

Instacart 판매 데이터로 연관성 분석과 추천 시스템 구현



브라이틱스 서포터즈
개인 프로젝트
1 조 STARBRIGHTICS 조하송

Agenda



1. Project Overview & Goals



2. Data Used & Analysis Process



3. Association Rules



4. ALS Recommend



5. Conclusion



1. Project Overview & Goals

Grocery Transaction Data Analysis

- Instacart 는 미국의 약 500개의 오프라인 retail 매장에서 판매하는 식료품 및 잡화를 온라인 앱으로 판매한다.
- 이번 프로젝트는 Instacart 의 상품 판매 데이터를 활용해 연관성 분석과 추천 시스템을 구현하고자 한다.
- 연관성 분석을 통해 Instacart 앱 과 오프라인 retail 매장은 함께 구매되는 상품을 가깝게 배치할 수 있다.
- 협업 필터링 모델을 통해 Instacart 는 개인화된 추천을 할 수 있다.



Overview & Goals

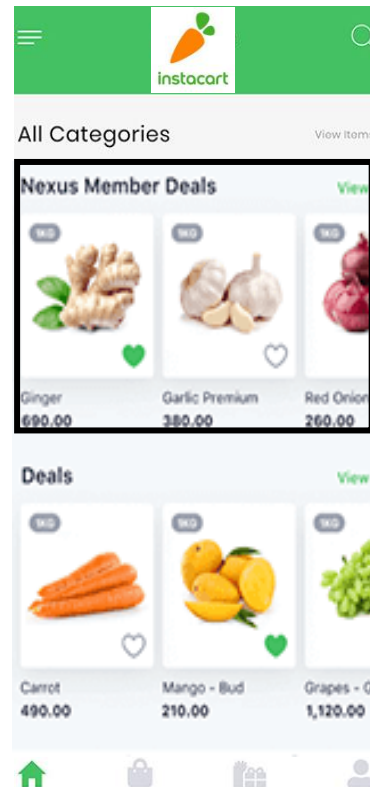
Association Rules Analysis

분석기법:

- 고객이 같은 장바구니에 구매한 상품들의 **연관성**을 알아보고 어떤 상품을 구매할 것인지 예측한다.

효과:

- 상품 배치 및 구성 최적화
- 연관성이 있는 상품들을 찾아 오프라인 매장 / 앱의 **상품 배치**를 최적화하고, 사은품 또는 **패키지 상품**을 구성하여 매출을 증가시킬 수 있다.



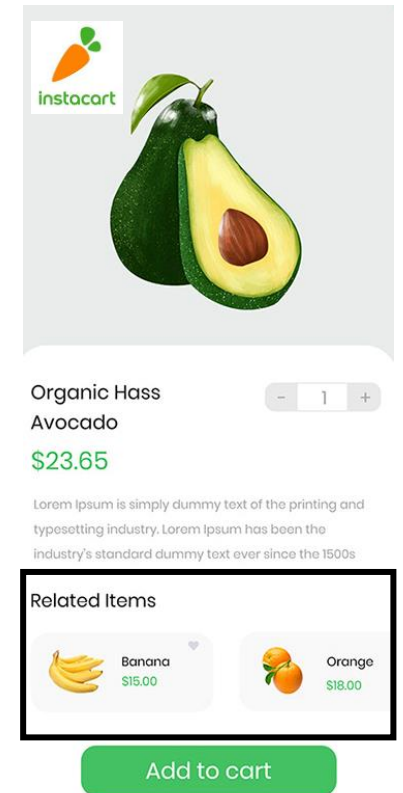
ALS Recommend

분석기법:

- 상품 평점 데이터를 통해 고객별 맞춤 상품을 추천해준다.

효과:

- 고객 만족도 증가
- Instacart 앱의 'Related items' 항목에 개인화된 추천 시스템을 통해 고객이 더 많은 상품을 구매하도록 유도할 수 있다.





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

고객 관련 데이터

- User id: 유저 아이디
- Order id: 구매 아이디

상품 관련 데이터

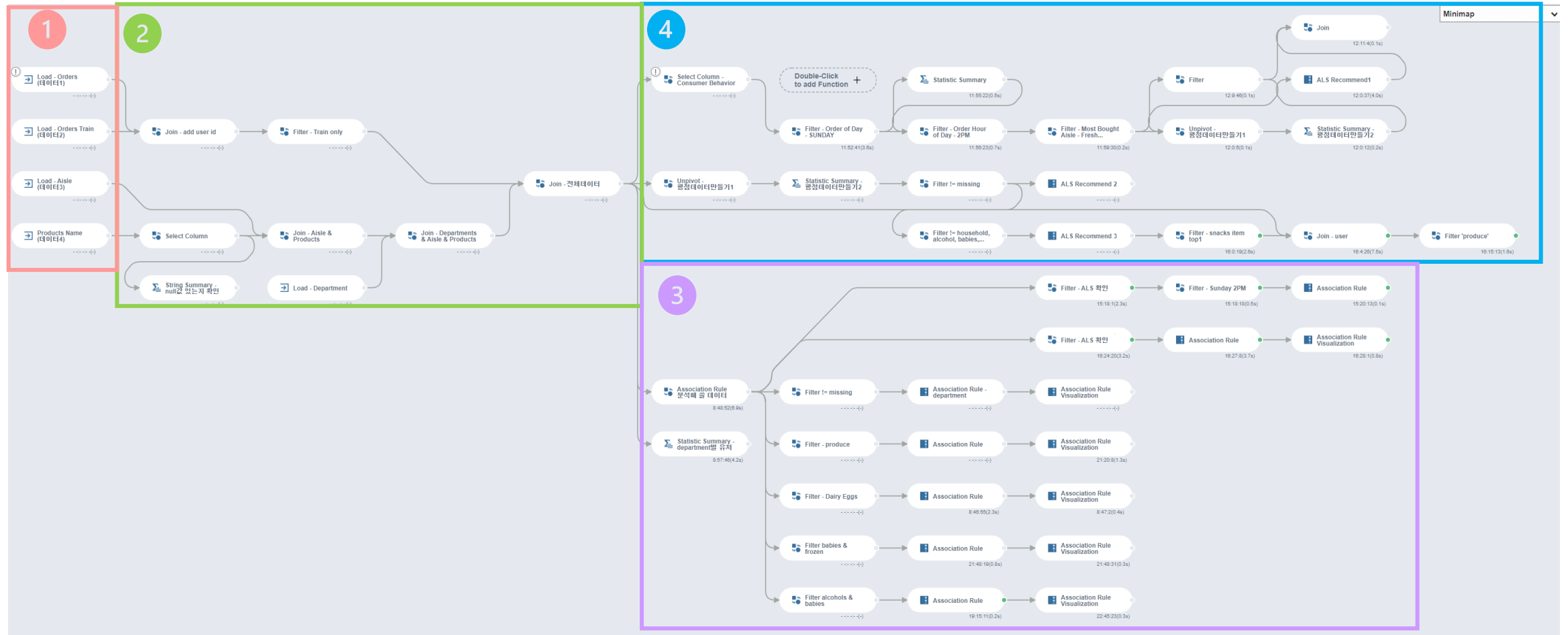
- Product id: 상품 아이디
- Department: 상품이 속한 main 카테고리
- Aisle: 상품이 속한 sub 카테고리
- Product name: 상품 이름

