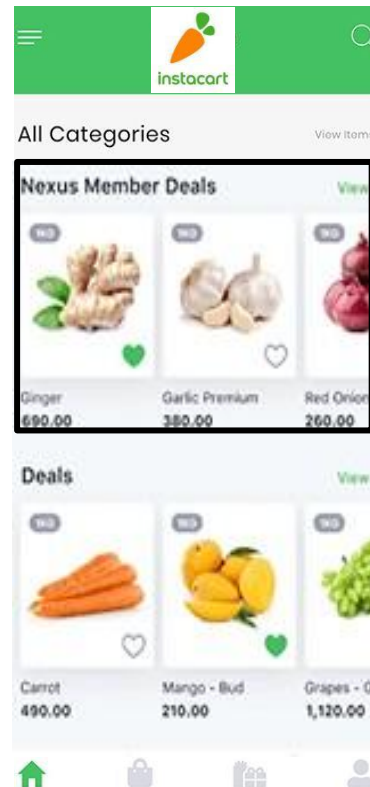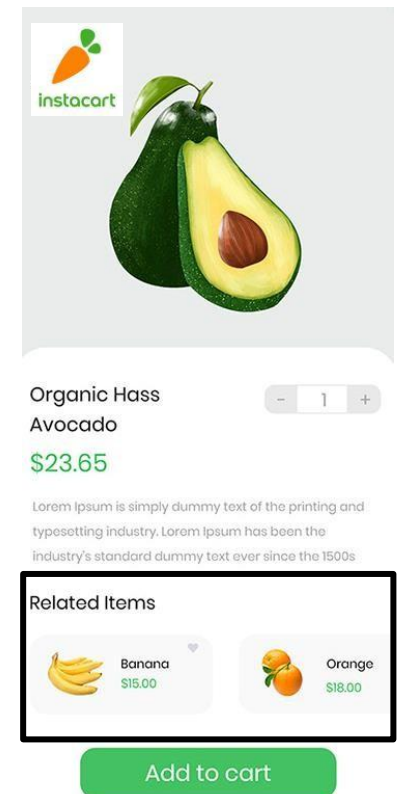
ALS Recommend

Analytical Technique:

- Implement customer-specific **customization** through product rating data goods

Effects:

- Increase in customer **satisfaction**
- Through a personalized recommendation system in the 'Related items' section of the Instacart app, customers can be encouraged to **purchase more products.**

2. Data Used & Analysis Process

Data Used

	user_id	order_id	product_id	department	aisle	product_name	add_to_cart_order	order_dow	order_hour_of_day
1	8382	317499	4	frozen	frozen m…	Smart Ones Cl…	13	2	19
2	102491	459014	4	frozen	frozen m…	Smart Ones Cl…	6	1	16
3	145313	508627	4	frozen	frozen m…	Smart Ones Cl…	13	1	17
4	167145	732095	4	frozen	frozen m…	Smart Ones Cl…	8	0	16
5	92944	770868	4	frozen	frozen m…	Smart Ones Cl…	6	0	15
6	21668	837726	4	frozen	frozen m…	Smart Ones Cl…	10	6	18
7	85136	864245	4	frozen	frozen m…	Smart Ones Cl…	3	6	13
8	109342	918568	4	frozen	frozen m…	Smart Ones Cl…	12	5	23
9	58746	934893	4	frozen	frozen m…	Smart Ones Cl…	6	3	16
10	177860	1083228	4	frozen	frozen m…	Smart Ones Cl…	1	3	14
11	38808	1418300	4	frozen	frozen m…	Smart Ones Cl…	5	4	13
12	160185	1633856	4	frozen	frozen m…	Smart Ones Cl…	8	6	18
13	44072	1815999	4	frozen	frozen m…	Smart Ones Cl…	10	3	14
14	200689	1941618	4	frozen	frozen m…	Smart Ones Cl…	21	2	17
15	150300	2153170	4	frozen	frozen m…	Smart Ones Cl…	4	2	11
16	29756	2179434	4	frozen	frozen m…	Smart Ones Cl…	14	0	14
17	67711	2372790	4	frozen	frozen m…	Smart Ones Cl…	2	2	19
18	91550	2439653	4	frozen	frozen m…	Smart Ones Cl…	16	0	15
19	47296	2472936	4	frozen	frozen m…	Smart Ones Cl…	9	6	12
20	167	2742787	4	frozen	frozen m…	Smart Ones Cl…	7	0	11
21	30889	2891043	4	frozen	frozen m…	Smart Ones Cl…	21	2	7
22	119229	3156707	4	frozen	frozen m…	Smart Ones Cl…	2	6	18
23	8382	317499	30	frozen	frozen m…	Three Cheese …	12	2	19
24	146034	941240	30	frozen	frozen m…	Three Cheese …	19	2	14
25	145522	1397014	30	frozen	frozen m…	Three Cheese …	6	1	10
26	38808	1418300	30	frozen	frozen m…	Three Cheese …	4	4	13
27	99996	1649913	30	frozen	frozen m…	Three Cheese …	12	4	11
28	7543	1711258	30	frozen	frozen m…	Three Cheese …	14	2	12
29	203269	1842155	30	frozen	frozen m…	Three Cheese …	15	0	13
30	45069	2237919	30	frozen	frozen m…	Three Cheese …	9	6	8
31	158212	2473363	30	frozen	frozen m…	Three Cheese …	3	4	9
32	65956	3344598	30	frozen	frozen m…	Three Cheese …	4	3	12

