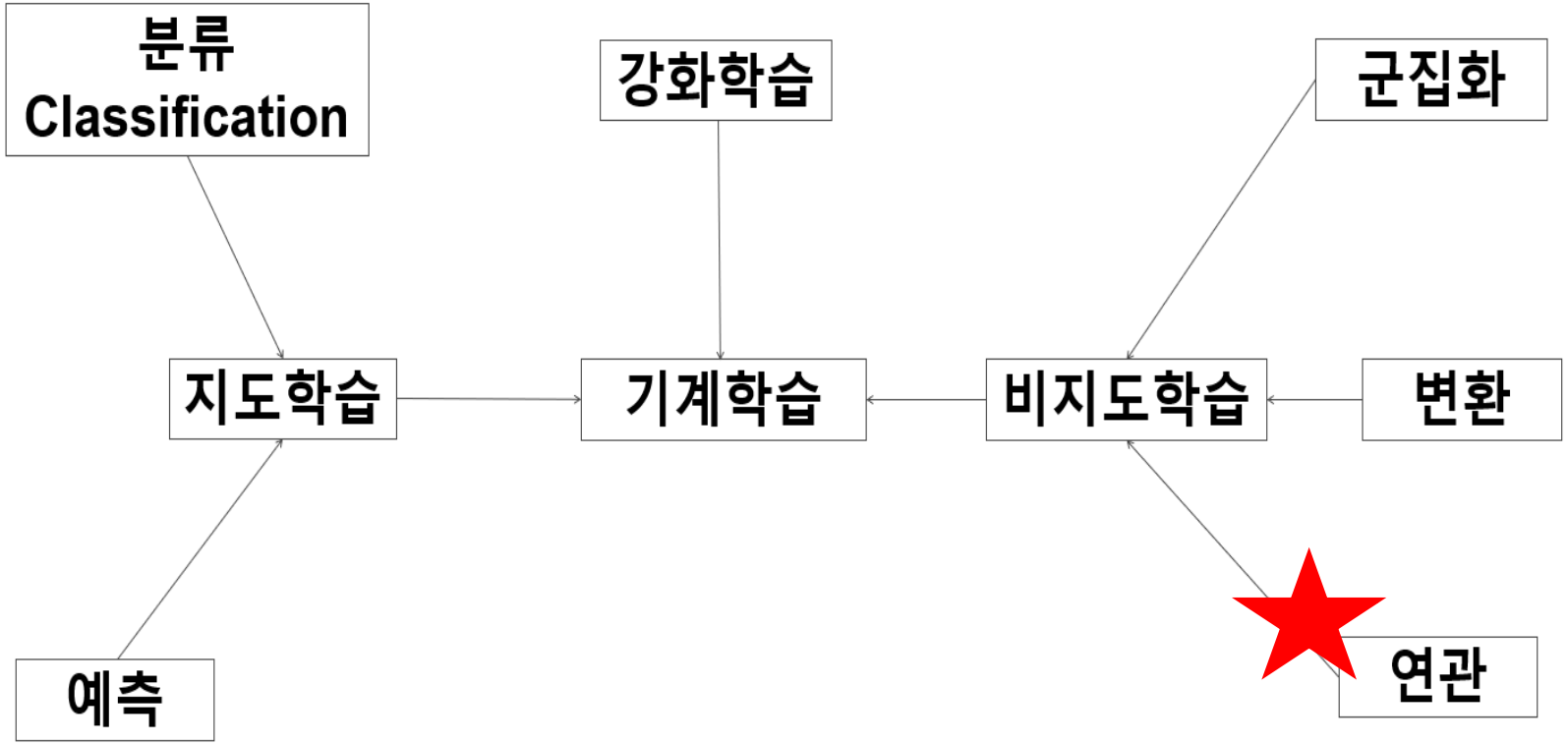
연관분석

홍익 대학교
Hyun-Sun Ryu

머신러닝의 분류



연관분석(Association Analysis)

연관 분석이란?

장바구니분석(Market Basket Analysis)

서열 분석(Sequence Analysis)

연관 분석이란?

장바구니분석 | 서열분석

장바구니에 무엇이 같이 들어 있는지에 대한 분석

연관 분석이란?

장바구니분석 | 서열분석

A를 산 다음에 B를 산다

연관 분석 적용 분야

유통업 | 호텔/숙박업 | 금융사 | 보험사

상품 제안

연관 분석 적용 분야

유통업 | **호텔/숙박업** | 금융사 | 보험사

서비스 제안

연관 분석 적용 분야

유통업 | 호텔/숙박업 | 금융사 | 보험사

고객 발굴

연관 분석 적용 분야

유통업 | 호텔/숙박업 | 금융사 | **보험사**

부정행위 적발

연관성 규칙(Association rule)

조건과 반응의 형태

If A then B : A가 일어나면 B가 일어난다

$A \rightarrow B$

연관성 규칙 유형

- Useful result
- Trivial result
- Inexplicable result

연관성 규칙 유형

Useful result	Trivial result	Inexplicable result

마케팅 전략상 유용한 결과가 나온 경우

연관성 규칙 유형

Useful result	Trivial result	Inexplicable result

	Explainable	Actionable
Useful Rules	O	O

(남성,금요일,기저귀) → (맥주)

연관성 규칙 유형

Useful result	**Trivial result**	Inexplicable result

일반적인 상식에 해당되거나 기존에 시행된
마케팅 전략에 의해 연관성이 높게 나온 경우

연관성 규칙 유형

Useful result	Trivial result	Inexplicable result

Explainable	Actionable

Trivial Rules	O (common sense)	X (result of previous marketing action)

(페인트) → (페인트붓)

(맥주) → (아이스박스)

연관성 규칙 유형

Useful result | Trivial result | **Inexplicable result**

의미를 발견하기 위해 많은 고민이 필요한 경우

연관성 규칙 유형

	Useful result	Trivial result	Inexplicable result
	Explainable	Actionable	
Inexplicable Rules	X	X	

(기초화장품) → (자동차와이퍼)

(강남점,케일) → (보석)

연관성 규칙 유형

	Explainable	Actionable
Useful Rules	O	O
Trivial Rules	O (common sense)	X (result of previous marketing action)
Inexplicable Rules	X	X

연관성 규칙 측도

- 지지도(support)
- 신뢰도(confidence)
- 향상도(lift)

연관성 규칙 측도

지지도(support) | 신뢰도(confidence) | 향상도(lift)

- 전체 거래 중 항목 A와 B를 동시에 포함하는 거래의 비율
- 항목 A와 B를 동시에 살 확률

$$\text{지지도} = P(A \cap B) = \frac{\text{A와 B가 동시에 포함된 거래수}}{\text{전체 거래수}} = \frac{A \cap B}{\text{전체}}$$

연관성 규칙 측도

지지도(support) | **신뢰도(confidence)** | 향상도(lift)

- 항목 A를 포함한 거래 중에서 항목 A와 B가 같이 포함된 거래의 비율
- 항목 A를 샀을 때 항목 B를 같이 살 확률

$$\text{신뢰도} = \frac{P(A \cap B)}{P(A)} = \frac{\text{A와 B가 동시에 포함된 거래수}}{\text{A를 포함하는 거래수}} = \frac{\text{지지도}}{P(A)}$$