그 외 추가 분석에 유용한 데이터

- Add to cart order: 장바구니에 추가한 순서
- Order dow (day of week): 상품을 구매한 요일
- Order hour of day: 상품을 구매한 시간

*2017년도 데이터

Analysis Process



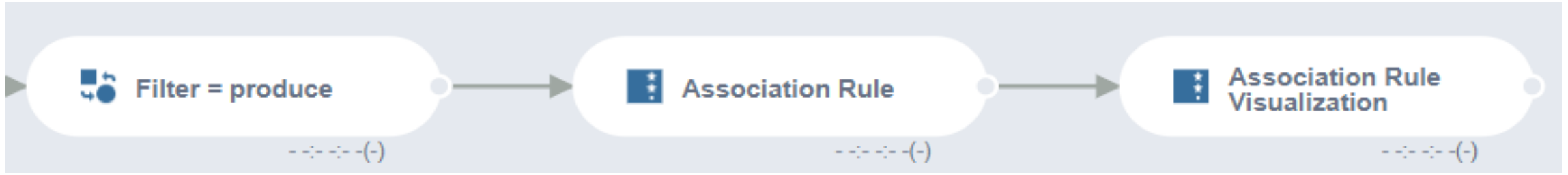
1) 데이터 업로드

2) 전처리 & EDA

3) 연관성 분석 모델링 및 시각화

4) 평점 데이터 생성 및 추천 시스템 구현

Analysis Process – 연관성 분석



Filter - produce

Inputs

- table
- Association Rule 분석때...
- out_table

Condition

department == 'produce'

Run

Table

	department	product_id	product_name	aisle
1	produce	31	White Pearl Oni...	packaged...
2	produce	31	White Pearl Oni...	packaged...
3	produce	31	White Pearl Oni...	packaged...
4	produce	31	White Pearl Oni...	packaged...
5	produce	31	White Pearl Oni...	packaged...
6	produce	31	White Pearl Oni...	packaged...
7	produce	43	Organic Clement...	packaged...
8	produce	43	Organic Clement...	packaged...
9	produce	43	Organic Clement...	packaged...
10	produce	66	European Style ...	packaged...
11	produce	66	European Style ...	packaged...
12	produce	66	European Style ...	packaged...
13	produce	66	European Style ...	packaged...
14	produce	66	European Style ...	packaged...
15	produce	66	European Style ...	packaged...
16	produce	66	European Style ...	packaged...
17	produce	66	European Style ...	packaged...
18	produce	66	European Style ...	packaged...
19	produce	66	European Style ...	packaged...
20	produce	66	European Style ...	packaged...
21	produce	89	Yogurt Fruit Dip ...	packaged...
22	produce	89	Yogurt Fruit Dip ...	packaged...
23	produce	89	Yogurt Fruit Dip ...	packaged...
24	produce	89	Yogurt Fruit Dip ...	packaged...
25	produce	120	Carillfinner Flore	packaged...

Association Rule

Inputs

- table
- Filter - produce
- out_table

Input Type

- ☒ User - single item
- ☐ User - multiple items
- ☐ Array

User Column

Double order_id

Item Column

String product_name

Min Support

0.01

Min Confidence

0.01

Min Lift

1

Max Lift

infinity

Table

	aisle	antecedent	consequent	supp
1	packaged...	["Organic Ba...	["Organic Gr...	0.01587
2	packaged...	["Organic Gr...	["Organic Ba...	0.01587
3	packaged...	["Organic Blu...	["Organic Ra...	0.01380
4	packaged...	["Organic Ra...	["Organic Blu...	0.01380
5	fresh veg...	["Organic Ga...	["Organic Yell...	0.01436
6	fresh veg...	["Organic Yell...	["Organic Ga...	0.01436
7	fresh fruits	["Bag of Org...	["Organic Na...	0.01304
8	fresh fruits	["Bag of Org...	["Organic Gal...	0.01122
9	fresh fruits	["Bag of Org...	["Organic Lar...	0.01346
10	fresh fruits	["Bag of Org...	["Organic Le...	0.01476
11	fresh fruits	["Bag of Org...	["Organic Ha...	0.03352
12	fresh fruits	["Bag of Org...	["Organic Str...	0.04256
13	fresh fruits	["Organic Na...	["Bag of Org...	0.01304
14	fresh fruits	["Organic Gal...	["Bag of Org...	0.01122
15	fresh fruits	["Organic Lar...	["Bag of Org...	0.01346
16	fresh fruits	["Banana"]	["Small Hass ...	0.01198
17	fresh fruits	["Banana"]	["Organic Fuj...	0.01576
18	fresh fruits	["Banana"]	["Honeycrisp ...	0.01705
19	fresh fruits	["Banana"]	["Strawberries"]	0.02396
20	fresh fruits	["Banana"]	["Organic Av...	0.03370
21	fresh fruits	["Banana"]	["Large Lemo...	0.02385
22	fresh fruits	["Small Hass ...	["Banana"]	0.01198
23	fresh fruits	["Organic Fuj...	["Banana"]	0.01576
24	fresh fruits	["Organic Ha...	["Organic Le...	0.01382
25	fresh fruits	["Organic Ha...	["Limes"]	0.01388
26	fresh fruits	["Organic Ha...	["Organic Str...	0.02132

Association Rule Visu...

Inputs

- table
- Association Rule
- out_table

Option

- ☒ single to single
- ☐ multiple to single
- ☐ multiple to multiple

Display Rule Number (except Single to Single)

- ☐ True
- ☒ False

Figure Size Multiplier

1 (0 < value)

Edge Length Multiplier

1 (0 < value)

Node Size Multiplier

1 (0 < value)

Font Size

15

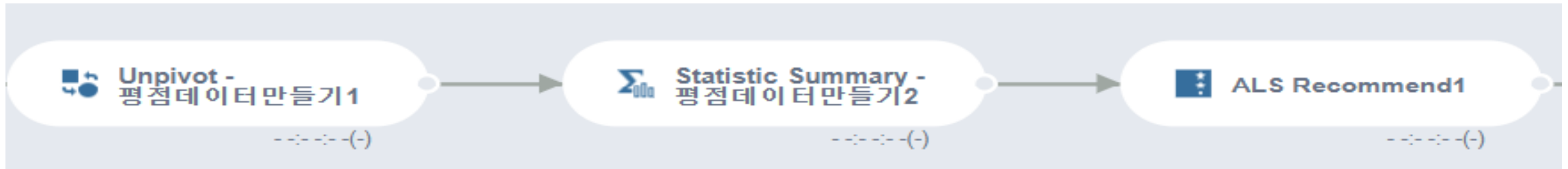
Network Digraph

Node color, size : support (0.027432181551165172~0.259)

Edge color : lift (1.0224773774762774~3.3650549906619)

Edge size : confidence (0.046192459681725945~0.37150)

Analysis Process – 추천 시스템 구현



Unpivot - 평점데이터만들기1

Inputs *

- table
- Join - 전체데이터
- table

Values *

1 columns selected

Select

String department

Identifiers

1 columns selected

Select

Double user_id

Variable Column Name

variable

Value Column Name

value

	user_id	variable	value
1	8382	department	frozen
2	102491	department	frozen
3	145313	department	frozen
4	167145	department	frozen
5	92944	department	frozen
6	21668	department	frozen
7	85136	department	frozen
8	109342	department	frozen
9	58746	department	frozen
10	177860	department	frozen
11	38808	department	frozen
12	160185	department	frozen
13	44072	department	frozen
14	200689	department	frozen
15	150300	department	frozen
16	29756	department	frozen
17	67711	department	frozen
18	91550	department	frozen
19	47296	department	frozen
20	167	department	frozen
21	30889	department	frozen
22	119229	department	frozen
23	8382	department	frozen
24	146034	department	frozen
25	145522	department	frozen
26	38808	department	frozen

Statistic Summary - 평...