*Kaggle's Data
As of 2017

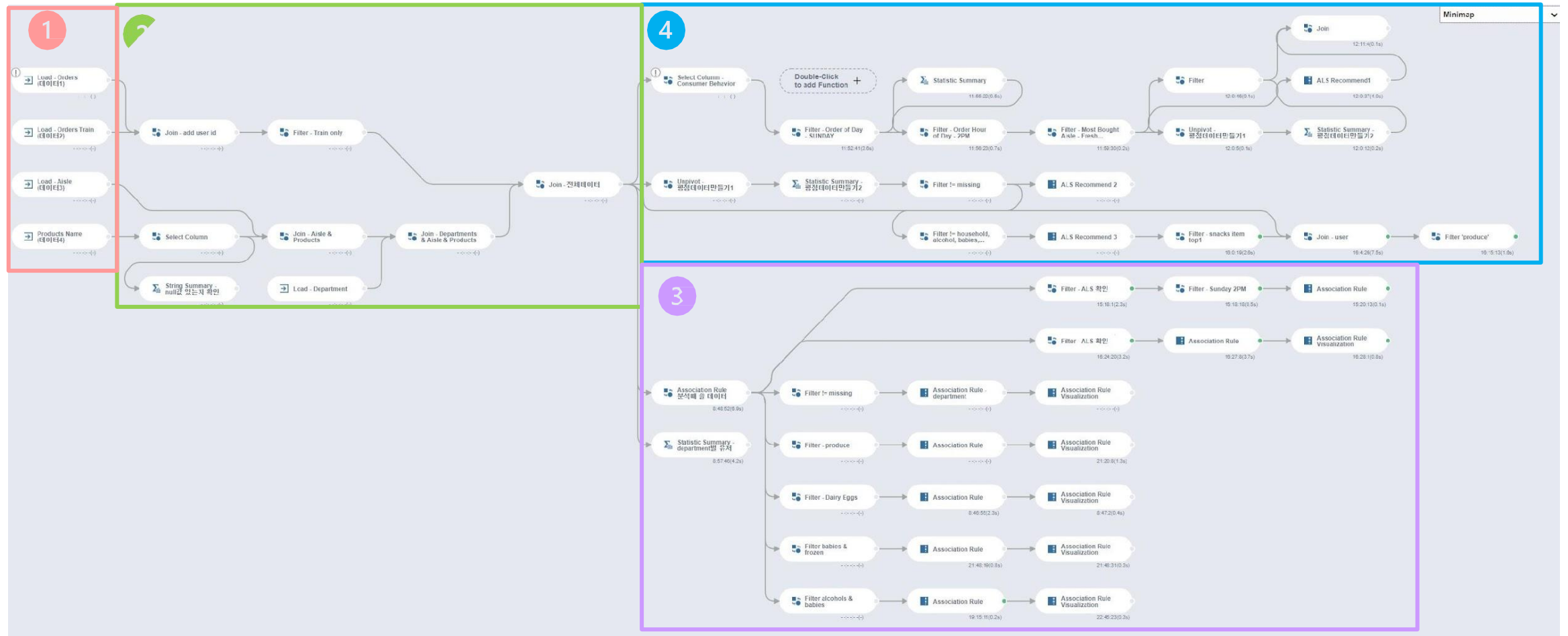## Customer Data
- User id
- Order id

Product Data
- Product id
- Department: main category where a product is classified
- Aisle: sub-category of a product
- Product name

Other Useful Data
- Add to cart order
- Order dow (day of week)
- Order hour of day

Analysis Process

1) Data Upload 2) EDA 3) Modeling & Visualization 4) Create Rating Data and build Recommendation System

Analysis Process - Association Rule Analysis

Analysis Process - Recommendation System

3. Association Rules

Association Rules Analysis Overview

Association between products from <u>same</u> department

Goal:

- Optimize product placement by finding the relevance of products belonging to the same department.

Method:

- Filter only one department out of the entire data and use Brightics' Association Rule function.

Data:

- Analyze the relationship between "Produce" and "Dairy & Eggs", which are the categories with the highest customer purchase frequency.

Association between products from a <u>different</u> department

Goal:

- Organize a gift or promotion by finding the relevance of products belonging to different departments.

Method:

- Filter several Departments of interest among the entire data and use the Association Rule function.

Data:

- Analyze the relationship between the most interesting categories: "Pantry" and "Frozen", and "Babies and Alcohol".

12

Tomatoes, Cucumbers

- When purchasing 'Baby Cucumbers', the probability of purchasing 'Packaged Grape Tomatoes' is **3.4 times (lift)** greater than the probability of purchasing 'Packaged Grape Tomatoes'.

- Since the color of the arrow leading from cucumber to cherry tomato is the darkest, **placing** the two products **next to each other** will encourage more customers to purchase these products together.

Banana, Strawberry, Avocado

- Of the entire purchase "Produce" history, transactions that include 'Bag of Organic Bananas' and 'Organic Strawberries' are **4.2% (support)**, and if 'Organic Strawberries' are purchased, the probability that 'Bag of Organic Bananas' will be purchased is **28% (confidence)**.