연관성 규칙 측도

지지도(support) | 신뢰도(confidence) | **향상도(lift)**

- 항목 B만 샀을 때 대비 항목 A를 구매하고 항목 B를 구매할 확률의 증가비

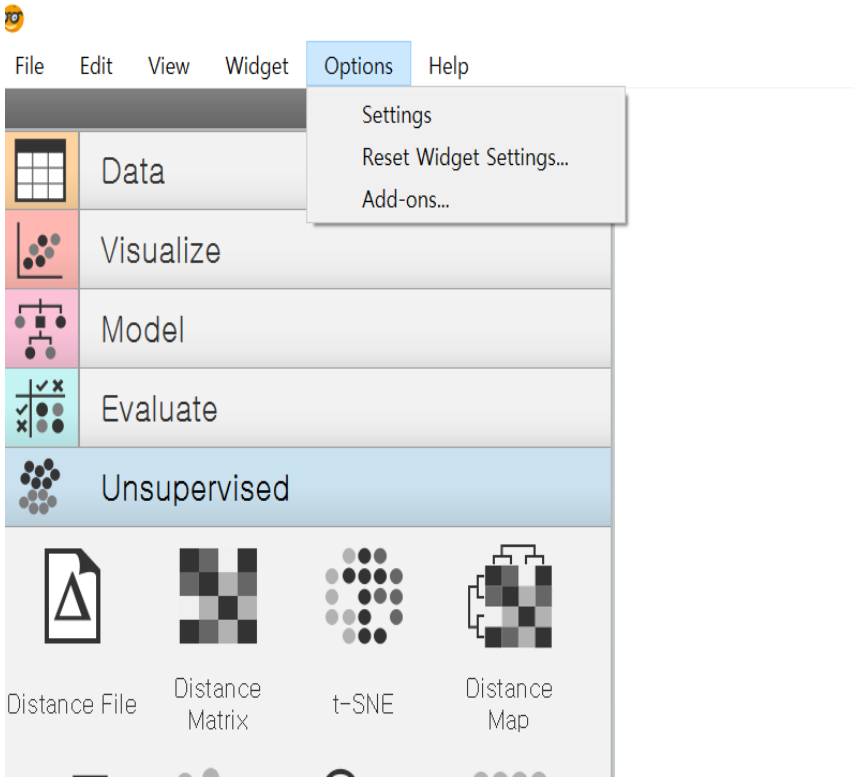
$$\text{향상도} = \frac{P(A|B)}{P(B)} = \frac{P(A \cap B)}{P(A)P(B)} = \frac{\text{A와 B를 동시에 포함하는 거래수}}{\text{A를 포함하는 거래수} \times \text{B를 포함하는 거래수}} = \frac{\text{신뢰도}}{P(B)}$$

연관성 규칙 측도

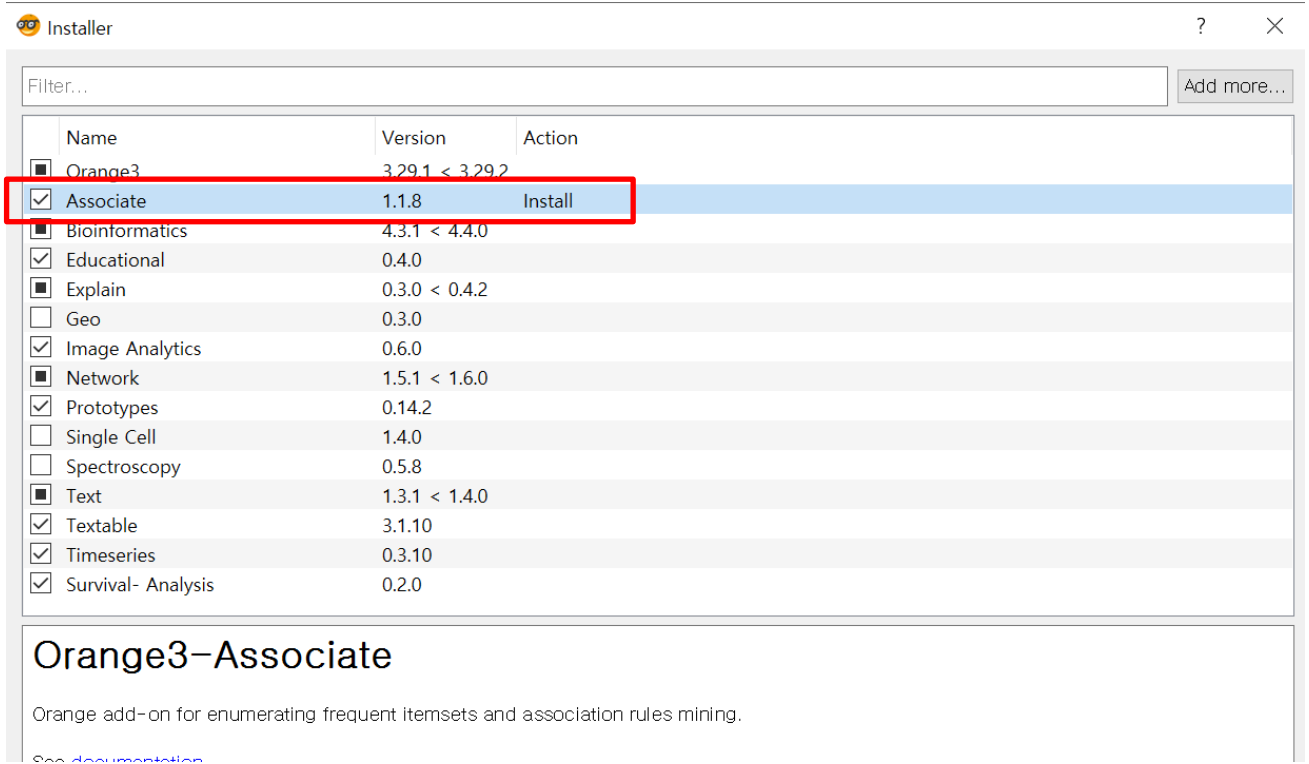
- 신뢰도의 값이 크면 좋지만 신뢰도가 크다고 최선의 연관성 규칙이라고 볼 수는 없음
- 신뢰도와 지지도는 자주 구매되는 항목에 대해서는 연관성 때문이 아니라 우연하게 높게 나올 수도 있음
- 신뢰도가 높을 경우에는 A→B에서 항목 B의 확률이 커야 이 연관성 규칙이 의미를 가짐

연관분석 위젯 알아보기

연관분석 위젯 설치하기


File
Edit
View
Widget
Options
Help
Settings
Reset Widget Settings...
Add-ons...
Data
Visualize
Model
Evaluate
Unsupervised
Distance File
Distance Matrix
t-SNE
Distance Map

연관분석 위젯 설치하기


Installer
Filter...
Add more...
Name
Version
Action
Orange3
3.29.1 < 3.29.2
Associate
1.1.8
Install
Bioinformatics
4.3.1 < 4.4.0
Educational
0.4.0
Explain
0.3.0 < 0.4.2
Geo
0.3.0
Image Analytics
0.6.0
Network
1.5.1 < 1.6.0
Prototypes
0.14.2
Single Cell
1.4.0
Spectroscopy
0.5.8
Text
1.3.1 < 1.4.0
Textable
3.1.10
Timeseries
0.3.10
Survival- Analysis
0.2.0
Orange3-Associate
Orange add-on for enumerating frequent itemsets and association rules mining.

연관분석 위젯 설치하기

연관분석 위젯 알아보기

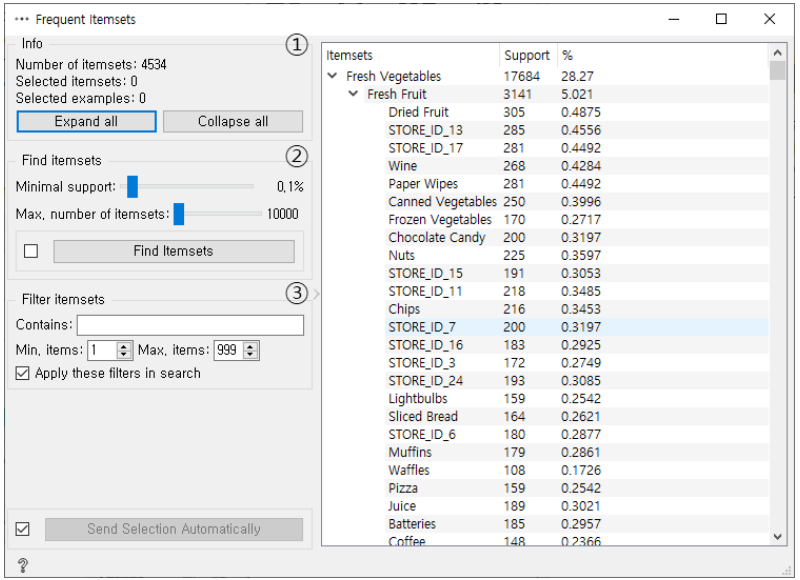
Frequent Itemsets | Association Rules

위젯	설명	입력	출력
Frequent Itemsets	자주 함께 나타나는 항목 집합을 탐색한다.	Data	Matching Data

- Frequent Itemsets 위젯을 사용하면 규칙에 대한 지원 척도를 기준으로 데이터 세트에서 빈 항목을 찾는다.