Inputs *

- table
- Unpivot - 평점데이터만들기1
- out_table

Input Columns *

1 columns selected

Select

Double user_id

Target statistic *

Select All

Unselect All

☐ Max

☐ Min

☐ Range

☐ Sum

☐ Average

☐ Variance of Population

☐ Standard Deviation of Population

☐ Skewness of Sample

☐ Kurtosis

☐ Number of Row

☒ Number of Value

	value	user_id	column_name	num...
1	alcohol	14	user_id	
2	alcohol	125	user_id	
3	alcohol	147	user_id	
4	alcohol	187	user_id	
5	alcohol	267	user_id	
6	alcohol	314	user_id	
7	alcohol	315	user_id	
8	alcohol	446	user_id	
9	alcohol	525	user_id	
10	alcohol	599	user_id	
11	alcohol	781	user_id	
12	alcohol	788	user_id	
13	alcohol	829	user_id	
14	alcohol	838	user_id	
15	alcohol	1072	user_id	
16	alcohol	1122	user_id	
17	alcohol	1227	user_id	
18	alcohol	1283	user_id	
19	alcohol	1397	user_id	
20	alcohol	1474	user_id	
21	alcohol	1705	user_id	
22	alcohol	1830	user_id	
23	alcohol	1947	user_id	
24	alcohol	2044	user_id	
25	alcohol	2091	user_id	
26	alcohol	2419	user_id	

ALS Recommend 2

Inputs *

- table
- Filter != missing
- out_table

User Column *

Double user_id

Item Column *

String value

Rating Column *

Long num_of_value

Number of recommendations

3

Filter already liked items

☒ True

☐ False

Target Users

All users

Implicit Preference

	user	item_top1	rating_top1	item
1	1	pantry	2.4401316367...	frozer
2	2	household	4.1654177737...	pantr
3	5	alcohol	3.1216764444...	babie
4	7	household	2.4766326588...	break
5	8	household	4.6917840879...	snack
6	9	household	3.1647997935...	frozer
7	10	household	1.3053397568...	cann
8	13	beverages	1.6136042442...	house
9	14	deli	3.9376524827...	pets
10	17	snacks	1.6589546262...	bever
11	18	household	4.3655878675...	pets
12	21	household	2.4330135697...	dairy
13	23	pantry	3.1409150502...	perso
14	24	household	0.5938005722...	dairy
15	27	household	4.1260989730...	pantr
16	29	canned goo...	4.2496355785...	alcoh
17	30	household	0.5938005722...	dairy
18	34	beverages	1.6185534012...	house
19	37	household	2.5539890919...	pets
20	38	beverages	2.2718824996...	pets
21	41	household	3.0281229387...	frozer
22	42	household	1.2796533453...	alcoh
23	43	babies	5.5334532919...	pets
24	44	canned ooo...	2.6012065920...	house



3. Association Rules

Association Rules Overview

같은 Department에 속한 상품들의 연관성 분석

목적:

- 같은 **Department**에 속한 상품들의 연관성을 찾아 상품 배치를 최적화 시킨다.

방법:

- 전체 데이터중 **Department** 한개만 filter 하고 Association Rule 함수를 사용한다.

데이터:

- 고객의 구매 빈도가 가장 높은 카테고리인 "Produce" 와 "Dairy & Eggs"의 연관성을 각각 분석한다.



다른 Department에 속한 상품들의 연관성 분석

목적:

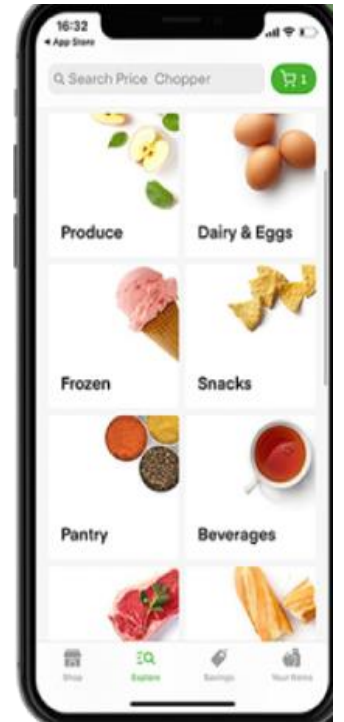
- 다른 **Department**에 속한 상품들의 연관성을 찾아 사은품 또는 프로모션을 구성해본다.

방법:

- 전체 데이터중 관심있는 **Department** 여러개를 filter하고 Association Rule 함수를 사용한다.

데이터:

- 가장 흥미로운 카테고리인 "Pantry" 와 "Frozen", "Babies" 와 Alcohol"의 연관성을 알아본다.