- Of the entire purchase "Produce" history, the transaction that includes 'Bananas' and 'Organic Avocado' is **3% (support)**, and if 'Organic Avocado' is purchased, the probability that 'Bananas' will be purchased is **30% (confidence).**

- Banana is a popular product that is included in many rules with high support or confidence. Placing it on the front page of the Instacart app will increase customer satisfaction, but even if not bananas will still be bought.

- Because the lift of bananas, strawberries, and avocados is smaller than that of cucumbers and cherry tomatoes, the effect of banana purchases on strawberry and avocado purchases is relatively small. After all, **placing cucumbers and cherry tomatoes next to each other** is more important than placing bananas, avocados, and strawberries next to each other.
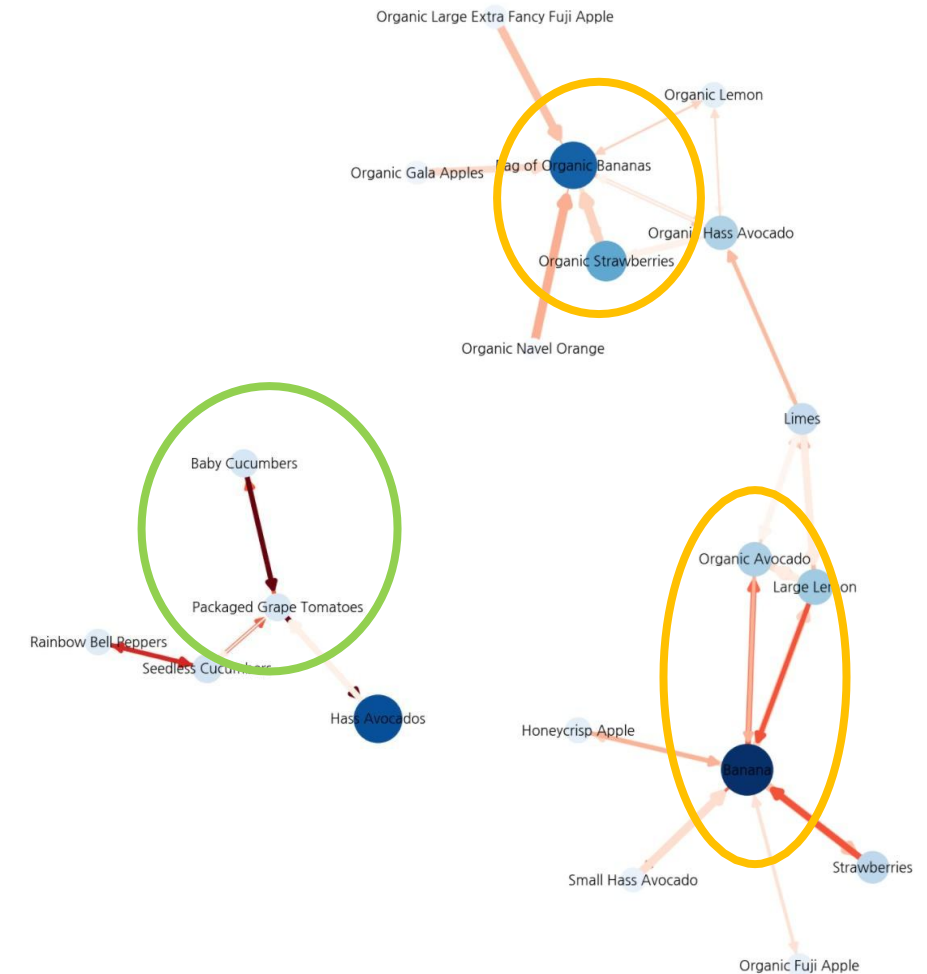
Yogurt, Milk, Cream

- Of the total purchase history of "Dairy & Eggs", transactions that include 'milk' and 'yogurt' are **13% (support)**, and if 'milk' is purchased, the probability that 'yogurt' will be purchased is **34% (confidence).** When purchasing 'milk', the probability of purchasing 'yogurt' is **1.02 times** greater than that of 'yogurt'.

- Of the total purchase history of "Dairy & Eggs", transactions that include 'milk' and 'cream' are **4.5% (support)**, and if 'milk' is purchased, the probability that 'cream' will be purchased is **13% (confidence)**. When purchasing 'milk', the probability of purchasing 'cream' is **1.03 times (lift)** greater than that of 'cream'.