연관분석 위젯 알아보기

Frequent Itemsets | Association Rules



①	데이터 세트에 대한 정보이다. '모두 확장'은 빈번한 항목 집합 트리를 확장하는 반면, '모두 축소'는 해당 트리를 축소한다.
②	다음에 의해 항목 집합 찾기에서는 항목 집합 검색에 대한 기준을 설정할 수 있다. Minimal support: 생성할 항목 집합을 지원(포함)해야 하는 데이터 인스턴스의 최소 비율이다. 대규모 데이터 세트의 경우 최소 지원(예: 2%-0.01%)을 낮게 설정하는 것이 일반적이다. Max. number of itemsets: 생성된 항목 세트의 상승 양을 제한한다. 항목 집합은 특정 순서로 생성되지 않는다. 항목 세트 자동 찾기가 켜져 있으면 매개 변수를 변경할 때마다 위젯이 검색을 실행한다. 큰 데이터 세트의 경우 속도가 느릴 수 있으므로 매개 변수가 설정된 경우에만 항목 세트 찾기를 누르는 것이 좋다.
③	Filter itemsets: 특정 항목 또는 항목 집합을 찾는 경우 정규식을 기준으로 결과를 필터링한다. 정규식을 쉼표로 구분하여 둘 이상의 단어로 필터링한다. Contains: 정규식을 기준으로 항목 집합을 필터링한다. Min. items: 항목 집합에 표시되어야 하는 최소 항목 수 1인 경우 모든 항목 세트가 표시된다. 예를 들어, 4로 늘리면 4개 이상의 항목이 있는 항목 집합만 표시된다. Max. items: 항목 집합에 표시할 최대 항목 수이다. 예를 들어, 항목이 5개 미만인 항목 집합만 찾으려면 이 매개변수를 5로 설정한다. 검색에 이 필터 적용을 선택하면 위젯이 실시간으로 결과를 필터링한다. 큰 데이터 세트에는 체크 표시를 하지 않는 것이 좋다.

연관분석 위젯 알아보기

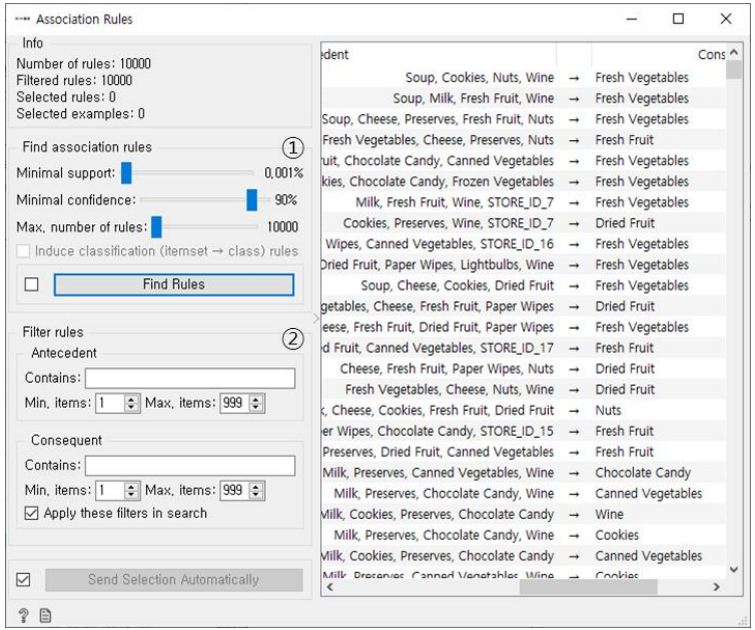
Frequent Itemsets | **Association Rules**

위젯	설명	입력	출력
Association Rules	데이터에서 연결 규칙을 유도한다.	Data	Matching Data, Rules

- 적은 항목의 조건부 데이터베이스에 대해 버킷 최적화를 사용하여 FP 성장 빈도 패턴 마이닝 알고리즘을 구현
- 분류 규칙을 유도하기 위해 전체 항목 집합에 대한 규칙을 생성하고 결과가 클래스 값 중 하나와 일치하지 않는 규칙은 건너뛴다.

연관분석 위젯 알아보기

Frequent Itemsets | **Association Rules**

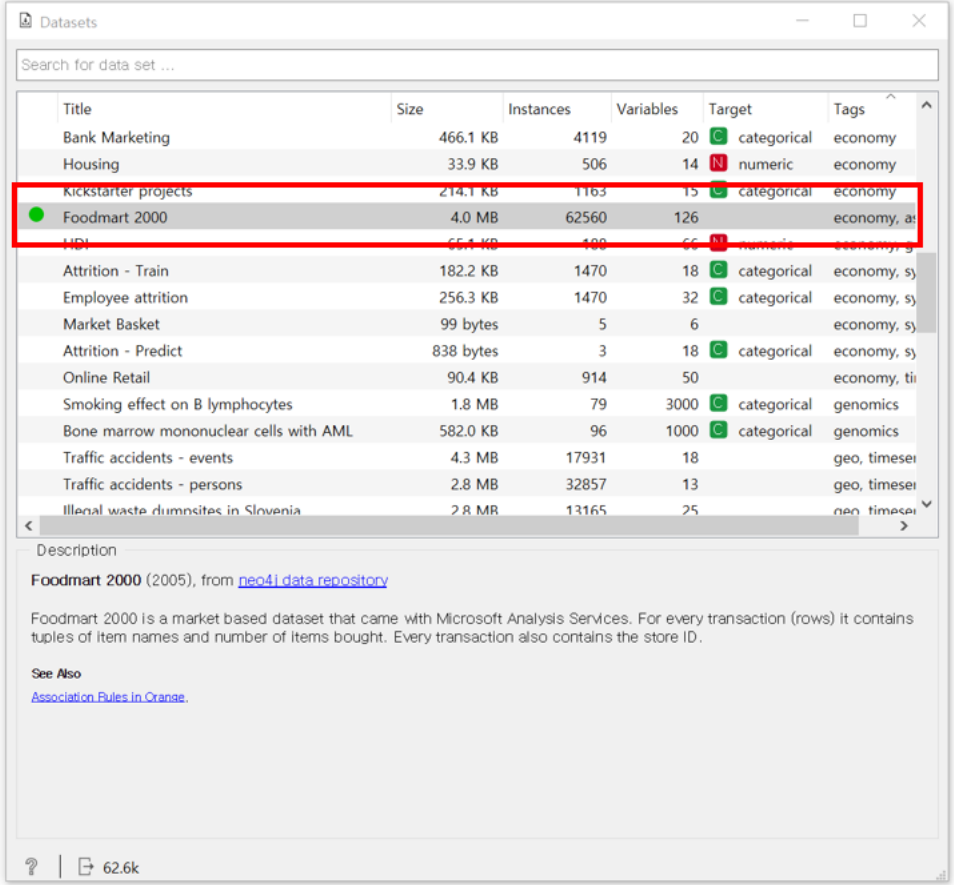


①	n 연결 규칙 찾기 규칙 유도 기준을 설정할 수 있다. Minimal support: 전체 규칙에서 적용되는 전체 데이터 세트의 백분율(이전 및 후속) Minimal confidence: 왼쪽(선행)에 맞는 예제 중 오른쪽(후행)에 맞는 예제 수의 비율 Max. number of rules: 알고리즘이 생성하는 규칙 수를 제한한다. 규칙이 너무 많으면 위젯 속도가 상당히 느려질 수 있다. 분류 유도(하위 집합 → 클래스) 규칙이 선택된 경우 위젯은 규칙의 오른쪽(하위)에 클래스 값이 있는 규칙만 생성한다. 자동 찾기 규칙이 켜져 있으면 매개 변수를 변경할 때마다 위젯이 검색을 실행한다. 속성이 많은 데이터 세트의 경우 속도가 느릴 수 있으므로 매개 변수가 설정된 경우에만 규칙 찾기를 누르는 것이 좋다.
②	규칙 필터링 기준 Antecedent Contains: 이전 항목의 공백으로 구분된 정규식을 일치시켜 규칙을 필터링한다. Min. items: 선행 항목에 나타나야 하는 최소 항목 수. Max. items: 선행 항목에 나타날 수 있는 최대 항목 수 Consequent Contains: 결과 항목의 공백으로 구분된 정규식을 일치시켜 규칙을 필터링한다. Min. items: 결과에 표시되어야 하는 최소 항목 수 Max. items: 결과에 나타날 수 있는 최대 항목 수 Apply these filters in search: 필터와 일치하는 규칙으로만 규칙 생성을 제한한다. 이 옵션을 선택하지 않으면 모든 규칙이 생성되지만 일치하는 규칙만 표시된다.