같은 Department에 속한 상품들의 연관성 분석

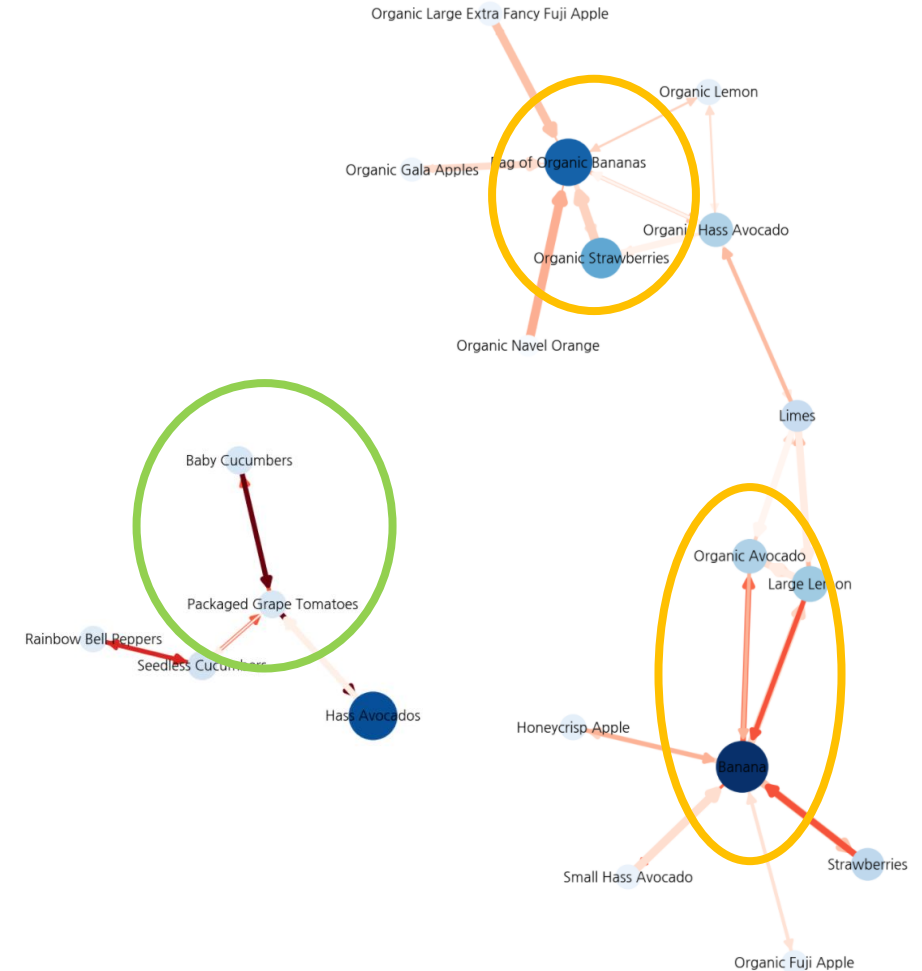
A. Department가 "Produce" 상품

Tomatoes, Cucumbers

- 'Baby Cucumbers'를 구매했을 때 'Packaged Grape Tomatoes'를 구매할 확률이 'Packaged Grape Tomatoes'를 구매할 확률에 비해 **3.4배(lift)** 크다.
- 오이에서 방울토마토로 이어지는 화살표의 색이 가장 진하기 때문에, 두 상품을 **옆에 배치**하면 더 많은 고객이 이 상품들을 같이 구매하게 될 것이다.

Banana, Strawberry, Avocado

- "Produce" 전체 내역중 'Bag of Organic Bananas'와 'Organic Strawberries'가 포함된 거래는 **4.2% (support)**이고, 'Organic Strawberries'가 구매된 경우 'Bag of Organic Bananas'가 구매될 확률은 **28% (confidence)**이다.
- "Produce" 전체 내역중 'Bananas'와 'Organic Avocado'가 포함된 거래는 **3% (support)**이고, 'Organic Avocado'가 구매된 경우 'Bananas'가 구매될 확률은 **30% (confidence)**이다.
- 이 외에도 바나나는 support 나 confidence 가 큰 규칙에 많이 포함되어 있는 인기 상품이다. Instacart 앱의 첫 페이지에 배치하면 고객 만족도가 올라가겠지만 그렇지 않아도 바나나는 계속 구매될 것이다.
- 바나나, 딸기, 아보카도의 lift 는 오이와 방울토마토의 lift 보다 작기 때문에, 바나나의 구매가 딸기나 아보카도의 구매에 미치는 영향이 비교적 작다. 결국, 바나나, 아보카도, 딸기를 서로 옆에 배치하는 것 보다 **오이와 방울 토마토를 옆에 배치하는게 더 중요하다.**

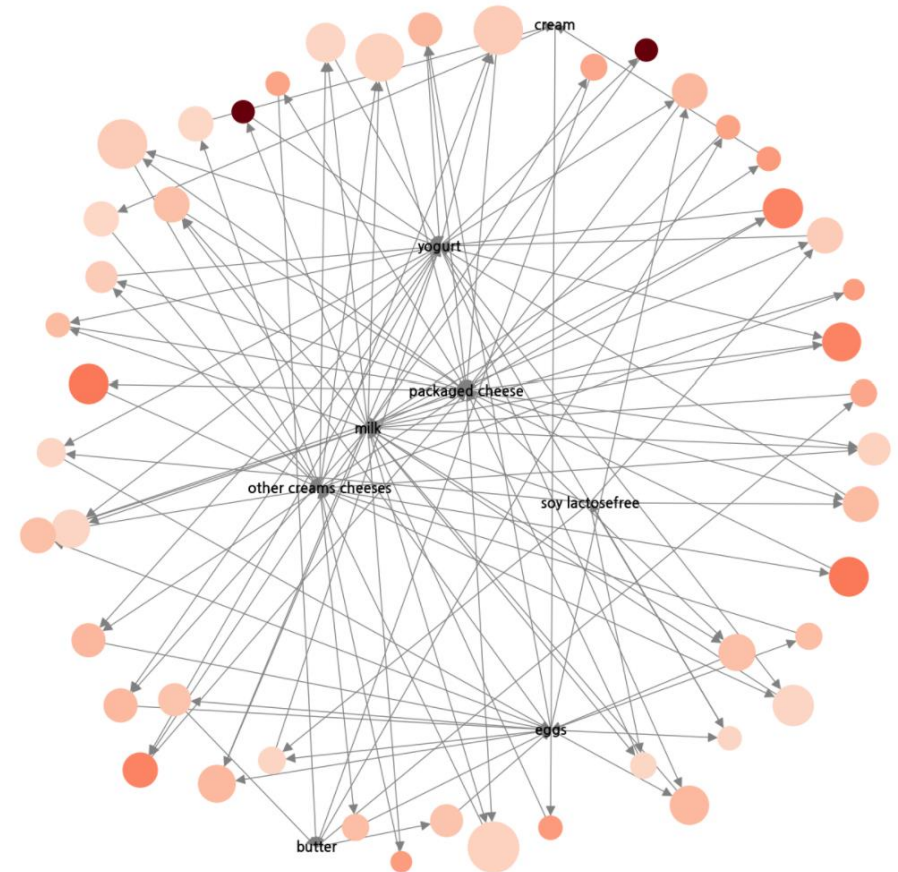


같은 Department에 속한 상품들의 연관성 분석

B. Department가 “Dairy & Eggs” 상품

Yogurt, Milk, Cream

- “Dairy & Eggs” 전체 구매 내역 중 ‘milk’와 ‘yogurt’가 포함된 거래는 **13% (support)** 이고, ‘milk’가 구매된 경우 ‘yogurt’가 구매될 확률은 **34% (confidence)** 이다. ‘milk’를 구매했을 때 ‘yogurt’를 구매할 확률이 ‘yogurt’를 구매할 확률에 비해 **1.02배 (lift)** 크다.
- “Dairy & Eggs” 전체 구매 내역 중 ‘milk’와 ‘cream’이 포함된 거래는 **4.5% (support)** 이고, ‘milk’가 구매된 경우 ‘cream’이 구매될 확률은 **13% (confidence)** 이다. ‘milk’를 구매했을 때 ‘cream’을 구매할 확률이 ‘cream’을 구매할 확률보다 **1.03배 (lift)** 크다.
- 이 외에도, 우유는 support와 confidence가 큰 규칙에 많이 포함되어있기 때문에 “Dairy & Eggs” 중 가장 인기 많은 상품이다. 반면, 우유가 포함된 규칙은 lift가 거의 1이기 때문에 우유의 구매가 요거트나 크림의 구매와 상관 없다고 볼 수 있다. 결국, 우유, 요거트, 크림을 **서로 옆에 배치할 필요는 없다**.