- Milk is the most popular product among "Dairy & Eggs" because it is included in many rules with great support and confidence. On the other hand, since the lift of the rule that includes milk is almost 1, it can be seen that the purchase of milk has nothing to do with the purchase of yogurt or cream. After all, milk, yogurt, and cream **don't have to be placed next to each other.**
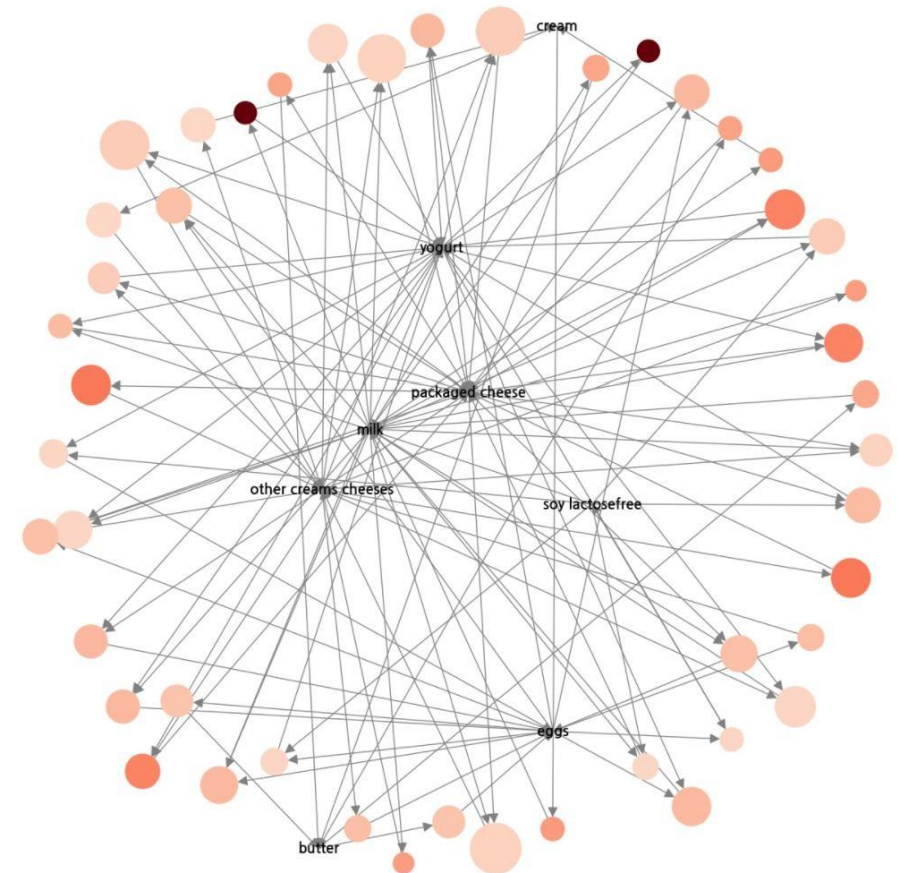
Condiments, Frozen Breakfast, Syrups

- Among the entire purchase history of "Frozen" and "Pantry", if 'honey syrups nectars' are purchased, the probability of 'frozen breakfast' being purchased is **14.2% (confidence)**, and 'frozen breakfast' is purchased when 'honey syrups nectars' is purchased. The probability of purchasing a 'frozen breakfast' is **1.3 times** greater than that of a 'frozen breakfast'.

- Among the total purchase history of "Frozen" and "Pantry", the probability of purchasing 'condiments' when 'frozen vegan vegetarian' is purchased is **12% (confidence),** and the probability of purchasing 'condiments' when 'frozen vegan vegetarian' is purchased is **1.1** times greater than that of 'condiments'.

- Recommendation: When customers buy frozen food, they buy sauce, and vice versa. It would be nice to have a promotion where they can buy sauce at a discounted price if they purchase frozen food over a certain price.

15
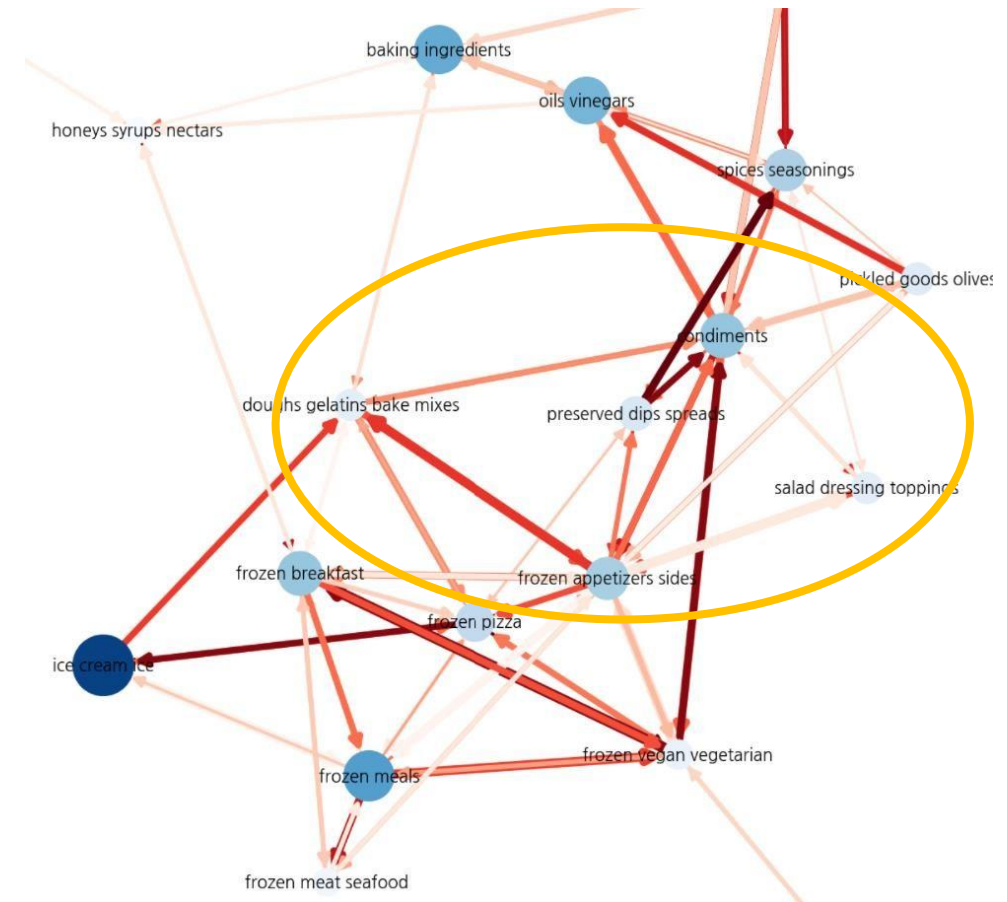
A. Products from "Alcohol" and "Babies" Department