연관분석의 실제

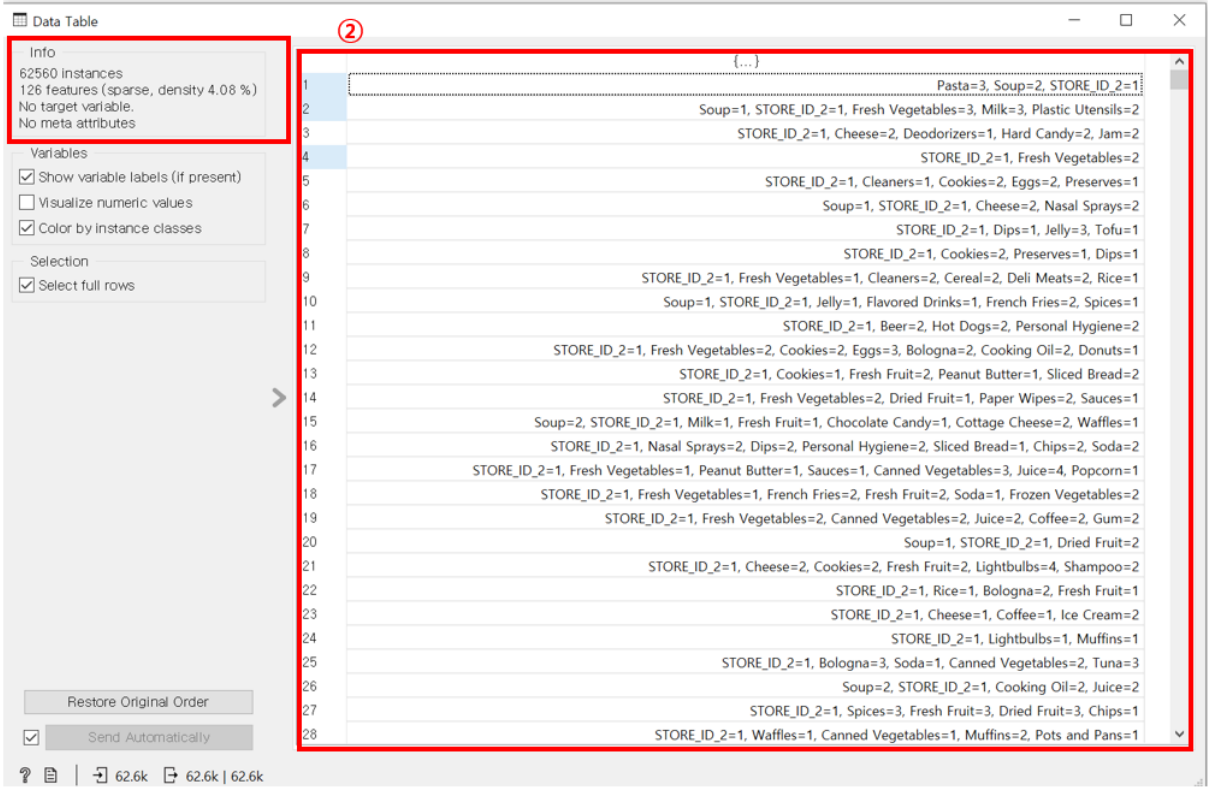
연관 분석 해보기

Datasets

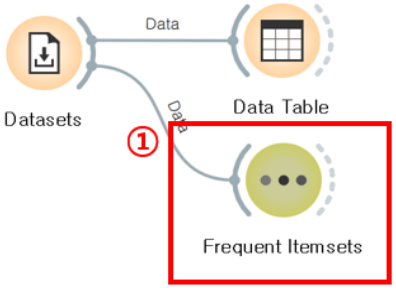


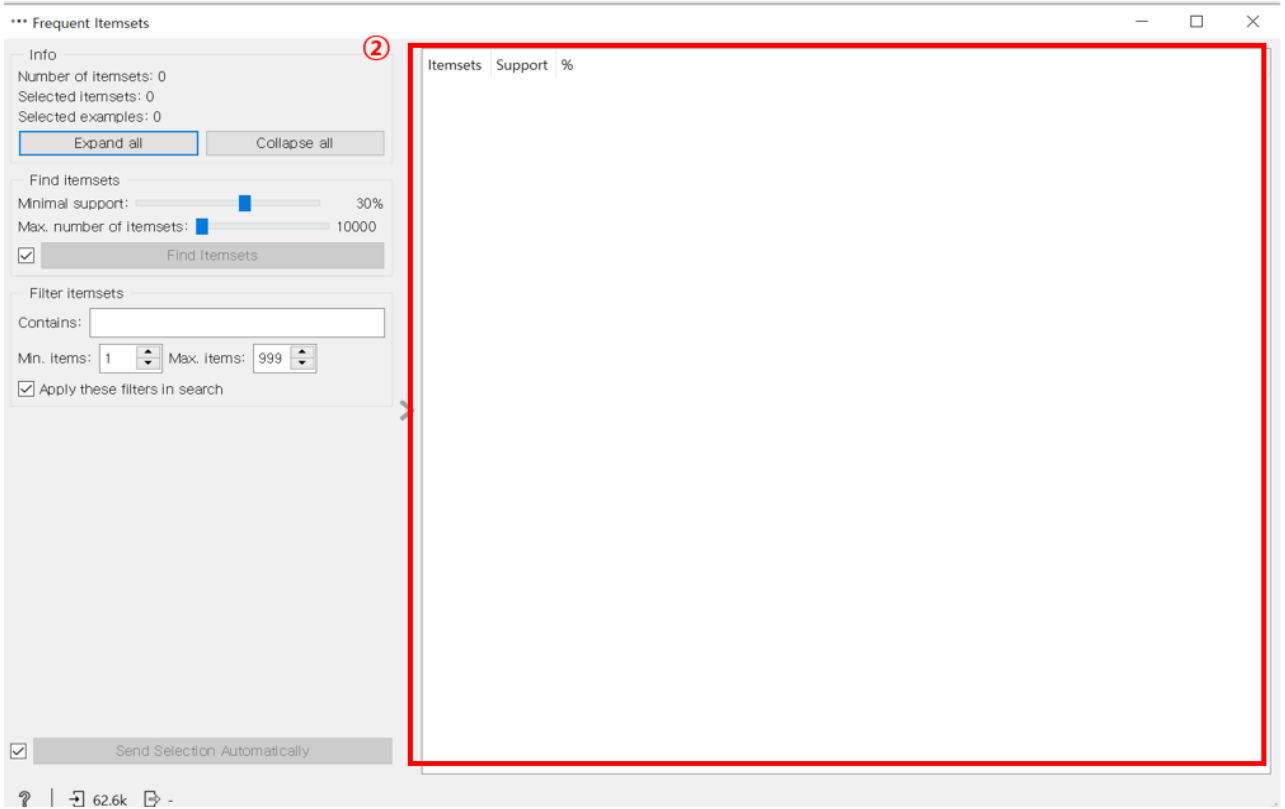
연관 분석 해보기

Data
Datasets
Data Table

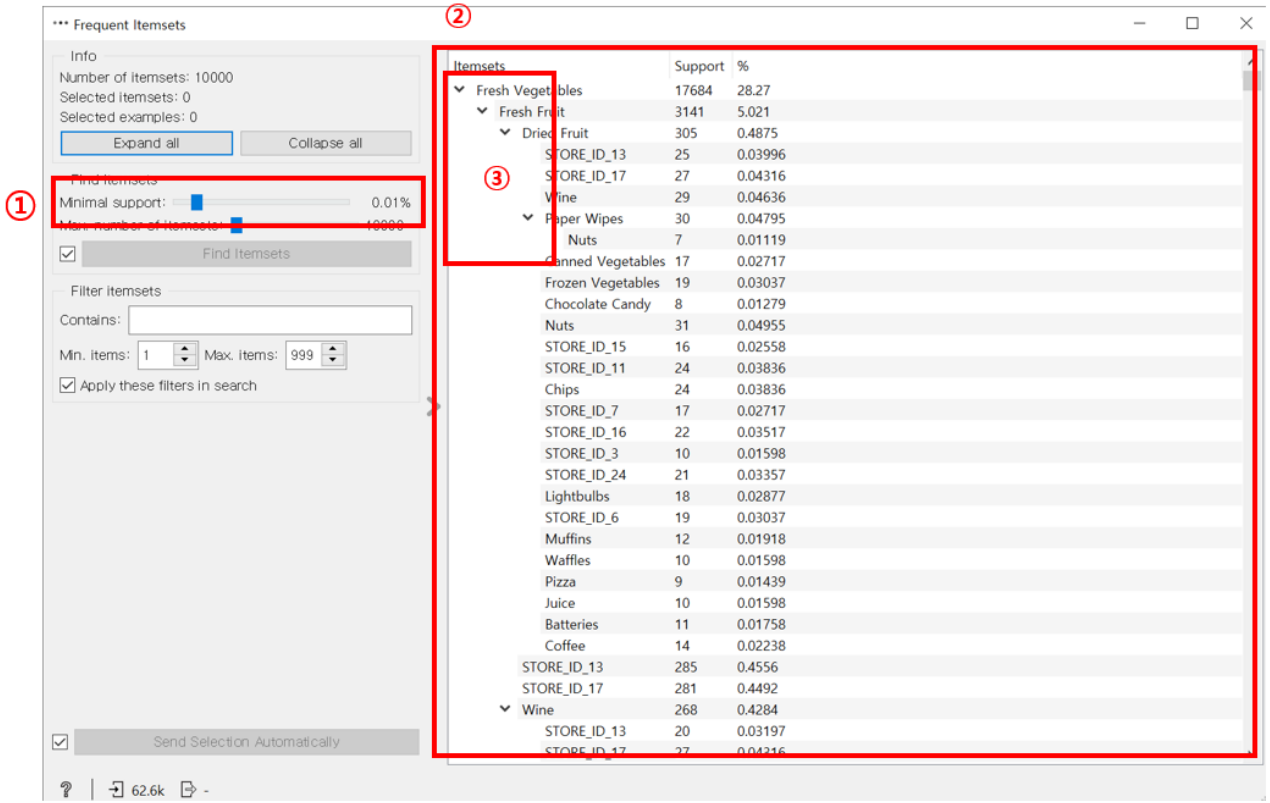

Data Table
①
②
Info
62560 instances
126 features (sparse, density 4.08 %)
No target variable.
No meta attributes
Variables
Show variable labels (if present)
Visualize numeric values
Color by instance classes
Selection
Select full rows
Restore Original Order
Send Automatically
62.6k
62.6k | 62.6k
{...}
1 Pasta=3, Soup=2, STORE_ID_2=1
2 Soup=1, STORE_ID_2=1, Fresh Vegetables=3, Milk=3, Plastic Utensils=2
3 STORE_ID_2=1, Cheese=2, Deodorizers=1, Hard Candy=2, Jam=2
4 STORE_ID_2=1, Fresh Vegetables=2
5 STORE_ID_2=1, Cleaners=1, Cookies=2, Eggs=2, Preserves=1
6 Soup=1, STORE_ID_2=1, Cheese=2, Nasal Sprays=2
7 STORE_ID_2=1, Dips=1, Jelly=3, Tofu=1
8 STORE_ID_2=1, Cookies=2, Preserves=1, Dips=1
9 STORE_ID_2=1, Fresh Vegetables=1, Cleaners=2, Cereal=2, Deli Meats=2, Rice=1
10 Soup=1, STORE_ID_2=1, Jelly=1, Flavored Drinks=1, French Fries=2, Spices=1
11 STORE_ID_2=1, Beer=2, Hot Dogs=2, Personal Hygiene=2