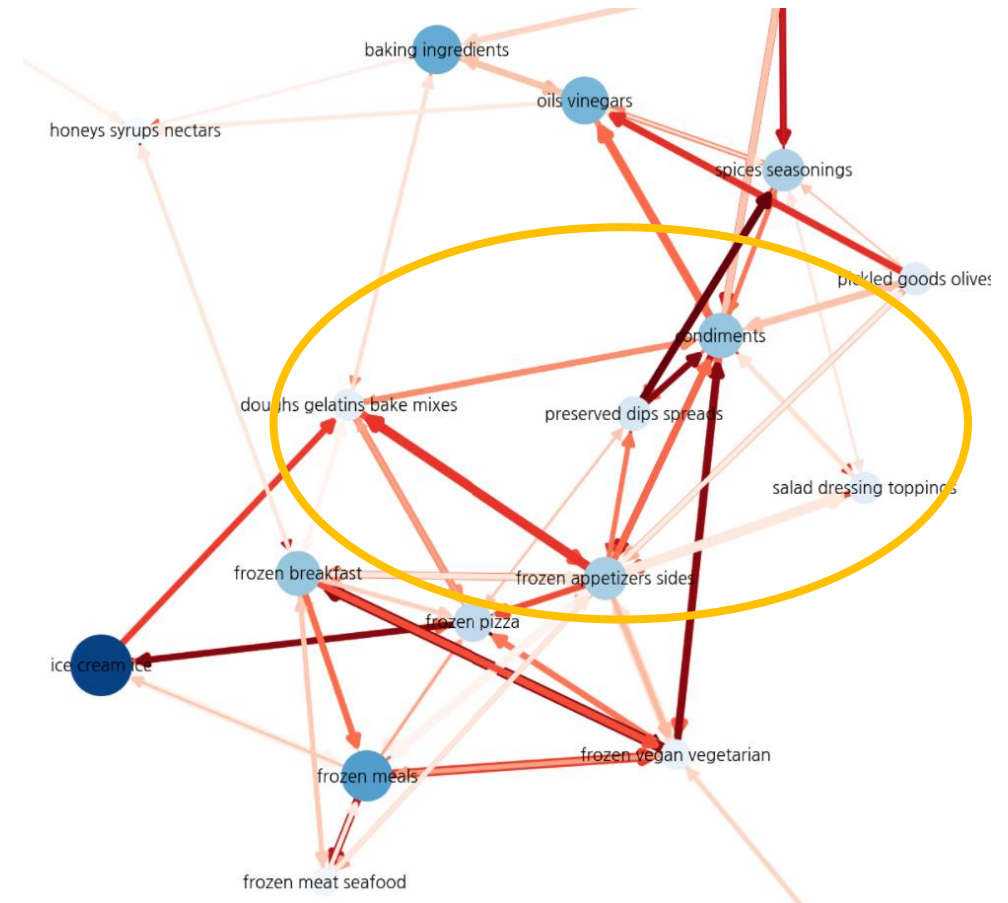


다른 Department에 속한 상품들의 연관성 분석

A. Department가 "Frozen" 과 "Pantry" 상품

Condiments, Frozen Breakfast, Syrups

- "Frozen" 과 "Pantry" 전체 구매 내역 중, 'honey syrups nectars' 가 구매된 경우 'frozen breakfast'가 구매될 확률은 **14.2% (confidence)** 이고 'honey syrups nectars'를 구매했을 때 'frozen breakfast'를 구매할 확률이 'frozen breakfast'를 구매할 확률에 비해 **1.3배(lift)** 크다.
- "Frozen" 과 "Pantry" 전체 구매 내역 중, 'frozen vegan vegetarian' 이 구매된 경우 'condiments'가 구매될 확률은 **12% (confidence)** 이고 'frozen vegan vegetarian'을 구매했을 때 'condiments'를 구매할 확률이 'condiments'를 구매할 확률에 비해 **1.1배(lift)** 크다.
- 냉동식품을 구매할 때 소스를 사는 경우와, 그 반대의 경우도 있다. 냉동식품을 일정 가격 이상 구매하면 소스를 **할인 가격에 구매할 수 있는 프로모션**이 있어도 좋을 것 같다.

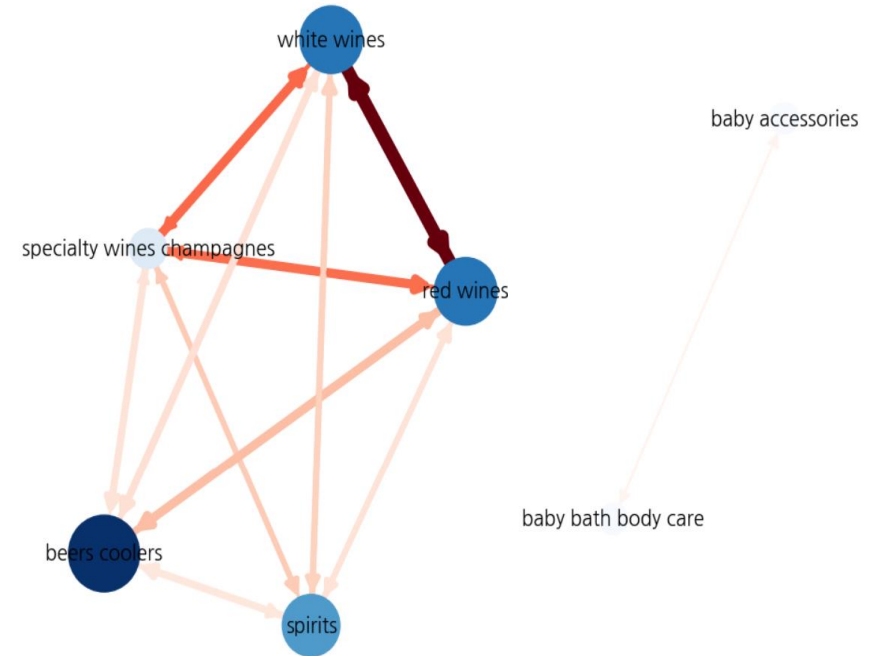


다른 Department에 속한 상품들의 연관성 분석

B. Department가 "Alcohol" 과 "Babies" 상품

Alcohol, Babies

- 1990년대 월마트가 발견한 연관성 분석의 가장 대표적인 사례는 맥주와 기저귀이다. 금요일 저녁에 맥주를 마시고 싶은 남편이 기저귀를 사러 간다는 핑계로 맥주도 함께 사 온다는 주장이다.
- Instacart 앱으로 장을 보는 2017년도 고객들의 데이터도 같은 패턴을 보일지 "Babies" 와 "Alcohol" Department 만 filter해서 Association Rule 함수를 사용했다.
- 우측에 있는 Network Digraph에서도 보이듯 minimum support 를 0.001로 진행한 결과 **술과 유아용품의 규칙은 없었다.**
- 단, 월마트에서 기저귀를 사는 주 고객은 남성이었고 Instacart 앱의 이용자의 성별은 모르기 때문에, 이 연관성은 Instacart 의 데이터에 제한되어있다.
- 더 낮은 minimum support 로 연관성을 찾을 수도 있겠지만, 그렇다면 "Alcohol" 과 "Babies" 전체 구매 내역 중 술과 유아용품이 포함된 거래는 0.1% 도 안 된다는 것이기 때문에 분석을 더 진행할 큰 의미가 없다.