Alcohol, Babies

- The most representative example of a correlation analysis discovered by Walmart in the 1990s is beer and diapers. The husband, who wants to drink beer on Friday evening, claims that he also buys beer with the excuse that he is going to buy diapers.

- The Association Rule function was used by filtering only the "Babies" and "Alcohol" departments to see if the data of 2017 customers shopping with the Instacart app would show the same pattern.

- As shown in the Network Digraph on the right, as a result of setting the minimum support to 0.001, there were **no rules for alcohol and baby products.**

- However, this association is limited to Instacart's data; the main customer for diapers at Walmart is male, and the gender of the user of the Instacart app is unknown.

- It is possible to find a correlation with a lower minimum support, but if so, it is not meaningful to proceed further with the analysis because the transaction that includes alcohol and baby products is less than 0.1% of the total purchase history of "Alcohol" and "Babies".
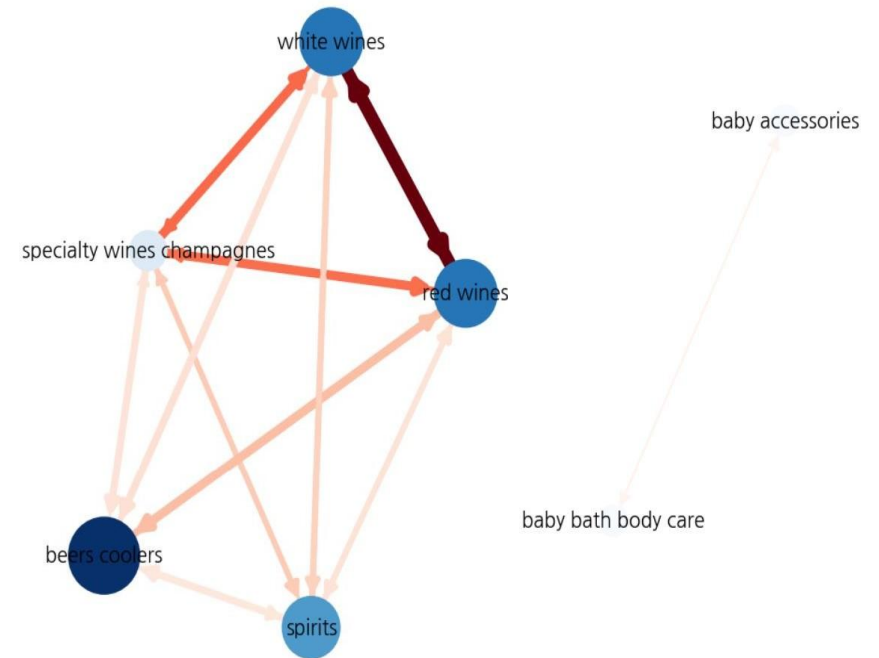
4. ALS Recommend

ALS Recommend Overview

Recommend products belonging to the <u>same</u> department as the purchased product

Goal:

- In the 'Related Items' section, products from the same department as the product purchased by the customer are recommended.

Method:

- Gather customer and product related data to create rating data.
- Filter only one department out of the entire data and use the ALS Recommend function as the customer's product rating data.

Data:

- Pick one specific product of interest and recommend a personalized product.
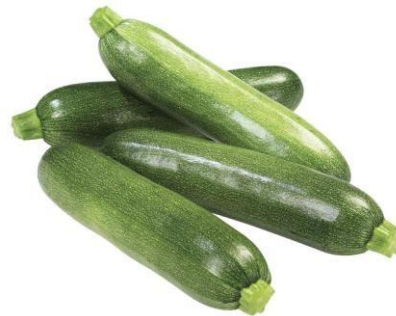

Organic Zucchini
$23.65

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s

Related Items

Banana $15.00   Orange $18.00

Add to cart

Recommend products belonging to a <u>different</u> department as the purchased product

Goal:

- In the 'Related Items' section, recommend products from a different department than the one the customer purchased.

Method:

- Gather customer and product related data to create rating data.
- After excluding a few departments from the entire data, the ALS Recommend function is used with the customer's product rating data.

Data:

- Pick two departments and make a little more detailed recommendation.