12 STORE_ID_2=1, Fresh Vegetables=2, Cookies=2, Eggs=3, Bologna=2, Cooking Oil=2, Donuts=1
13 STORE_ID_2=1, Cookies=1, Fresh Fruit=2, Peanut Butter=1, Sliced Bread=2
14 STORE_ID_2=1, Fresh Vegetables=2, Dried Fruit=1, Paper Wipes=2, Sauces=1
15 Soup=2, STORE_ID_2=1, Milk=1, Fresh Fruit=1, Chocolate Candy=1, Cottage Cheese=2, Waffles=1
16 STORE_ID_2=1, Nasal Sprays=2, Dips=2, Personal Hygiene=2, Sliced Bread=1, Chips=2, Soda=2
17 STORE_ID_2=1, Fresh Vegetables=1, Peanut Butter=1, Sauces=1, Canned Vegetables=3, Juice=4, Popcorn=1
18 STORE_ID_2=1, Fresh Vegetables=1, French Fries=2, Fresh Fruit=2, Soda=1, Frozen Vegetables=2
19 STORE_ID_2=1, Fresh Vegetables=2, Canned Vegetables=2, Juice=2, Coffee=2, Gum=2
20 Soup=1, STORE_ID_2=1, Dried Fruit=2
21 STORE_ID_2=1, Cheese=2, Cookies=2, Fresh Fruit=2, Lightbulbs=4, Shampoo=2
22 STORE_ID_2=1, Rice=1, Bologna=2, Fresh Fruit=1
23 STORE_ID_2=1, Cheese=1, Coffee=1, Ice Cream=2
24 STORE_ID_2=1, Lightbulbs=1, Muffins=1
25 STORE_ID_2=1, Bologna=3, Soda=1, Canned Vegetables=2, Tuna=3
26 Soup=2, STORE_ID_2=1, Cooking Oil=2, Juice=2
27 STORE_ID_2=1, Spices=3, Fresh Fruit=3, Dried Fruit=3, Chips=1
28 STORE_ID_2=1, Waffles=1, Canned Vegetables=1, Muffins=2, Pots and Pans=1