4. ALS Recommend

ALS Recommend Overview

구매한 상품과 같은 Department에 속한 상품을 추천

목적:

- 'Related Items' 섹션에 고객이 구매한 상품과 **같은 Department**인 상품을 추천한다.

방법:

- 고객과 상품 관련 데이터를 취합해 평점 데이터를 생성한다.
- 전체 데이터 중 **한개의 Department**만 filter 하고 고객의 상품 평점 데이터로 ALS Recommend 함수를 사용한다.

데이터:

- 가장 흥미로운 특정 상품 한개를 골라 맞춤형 상품을 추천해준다.

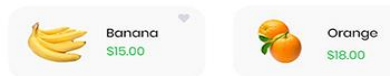


Organic Zucchini

\$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



Add to cart

구매한 상품과 다른 Department에 속한 상품을 추천

목적:

- 'Related Items' 섹션에 고객이 구매한 상품과 **다른 Department**인 상품을 추천한다.

방법:

- 고객과 상품 관련 데이터를 취합해 평점 데이터를 생성한다.
- 전체 데이터 중 **몇개의 Department**만 제외한 후 고객의 상품 평점 데이터로 ALS Recommend 함수를 사용한다.

데이터:

- 두가지 Department를 골라 조금 더 자세한 추천을 해준다.