Frozen Pizza
$23.65

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s

Related Items

Banana $15.00   Orange $18.00

Add to cart

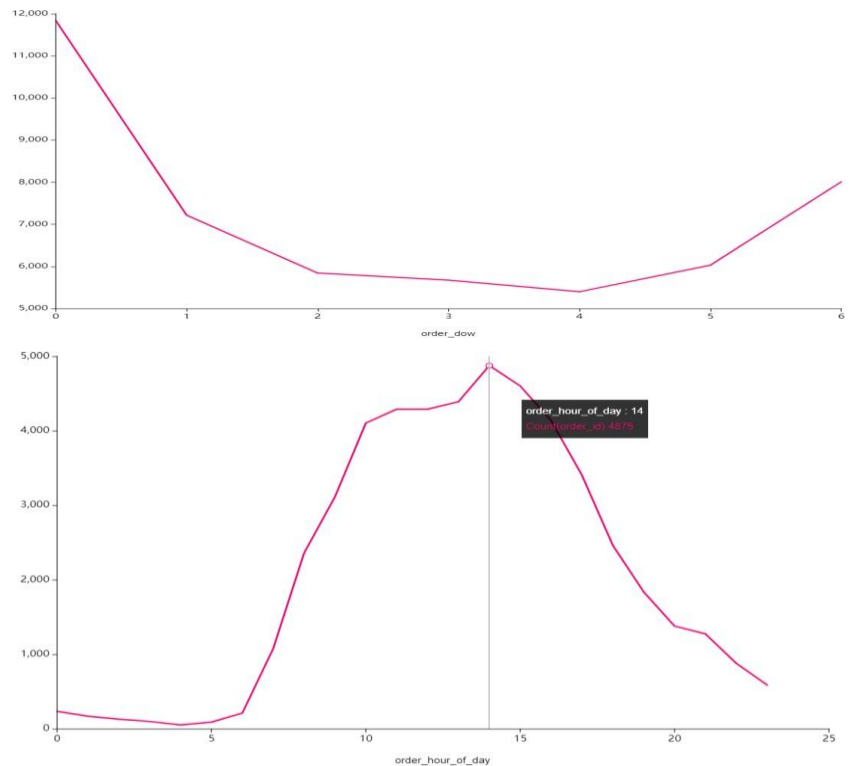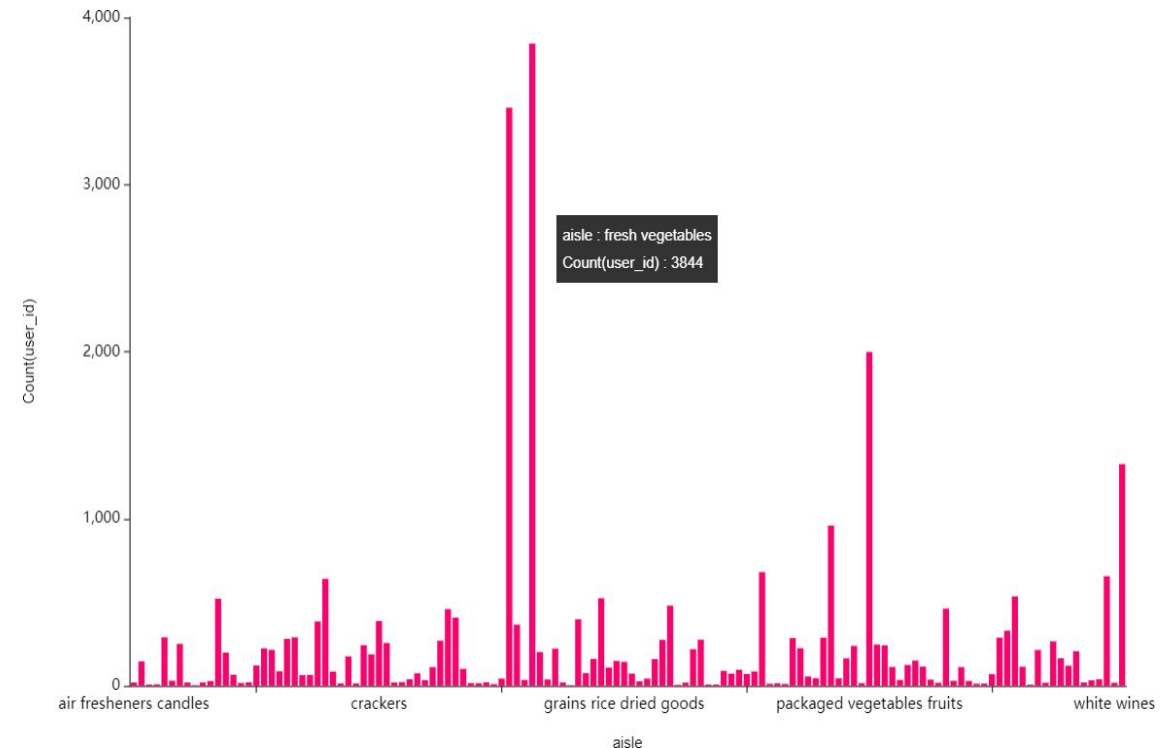# Recommending products from the same department as purchased products
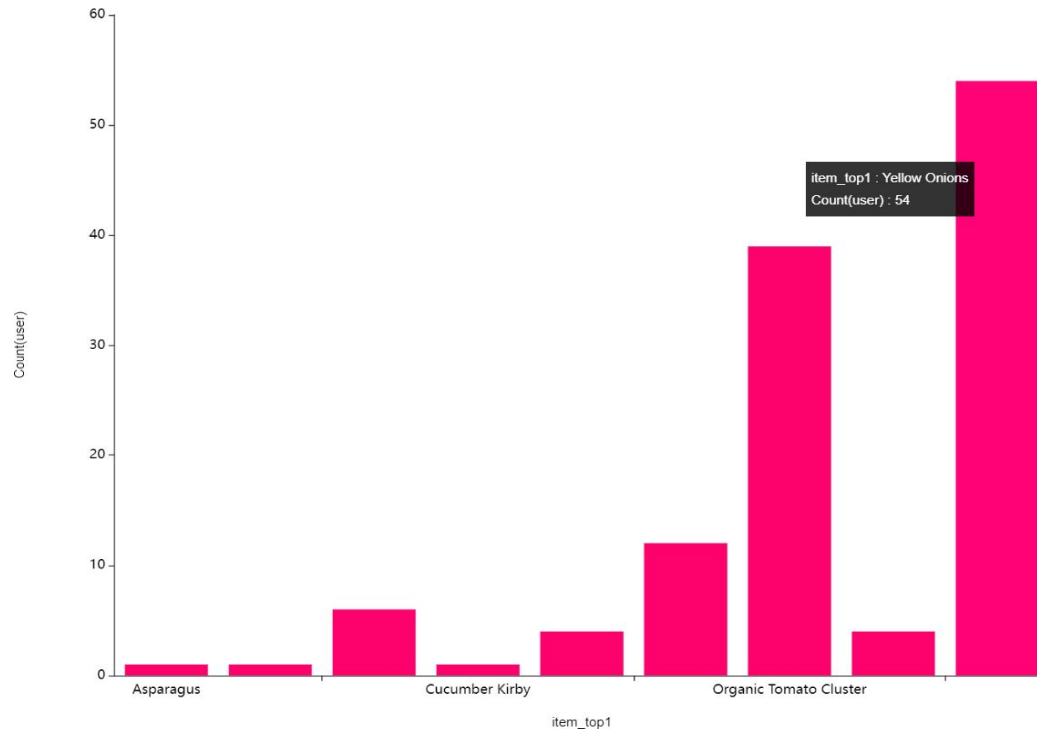
Customer who bought Organic Zucchini

