Market Basket Analysis

May 17, 2023

[6]:
```python
import pandas as pd
import numpy as np
```

[7]:
```python
df=pd.read_csv('supermarket.csv')
```

[8]:
```python
df.head(10)
```

[8]:
```
                MILK,BREAD,BISCUIT
0   BREAD,MILK,BISCUIT,CORNFLAKES
1             BREAD,TEA,BOURNVITA
2            JAM,MAGGI,BREAD,MILK
3               MAGGI,TEA,BISCUIT
4             BREAD,TEA,BOURNVITA
5            MAGGI,TEA,CORNFLAKES
6         MAGGI,BREAD,TEA,BISCUIT
7             JAM,MAGGI,BREAD,TEA
8                      BREAD,MILK
9  COFFEE,COCK,BISCUIT,CORNFLAKES
```

[15]:
```python
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori,association_rules
```

[10]:
```python
df1=df["MILK,BREAD,BISCUIT"].apply(lambda t: t.split(","))
```

[11]:
```python
df1
```

[11]:
```
0       [BREAD, MILK, BISCUIT, CORNFLAKES]
1                  [BREAD, TEA, BOURNVITA]
2                [JAM, MAGGI, BREAD, MILK]
3                    [MAGGI, TEA, BISCUIT]
4                  [BREAD, TEA, BOURNVITA]
5                 [MAGGI, TEA, CORNFLAKES]
6             [MAGGI, BREAD, TEA, BISCUIT]
7                 [JAM, MAGGI, BREAD, TEA]
8                            [BREAD, MILK]
9      [COFFEE, COCK, BISCUIT, CORNFLAKES]
10     [COFFEE, COCK, BISCUIT, CORNFLAKES]
11              [COFFEE, SUGER, BOURNVITA]
```

```
12                   [BREAD, COFFEE, COCK]
13                 [BREAD, SUGER, BISCUIT]
14             [COFFEE, SUGER, CORNFLAKES]
15               [BREAD, SUGER, BOURNVITA]
16                  [BREAD, COFFEE, SUGER]
17                  [BREAD, COFFEE, SUGER]
18         [TEA, MILK, COFFEE, CORNFLAKES]
Name: MILK,BREAD,BISCUIT, dtype: object
```

[12]:
```python
df1=list(df1)
df1
```

[12]:
```
[['BREAD', 'MILK', 'BISCUIT', 'CORNFLAKES'],
 ['BREAD', 'TEA', 'BOURNVITA'],
 ['JAM', 'MAGGI', 'BREAD', 'MILK'],
 ['MAGGI', 'TEA', 'BISCUIT'],
 ['BREAD', 'TEA', 'BOURNVITA'],
 ['MAGGI', 'TEA', 'CORNFLAKES'],
 ['MAGGI', 'BREAD', 'TEA', 'BISCUIT'],
 ['JAM', 'MAGGI', 'BREAD', 'TEA'],
 ['BREAD', 'MILK'],
 ['COFFEE', 'COCK', 'BISCUIT', 'CORNFLAKES'],
 ['COFFEE', 'COCK', 'BISCUIT', 'CORNFLAKES'],
 ['COFFEE', 'SUGER', 'BOURNVITA'],
 ['BREAD', 'COFFEE', 'COCK'],
 ['BREAD', 'SUGER', 'BISCUIT'],
 ['COFFEE', 'SUGER', 'CORNFLAKES'],
 ['BREAD', 'SUGER', 'BOURNVITA'],
 ['BREAD', 'COFFEE', 'SUGER'],
 ['BREAD', 'COFFEE', 'SUGER'],
 ['TEA', 'MILK', 'COFFEE', 'CORNFLAKES']]
```

[16]:
```python
te=TransactionEncoder()
```

[17]:
```python
df2=te.fit(df1).transform(df1)
```

[22]:
```python
fte=pd.DataFrame(df2,columns=te.columns_)
```

[23]:
```python
fte
```

[23]:

	BISCUIT	BOURNVITA	BREAD	COCK	COFFEE	CORNFLAKES	JAM	MAGGI	MILK	\
0	True	False	True	False	False	True	False	False	True	
1	False	True	True	False	False	False	False	False	False	
2	False	False	True	False	False	False	True	True	True	
3	True	False	False	False	False	False	False	True	False	
4	False	True	True	False	False	False	False	False	False	
5	False	False	False	False	False	True	False	True	False	

```
6    True   False   True  False  False    False  False   True  False
7   False   False   True  False  False    False   True   True  False
8   False   False   True  False  False    False  False  False   True
9    True   False  False   True   True     True  False  False  False
10   True   False  False   True   True     True  False  False  False
11  False    True  False  False   True    False  False  False  False
12  False   False   True   True   True    False  False  False  False
13   True   False   True  False  False    False  False  False  False
14  False   False  False  False   True     True  False  False  False
15  False    True   True  False  False    False  False  False  False
16  False   False   True  False   True    False  False  False  False
17  False   False   True  False   True    False  False  False  False
18  False   False  False  False   True     True  False  False   True

   SUGER    TEA
0  False  False
1  False   True
2  False  False
3  False   True
4  False   True
5  False   True
6  False   True
7  False   True
8  False  False
9  False  False
10 False  False
11  True  False
12 False  False
13  True  False
14  True  False
15  True  False
16  True  False
17  True  False
18 False   True
```