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

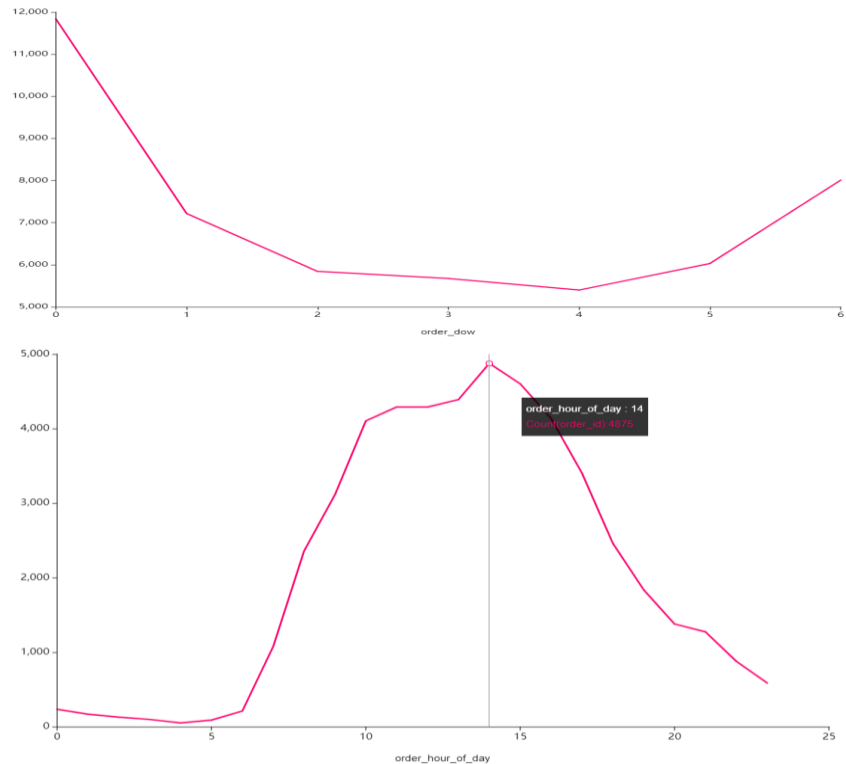
Related Items



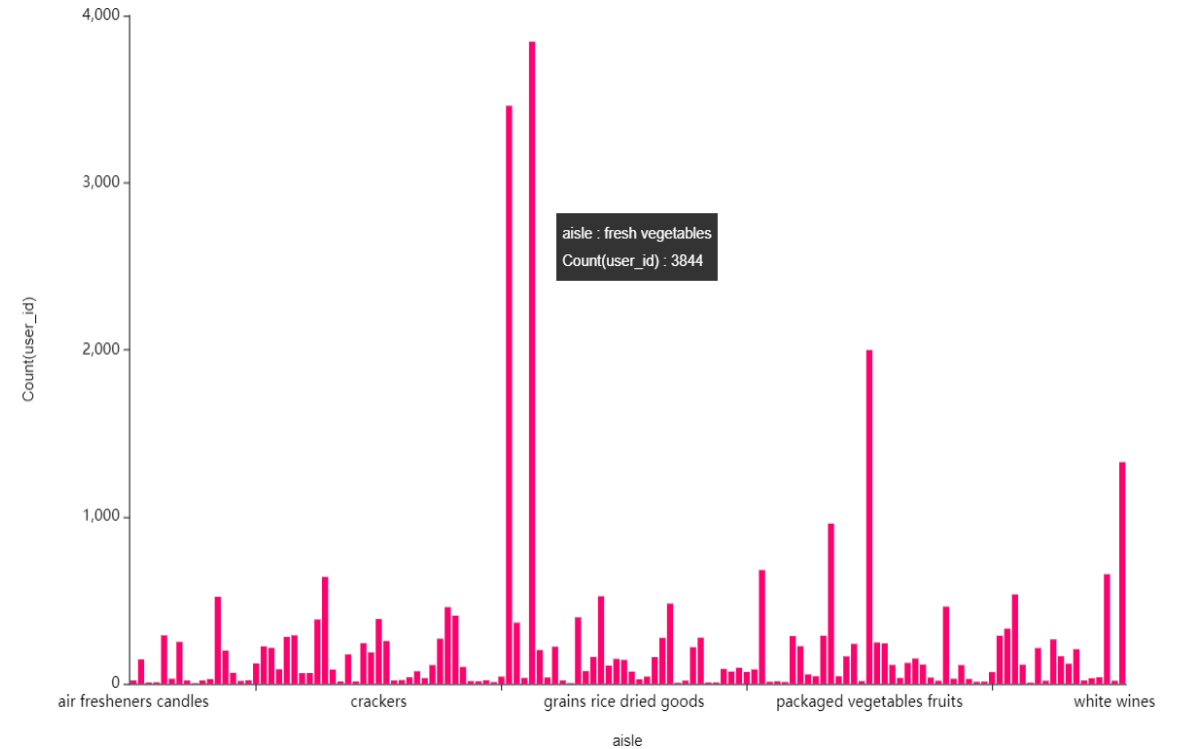
Add to cart

구매한 상품과 같은 Department 상품 추천 하기

Organic Zucchini (유기농 애호박) 을 구매한 경우



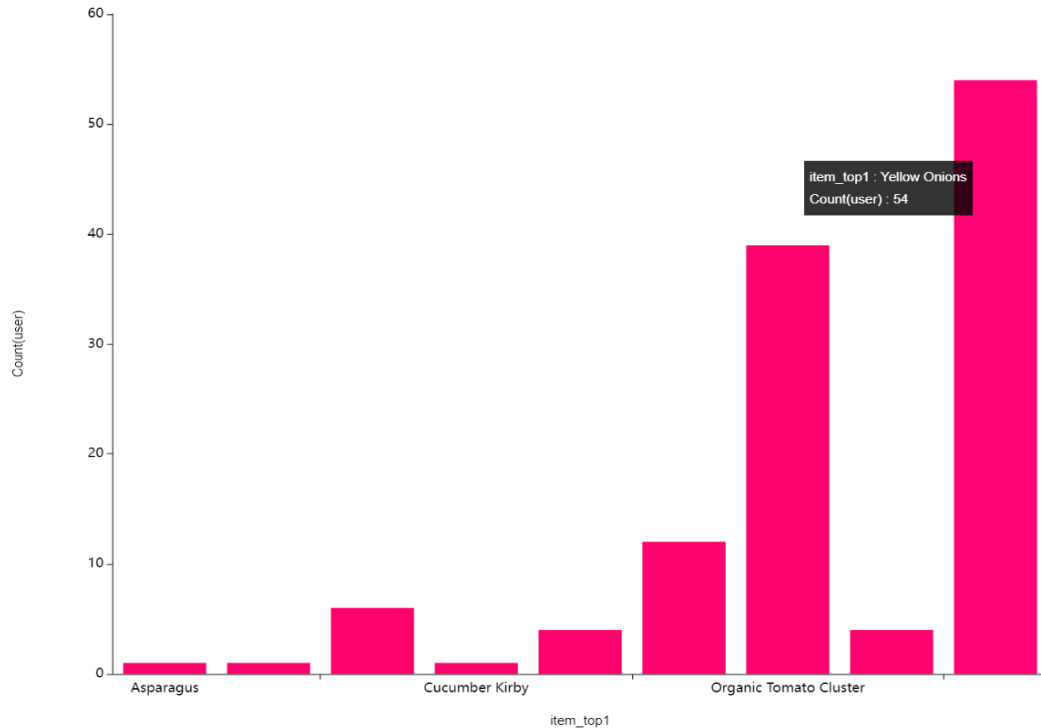
- Order day of Week, Order hour of day 데이터로 가장 많은 고객이 주문한 시간대를 찾고, 가장 많이 팔린 상품을 선택했다.



- 일요일 2시에 가장 많이 팔린 Department 는 "Produce", Aisle 은 "fresh vegetables", 상품은 Organic Zucchini 이다.

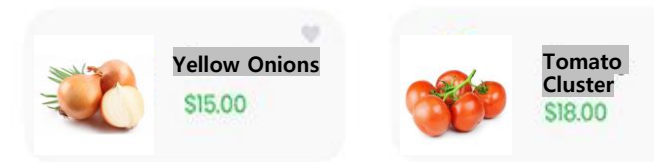
구매한 상품과 같은 Department에 속한 상품 추천 하기

Organic Zucchini 를 구매한 고객에게 Yellow Onion, Tomato Cluster, Red Bell Pepper를 추천



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



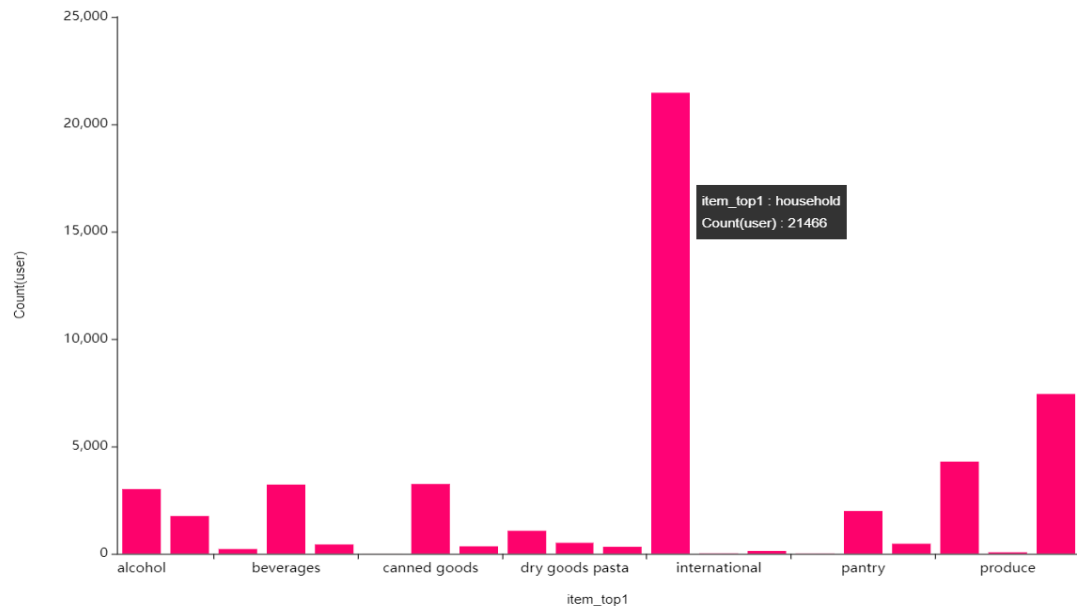
Add to cart

- Organic Zucchini를 구매한 고객에게 가장 많이 추천된 item top 1 상품은 Yellow Onions, Organic Tomato Cluster가 있다.

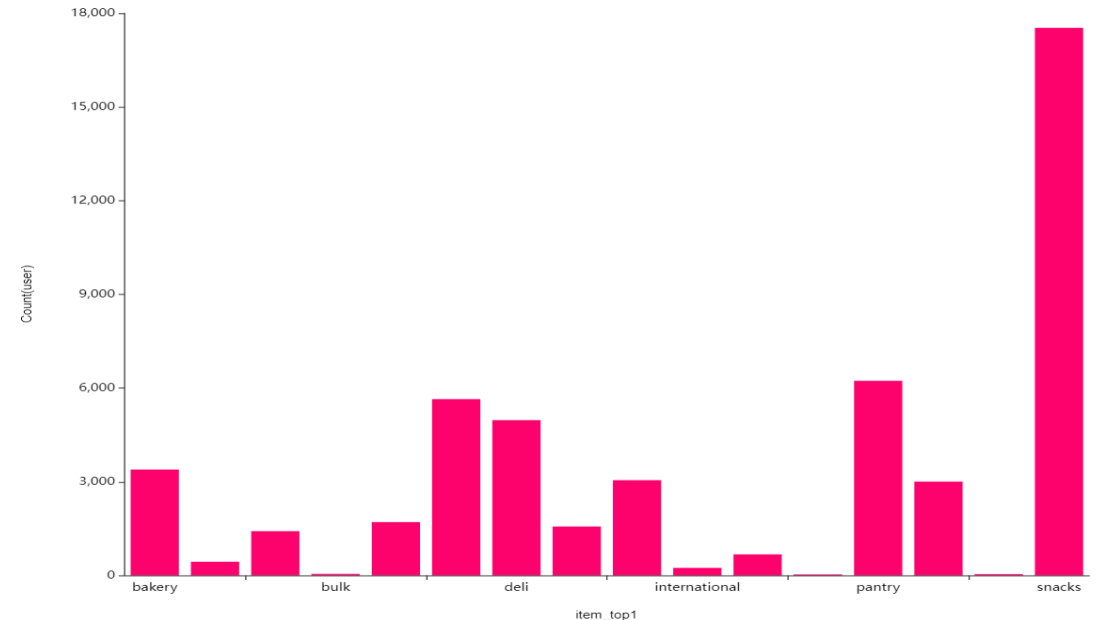
- 연관성 분석을 진행한 결과, Organic Zucchini 의 구매는 양파, 토마토, 후추의 구매에 양 (+)의 영향력이 있다. Organic Zucchini 를 산 고객의 'Related Items' 항목에 이 상품들을 추천해주면 같이 구매할 확률이 높아질 것이다.