연관 분석 해보기

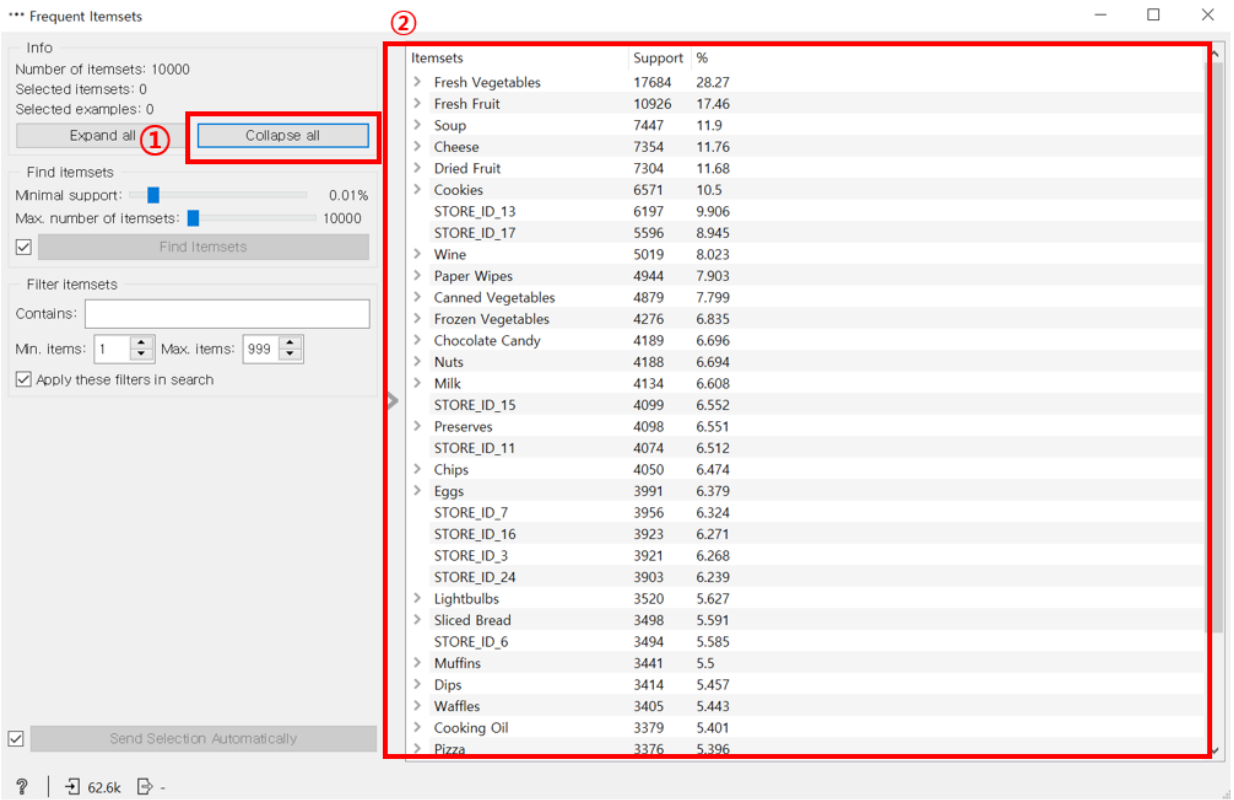

Data
Datasets
Data Table
Data
①
Frequent Itemsets


Frequent Itemsets
②
Info
Number of itemsets: 0
Selected itemsets: 0
Selected examples: 0
Expand all
Collapse all
Find itemsets
Minimal support:
30%
Max. number of itemsets:
10000
Find Itemsets
Filter itemsets
Contains:
Min. items:
1
Max. items:
999
Apply these filters in search
Send Selection Automatically
Itemsets
Support
%
62.6k

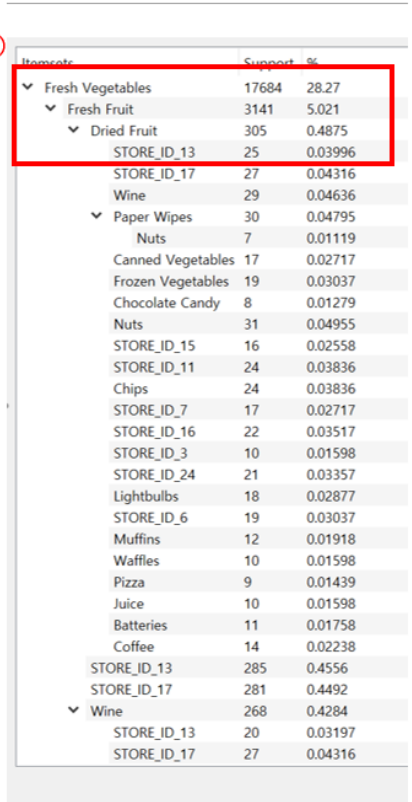
연관 분석 해보기


Frequent Itemsets
②
Info
Number of itemsets: 10000
Selected itemsets: 0
Selected examples: 0
Expand all
Collapse all
①
Minimal support:
0.01%
Find Itemsets
Filter itemsets
Contains:
Min. items: 1
Max. items: 999
Apply these filters in search
Send Selection Automatically
62.6k
③
Itemsets
Support
%
Fresh Vegetables 17684 28.27
Fresh Fruit 3141 5.021
Dried Fruit 305 0.4875
STORE_ID_13 25 0.03996
STORE_ID_17 27 0.04316
Wine 29 0.04636
Paper Wipes 30 0.04795
Nuts 7 0.01119
Canned Vegetables 17 0.02717
Frozen Vegetables 19 0.03037
Chocolate Candy 8 0.01279
Nuts 31 0.04955
STORE_ID_15 16 0.02558
STORE_ID_11 24 0.03836
Chips 24 0.03836
STORE_ID_7 17 0.02717
STORE_ID_16 22 0.03517
STORE_ID_3 10 0.01598
STORE_ID_24 21 0.03357
Lightbulbs 18 0.02877
STORE_ID_6 19 0.03037
Muffins 12 0.01918
Waffles 10 0.01598
Pizza 9 0.01439
Juice 10 0.01598
Batteries 11 0.01758
Coffee 14 0.02238
STORE_ID_13 285 0.4556
STORE_ID_17 281 0.4492
Wine 268 0.4284
STORE_ID_13 20 0.03197

연관 분석 해보기

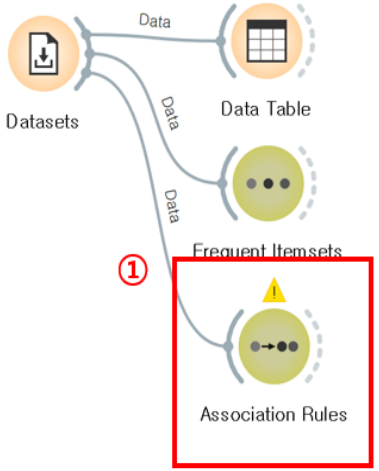


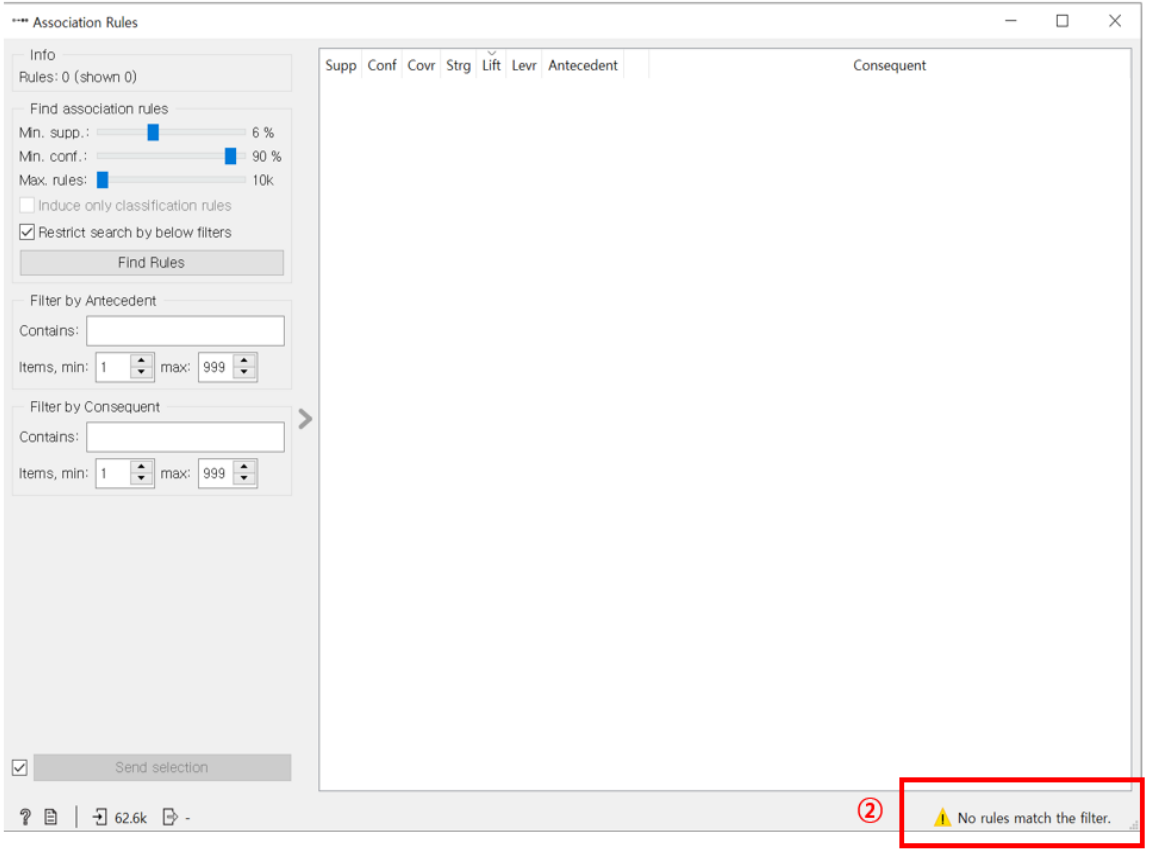
Itemsets	Support	%
Fresh Vegetables	17684	28.27
Fresh Fruit	10926	17.46
Soup	7447	11.9
Cheese	7354	11.76
Dried Fruit	7304	11.68
Cookies	6571	10.5
STORE_ID_13	6197	9.906
STORE_ID_17	5596	8.945
Wine	5019	8.023
Paper Wipes	4944	7.903
Canned Vegetables	4879	7.799
Frozen Vegetables	4276	6.835
Chocolate Candy	4189	6.696
Nuts	4188	6.694
Milk	4134	6.608
STORE_ID_15	4099	6.552
Preserves	4098	6.551
STORE_ID_11	4074	6.512
Chips	4050	6.474
Eggs	3991	6.379
STORE_ID_7	3956	6.324
STORE_ID_16	3923	6.271
STORE_ID_3	3921	6.268
STORE_ID_24	3903	6.239
Lightbulbs	3520	5.627
Sliced Bread	3498	5.591
STORE_ID_6	3494	5.585
Muffins	3441	5.5
Dips	3414	5.457
Waffles	3405	5.443
Cooking Oil	3379	5.401
Pizza	3376	5.396



