

08

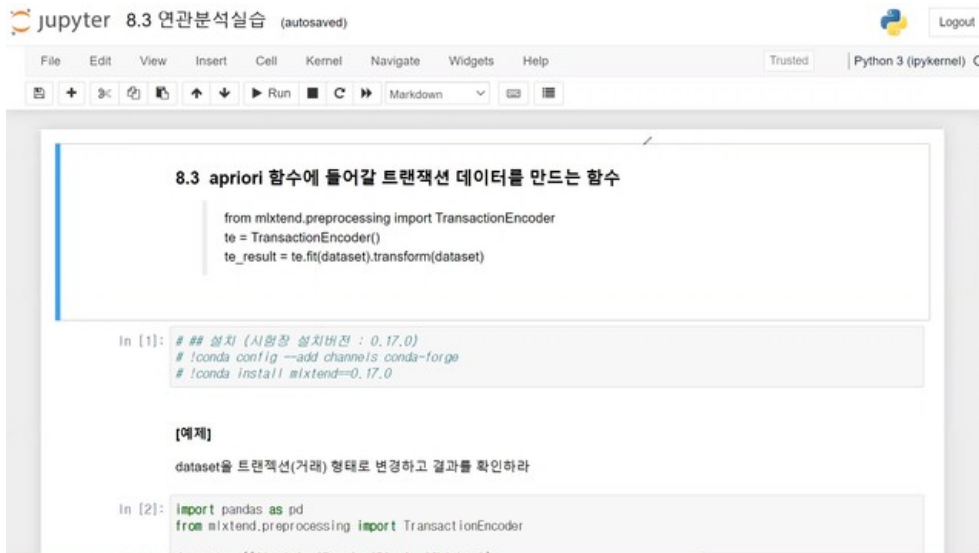
연관분석 실습

[거래형식으로 전처리 하기]

장바구니 분석을 할 때에는, 거래형식으로 인코딩을 해줘야 합니다.

- 전처리가 매우 중요합니다!

01:56



[Apriori]

Apriori 함수를 만드는 방법, 최소지지도의 개념을 기억해주세요~

05:00

jupyter 8.3 연관분석실습 (autosaved)

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (ipykernel)

Run

4	False	False	True	True	True	True
5	False	False	True	False	True	True
6	False	False	True	False	True	False
7	True	True	False	False	False	False

8.3.0.0.1 mixtend의 apriori 함수

One-Hot 형식의 DataFrame에서 빈번 항목 집합을 출력하는 함수

```
from mixtend.frequent_patterns import apriori
apriori(df)
```

- df : 값이 0/1 또는 True / False인 pandas DataFrame
- min_support : 최소 지지도 (default = 0.5)
- use_colnames : True이면 열 인덱스 대신 반환된 DataFrame의 열 이름을 사용 (default = False)
- max_len : 생성된 항목 세트의 최대 길이로 None인 경우 가능한 모든 항목 세트 길이를 평가 (default = None)
- low_memory : 메모리 리소스가 제한될 경우에 사용할 수 있으나 속도가 기존의 3-6배 느림 (default = False)

[연관 규칙 분석 개념]

연관 규칙을 찾는 것이 저희의 목적입니다!,

연관 규칙을 발견하고 해석할 수 있어야 합니다.

13:07

jupyter 8.3 연관분석실습 (unsaved changes)

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (ipykernel)

Code

24	0.104931	(tropical fruit)
25	0.071683	(whipped/sour cream)
26	0.255516	(whole milk)
27	0.139502	(yogurt)
28	0.074835	(other vegetables, whole milk)
29	0.056634	(rolls/buns, whole milk)
30	0.056024	(yogurt, whole milk)

8.3.0.0.2 mixtend의 association_rules 함수

'score(점수)', 'confidence(신뢰도)' 및 'lift(상승도)'를 포함하는 연관 규칙의 DataFrame 생성

```
from mixtend.frequent_patterns import association_rules
association_rules(df, metric='confidence', min_threshold=0.8, support_only=False)
```

- df : ['support', 'itemsets'] 열이 있는 빈번 항목 집합의 DataFrame
- metric : ('support', 'confidence', 'lift')
support_only = True 인 경우 자동으로 'support'로 설정됨
- min_threshold : metric에 지정된 rule의 최소값 (default: 0.8)