구매한 상품과 다른 Department에 속한 상품 추천 하기

구매한 상품의 평점 데이터 중 일부 Department 제외



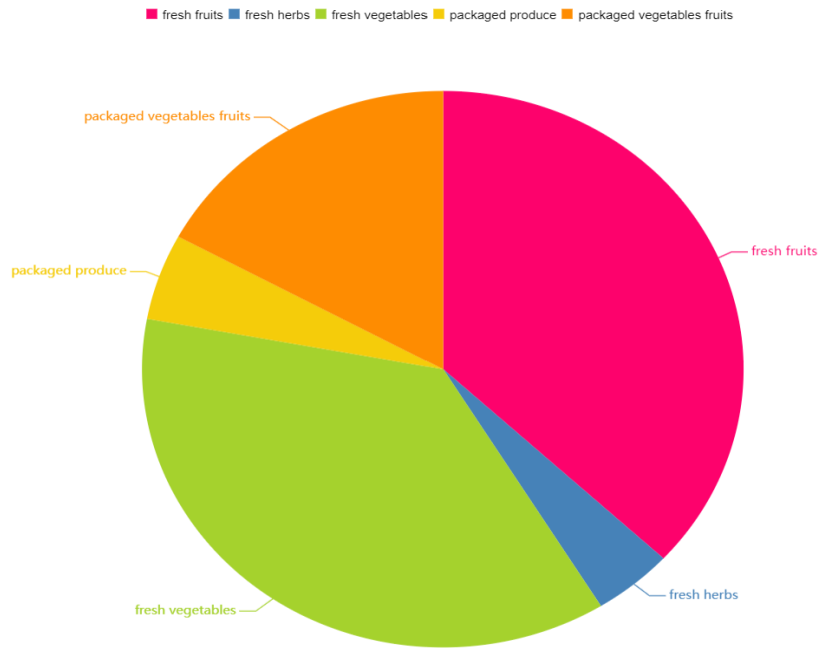
- 모든 Department 가 포함된 상품의 평점 데이터로 ALS Recommend 함수를 실행했다.
- Item top 1 중 가장 많은 고객에게 추천한 Department 는 "Household", 그다음은 "Pets", 그리고 "alcohol" 이었다.



- 다른 Department 의 상품을 구매한 고객에게 추천하면 불필요할 확률이 비교적 높은 "Alcohol", "Pets", "Babies" 를 제외하고 온라인으로 구매하는 편이 낮은 "Household" (*데이터 2017년도 기준) 도 제외했다.
- Item top 1 중 가장 많은 고객에게 추천한 Department 는 "Snacks" 였다.

구매한 상품과 다른 Department에 속한 상품 추천 하기

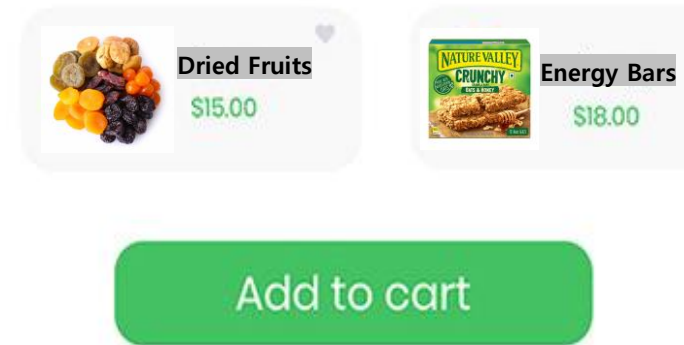
“Produce” Department 에 있는 상품을 구매한 고객에게 “Snacks” Department 에 있는 상품을 추천



- “Snacks”를 추천받은 고객만 filter 했을 때 가장 많이 구매한 음식의 Department는 “Produce” 였다.
- 그중에서도 fresh fruits, fresh herbs, fresh vegetables, packaged produce 와 packaged vegetable fruits 를 구매했다.

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



- 연관성 분석을 진행한 결과, 고객이 구매한 상품의 구매는 “Snacks”의 구매에 양 (+)의 영향력이 있다.
- Item top 1로 “Snacks”를 추천받은 고객 중 과일과 야채를 구매했다면, ‘Related Items’ 항목에 **말린 과일과 에너지바** 같은 간식을 추천해 준다.

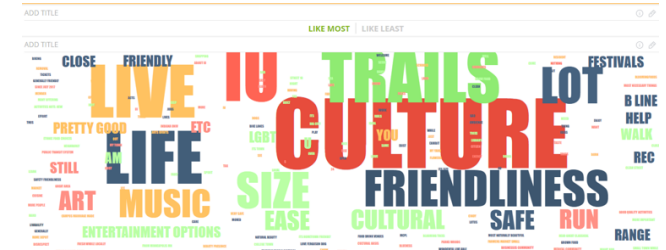


5. Conclusion

Evaluation and What I learned

Evaluation:

- Instacart는 연관성 분석을 통해 최적화된 상품 배치를 하고, 추천 시스템 구현을 통해 고객 만족도와 매출을 증가시키는 효과를 얻을 수 있다.
- 단, 연관성 분석을 할 때 'bag of bananas', 'bananas', 'organic bananas' 등 같은 바나나임에도 불구하고 '이름'이 다르기 때문에 다른 상품으로 진행됐다. 이런 경우 further research로 text analysis 를 통해 grouping 을 진행하고 더 정확한 연관성 분석을 할 수 있다.



What I learned:

- 이번 프로젝트를 통해 처음으로 데이터 분석 플랫폼을 경험하고, 데이터 분석의 절차를 알게 된 것이 가장 유익했다.
- 또, Raw data 중 분석에 필요한 데이터만 추출할 수 있는 능력을 키우고, 상황에 따라 새로운 형태의 데이터로 가공하는 법을 알게 되었다.
- 브라이틱스의 다양한 함수로 추천 알고리즘 분석에 필요한 모델링과 시각화를 실행할 수 있었다.

brightics/studio

Component based analytics studio on the web
browser



THANK YOU