- With Order day of Week and Order hour of day data, I found a time zone where the most customers ordered and selected the most sold products.

- The department that sold the most at 2:00pm on Sunday is "Produce", Aisle is "fresh vegetables", and the product is Organic Zucchini.

19

Recommending products from the same department as purchased products

Recommend Yellow Onion, Tomato Cluster, Red Bell Pepper to customers who bought Organic Zucchini

antecedent	consequent	confidence	lift ↑
["Organic Zucchini"]	["Organic Red Bell Pepper"]	0.08196721...	1.96855...
["Organic Zucchini"]	["Yellow Onions"]	0.09016393...	1.42032...
["Organic Zucchini"]	["Organic Tomato Cluster"]	0.06557377...	1.39225...

Related Items

- Yellow Onions and Organic Tomato Cluster are the most recommended products for customers who have purchased Organic Zucchini.

- As a result of association rule analysis, the purchase of Organic Zucchini has a positive (+) effect on the purchase of onions, tomatoes, and peppers. If recommended these products in the 'Related Items' section of customers who bought Organic Zucchini, they will be more likely to purchase them together.

Recommending products from different department as purchased products

Excluding some departments from the data
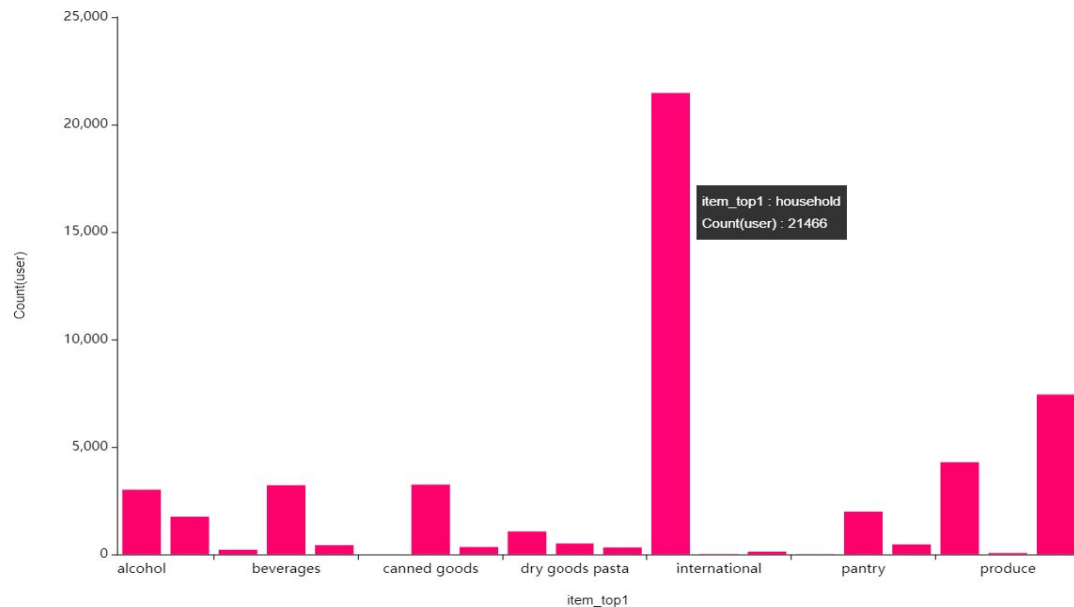
- The ALS Recommend function was executed with the rating data of products that included all departments.