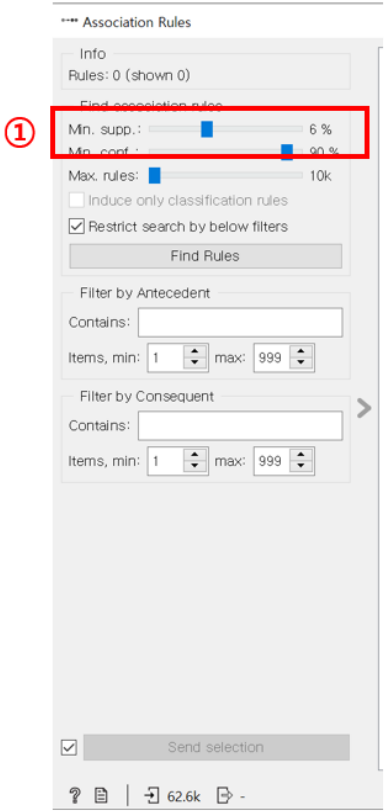
Itemsets	Support	%
Fresh Vegetables	17684	28.27
Fresh Fruit	3141	5.021
Dried Fruit	305	0.4875
STORE_ID_13	25	0.03996
STORE_ID_17	27	0.04316
Wine	29	0.04636
Paper Wipes	30	0.04795
Nuts	7	0.01119
Canned Vegetables	17	0.02717
Frozen Vegetables	19	0.03037
Chocolate Candy	8	0.01279
Nuts	31	0.04955
STORE_ID_15	16	0.02558
STORE_ID_11	24	0.03836
Chips	24	0.03836
STORE_ID_7	17	0.02717
STORE_ID_16	22	0.03517
STORE_ID_3	10	0.01598
STORE_ID_24	21	0.03357
Lightbulbs	18	0.02877
STORE_ID_6	19	0.03037
Muffins	12	0.01918
Waffles	10	0.01598
Pizza	9	0.01439
Juice	10	0.01598
Batteries	11	0.01758
Coffee	14	0.02238
STORE_ID_13	285	0.4556
STORE_ID_17	281	0.4492
Wine	268	0.4284
STORE_ID_13	20	0.03197
STORE_ID_17	27	0.04316

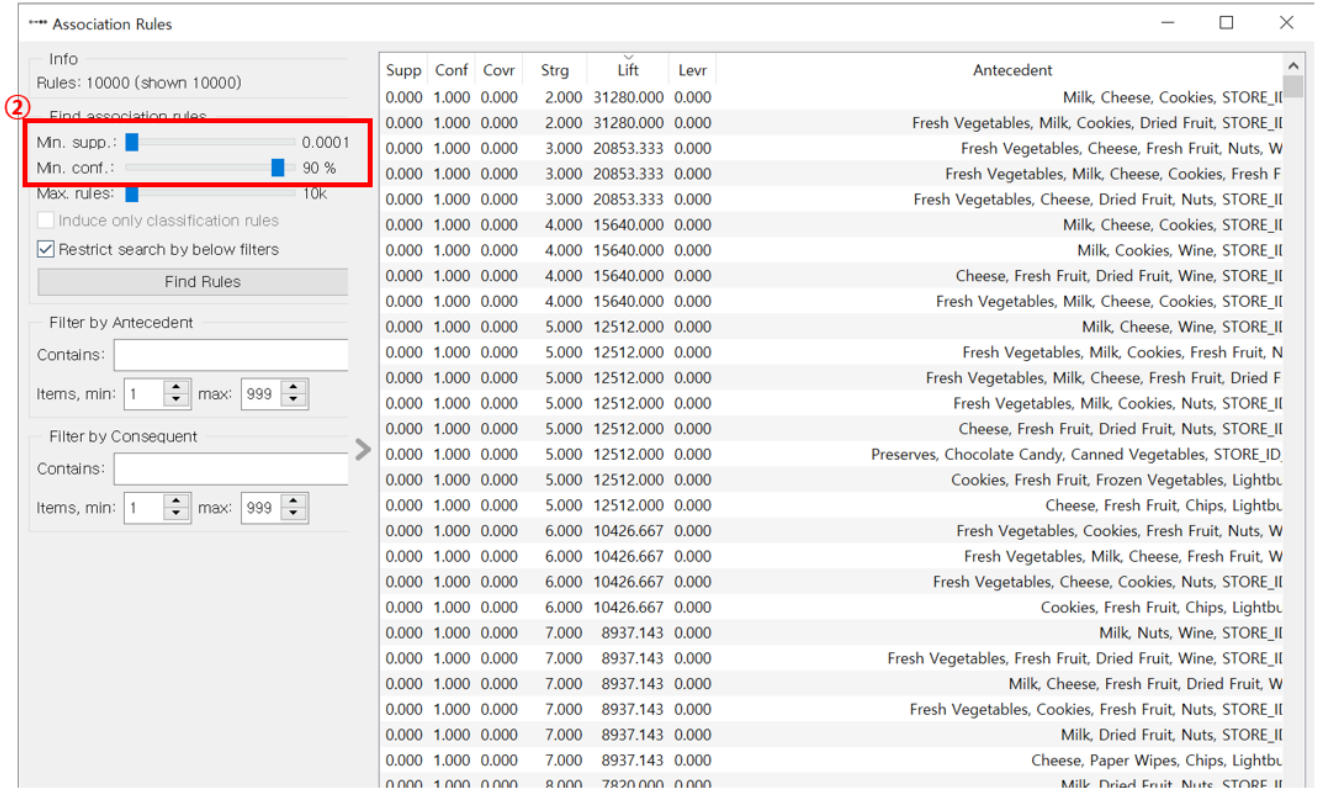
연관 분석 해보기


Datasets
Data
Data Table
Data
Frequent Itemsets
Data
①
Association Rules


Association Rules
Info
Rules: 0 (shown 0)
Find association rules
Min. supp.: 6 %
Min. conf.: 90 %
Max. rules: 10k
Induce only classification rules
Restrict search by below filters
Find Rules
Filter by Antecedent
Contains:
Items, min: 1 max: 999
Filter by Consequent
Contains:
Items, min: 1 max: 999
Send selection
Supp Conf Covr Strg Lift Levr Antecedent Consequent
62.6k
②
No rules match the filter.