- Among Item top 1, the department recommended by the most customers was "Household", followed by "Pets" and "alcohol".
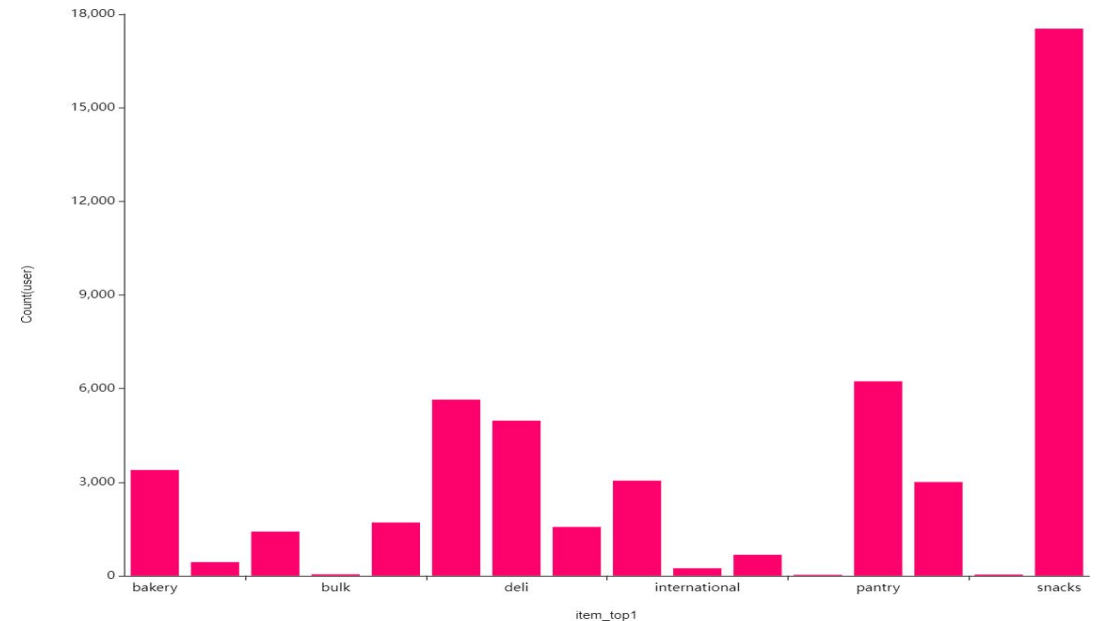
- "Alcohol," "Pets," and "Babies," which are unnecessary if recommended to customers who have purchased products from other departments, are excluded, and "Household," which is less likely to be purchased online (*based on 2017 data), was also excluded. .

- Among Item top 1, the Department recommended by most customers was "Snacks".

Recommending products from different department as purchased products

Recommend products in "Snacks" Department to customers who have purchased products in "Produce" Department

fresh fruits ■ fresh herbs ■ fresh vegetables ■ packaged produce ■ packaged vegetables fruits

| antecedent	consequent	confidence	lift ↓
["fresh fruits","fresh vegetables","packaged vegetables fruits"]	["nuts seeds dried fruit"]	0.12348632...	1.28834...
["fresh fruits","packaged vegetables fruits"]	["energy granola bars"]	0.11481933...	1.14061...

Related Items

Dried Fruits $15.00

Energy Bars $18.00

Add to cart

- When only customers who recommended "Snacks" were filtered, the most purchased Department was "Produce".

- Among them, fresh fruits, fresh herbs, fresh vegetables, packaged produce and packaged vegetable fruits were purchased.
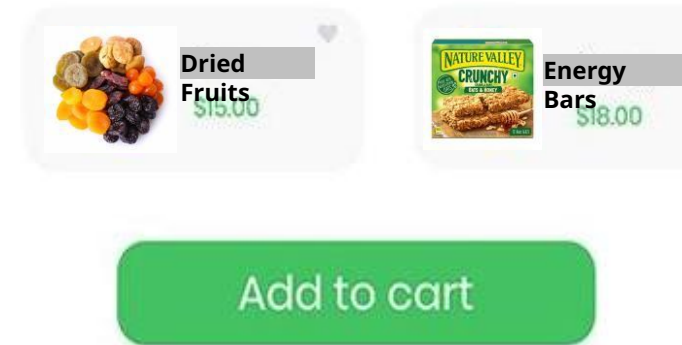
- As a result of the association rule analysis, the purchase of the product purchased by the customer has a positive (+) effect on the purchase of "Snacks".

- Among the customers who were recommended "Snacks", if they purchased fruits and vegetables, dried fruits and snacks such as energy bars are recommended in the "Related Items" category.

5. Conclusion

Summary

Evaluation:

- Using the results from association rule analysis and implementation of a recommendation system, Instacart can achieve the effect of increasing customer satisfaction and sales by optimizing product placement.

- However, different products were used during the process because same products had different names even though they were the same bananas: 'bag of bananas', 'bananas', and 'organic bananas'. In this case, grouping can be performed through text analysis and more accurate correlation analysis can be performed.

What I learned throughout the ambassador program:

- Through this program, I experienced an ML platform for the first time and learned full process of a data science project.

- In addition, I developed the ability to extract only the necessary information from raw data, and learned how to process it into new types of data depending on the situation.

- With the various functions of Brightics, I was able to build models and create visualizations for the recommendation algorithm analysis.

brightics/**studio**

Component based analytics studio on the web browser

THANK YOU