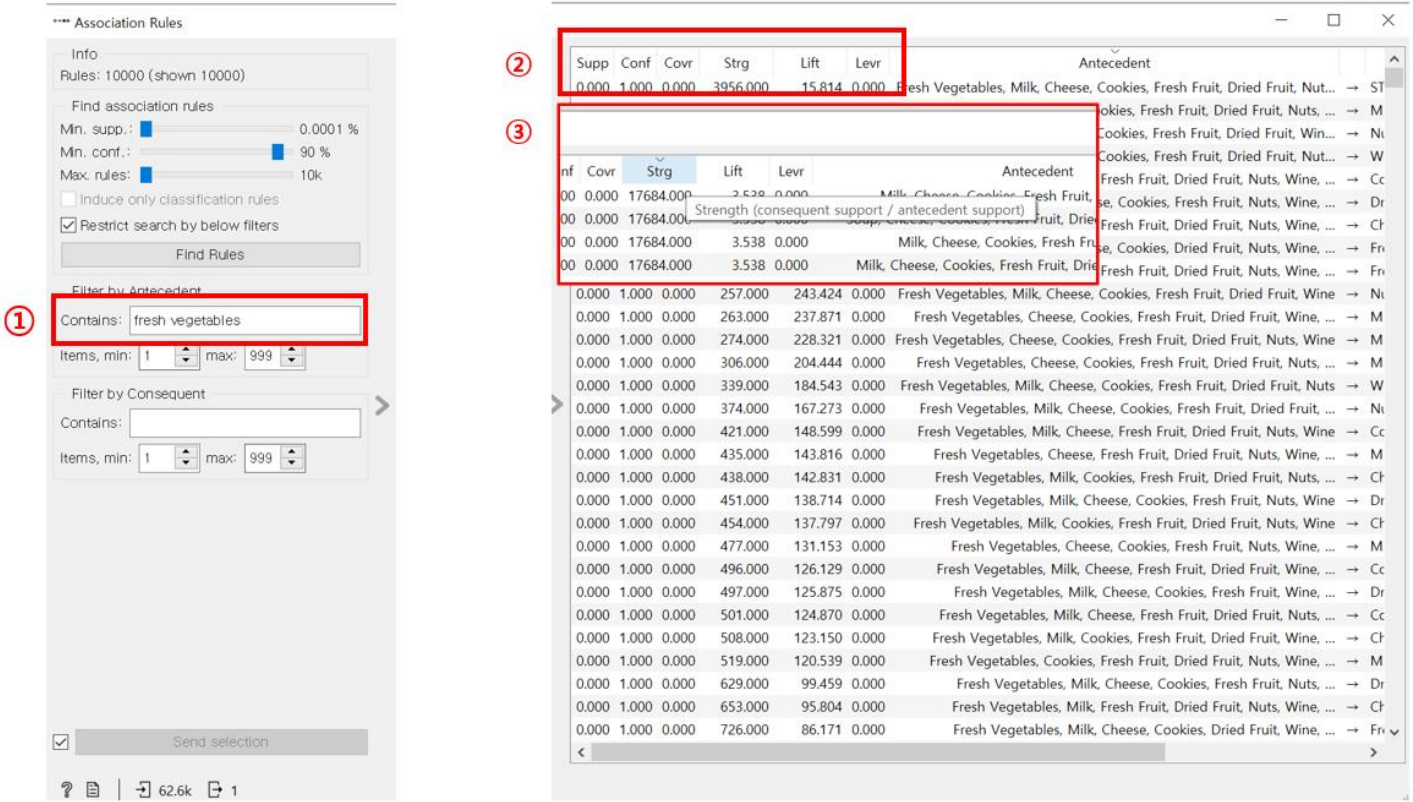
연관 분석 해보기





Supp	Conf	Covr	Strg	Lift	Levr	Antecedent
0.000	1.000	0.000	2.000	31280.000	0.000	Milk, Cheese, Cookies, STORE_II
0.000	1.000	0.000	2.000	31280.000	0.000	Fresh Vegetables, Milk, Cookies, Dried Fruit, STORE_II
0.000	1.000	0.000	3.000	20853.333	0.000	Fresh Vegetables, Cheese, Fresh Fruit, Nuts, W
0.000	1.000	0.000	3.000	20853.333	0.000	Fresh Vegetables, Milk, Cheese, Cookies, Fresh F
0.000	1.000	0.000	3.000	20853.333	0.000	Fresh Vegetables, Cheese, Dried Fruit, Nuts, STORE_II
0.000	1.000	0.000	4.000	15640.000	0.000	Milk, Cheese, Cookies, STORE_II
0.000	1.000	0.000	4.000	15640.000	0.000	Milk, Cookies, Wine, STORE_II
0.000	1.000	0.000	4.000	15640.000	0.000	Cheese, Fresh Fruit, Dried Fruit, Wine, STORE_II
0.000	1.000	0.000	4.000	15640.000	0.000	Fresh Vegetables, Milk, Cheese, Cookies, STORE_II
0.000	1.000	0.000	5.000	12512.000	0.000	Milk, Cheese, Wine, STORE_II
0.000	1.000	0.000	5.000	12512.000	0.000	Fresh Vegetables, Milk, Cookies, Fresh Fruit, N
0.000	1.000	0.000	5.000	12512.000	0.000	Fresh Vegetables, Milk, Cheese, Fresh Fruit, Dried F
0.000	1.000	0.000	5.000	12512.000	0.000	Fresh Vegetables, Milk, Cookies, Nuts, STORE_II
0.000	1.000	0.000	5.000	12512.000	0.000	Cheese, Fresh Fruit, Dried Fruit, Nuts, STORE_II
0.000	1.000	0.000	5.000	12512.000	0.000	Preserves, Chocolate Candy, Canned Vegetables, STORE_ID
0.000	1.000	0.000	5.000	12512.000	0.000	Cookies, Fresh Fruit, Frozen Vegetables, Lightbu
0.000	1.000	0.000	5.000	12512.000	0.000	Cheese, Fresh Fruit, Chips, Lightbu
0.000	1.000	0.000	6.000	10426.667	0.000	Fresh Vegetables, Cookies, Fresh Fruit, Nuts, W
0.000	1.000	0.000	6.000	10426.667	0.000	Fresh Vegetables, Milk, Cheese, Fresh Fruit, W
0.000	1.000	0.000	6.000	10426.667	0.000	Fresh Vegetables, Cheese, Cookies, Nuts, STORE_II
0.000	1.000	0.000	6.000	10426.667	0.000	Cookies, Fresh Fruit, Chips, Lightbu
0.000	1.000	0.000	7.000	8937.143	0.000	Milk, Nuts, Wine, STORE_II
0.000	1.000	0.000	7.000	8937.143	0.000	Fresh Vegetables, Fresh Fruit, Dried Fruit, Wine, STORE_II
0.000	1.000	0.000	7.000	8937.143	0.000	Milk, Cheese, Fresh Fruit, Dried Fruit, W
0.000	1.000	0.000	7.000	8937.143	0.000	Fresh Vegetables, Cookies, Fresh Fruit, Nuts, STORE_II
0.000	1.000	0.000	7.000	8937.143	0.000	Milk, Dried Fruit, Nuts, STORE_II
0.000	1.000	0.000	7.000	8937.143	0.000	Cheese, Paper Wipes, Chips, Lightbu
0.000	1.000	0.000	8.000	7820.000	0.000	Milk, Dried Fruit, Nuts, STORE_II

연관 분석 해보기


Association Rules
Info
Rules: 10000 (shown 10000)
Find association rules
Min. supp.: 0.0001 %
Min. conf.: 90 %
Max. rules: 10k
Induce only classification rules
Restrict search by below filters
Find Rules
Filter by Antecedent
①
Contains: fresh vegetables
Items, min: 1 max: 999
Filter by Consequent
Contains:
Items, min: 1 max: 999
Send selection
62.6k 1
②
③
Strength (consequent support / antecedent support)

연관 분석 해보기

① Antecedent		② Consequent
Frozen Vegetables, Shampoo, Candles	→	Gum
Plastic Utensils, Fresh Fruit, Ibuprofen	→	Cold Remedies
Fresh Vegetables, Cleaners, Pretzels	→	Jam
Fresh Fruit, Tools, Sports Magazines	→	Coffee
Pasta, Canned Vegetables, Pretzels	→	Dips
Bologna, Fresh Fruit, Yogurt	→	Chocolate Candy
Fresh Fruit, Bagels, Acetominifen	→	Wine
Frozen Vegetables, Coffee, Aspirin	→	Cookies
Soup, Frozen Vegetables, Auto Magazines	→	Cookies
Fresh Vegetables, Deodorizers, Pots and Pans	→	Dried Fruit
Flavored Drinks, Spices, Canned Vegetables	→	Cheese
Deli Meats, Frozen Vegetables, Coffee	→	Cookies
Frozen Vegetables, Tuna, Pancakes	→	Cheese
Hot Dogs, Canned Vegetables, Dried Meat	→	Dried Fruit
Frozen Vegetables, Coffee, Computer Magazines	→	Soup
Plastic Utensils, Frozen Vegetables, Pot Cleaners	→	Fresh Fruit

질문 있나요?

hsryu13@hongik.ac.kr