[66]:
```
freq=apriori(fte,min_support=0.1,use_colnames=True)
```

[67]:
```
freq
```

[67]:
```
     support                    itemsets
0   0.315789                   (BISCUIT)
1   0.210526                 (BOURNVITA)
2   0.631579                     (BREAD)
3   0.157895                      (COCK)
4   0.421053                    (COFFEE)
5   0.315789                (CORNFLAKES)
6   0.105263                       (JAM)
```

```
7   0.263158                                (MAGGI)
8   0.210526                                 (MILK)
9   0.315789                                (SUGER)
10  0.368421                                  (TEA)
11  0.157895                     (BREAD, BISCUIT)
12  0.105263                      (COCK, BISCUIT)
13  0.105263                    (COFFEE, BISCUIT)
14  0.157895                (CORNFLAKES, BISCUIT)
15  0.105263                     (MAGGI, BISCUIT)
16  0.105263                       (TEA, BISCUIT)
17  0.157895                   (BREAD, BOURNVITA)
18  0.105263                   (SUGER, BOURNVITA)
19  0.105263                     (TEA, BOURNVITA)
20  0.157895                      (BREAD, COFFEE)
21  0.105263                         (BREAD, JAM)
22  0.157895                       (MAGGI, BREAD)
23  0.157895                        (BREAD, MILK)
24  0.210526                       (SUGER, BREAD)
25  0.210526                         (TEA, BREAD)
26  0.157895                       (COFFEE, COCK)
27  0.105263                   (CORNFLAKES, COCK)
28  0.210526                 (CORNFLAKES, COFFEE)
29  0.210526                      (SUGER, COFFEE)
30  0.105263                   (CORNFLAKES, MILK)
31  0.105263                    (CORNFLAKES, TEA)
32  0.105263                         (MAGGI, JAM)
33  0.210526                         (MAGGI, TEA)
34  0.105263               (COFFEE, COCK, BISCUIT)
35  0.105263           (CORNFLAKES, COCK, BISCUIT)
36  0.105263         (CORNFLAKES, COFFEE, BISCUIT)
37  0.105263                (MAGGI, TEA, BISCUIT)
38  0.105263               (TEA, BREAD, BOURNVITA)
39  0.105263                (SUGER, BREAD, COFFEE)
40  0.105263                   (MAGGI, BREAD, JAM)
41  0.105263                   (MAGGI, TEA, BREAD)
42  0.105263            (CORNFLAKES, COFFEE, COCK)
43  0.105263  (CORNFLAKES, COFFEE, COCK, BISCUIT)
```

[68]:
```
asr=association_rules(freq,metric="confidence",min_threshold=0.3)
```

[69]:
```
asr
```

[69]:
```
          antecedents                    consequents  antecedent support  \
0           (BISCUIT)                        (BREAD)            0.315789
1              (COCK)                      (BISCUIT)            0.157895
2           (BISCUIT)                         (COCK)            0.315789
3           (BISCUIT)                       (COFFEE)            0.315789
```

```
4       (CORNFLAKES)                          (BISCUIT)             0.315789
..               …                                   …                    …
87   (COFFEE, BISCUIT)              (CORNFLAKES, COCK)             0.105263
88     (COCK, BISCUIT)            (CORNFLAKES, COFFEE)             0.105263
89        (CORNFLAKES)         (COCK, COFFEE, BISCUIT)             0.315789
90              (COCK)   (CORNFLAKES, COFFEE, BISCUIT)             0.157895
91           (BISCUIT)      (COCK, CORNFLAKES, COFFEE)             0.315789

    consequent support   support  confidence      lift  leverage  conviction  \
0             0.631579  0.157895    0.500000  0.791667 -0.041551    0.736842
1             0.315789  0.105263    0.666667  2.111111  0.055402    2.052632
2             0.157895  0.105263    0.333333  2.111111  0.055402    1.263158
3             0.421053  0.105263    0.333333  0.791667 -0.027701    0.868421
4             0.315789  0.157895    0.500000  1.583333  0.058172    1.368421
..                 …         …           …         …         …           …
87            0.105263  0.105263    1.000000  9.500000  0.094183         inf
88            0.210526  0.105263    1.000000  4.750000  0.083102         inf
89            0.105263  0.105263    0.333333  3.166667  0.072022    1.342105
90            0.105263  0.105263    0.666667  6.333333  0.088643    2.684211
91            0.105263  0.105263    0.333333  3.166667  0.072022    1.342105

    zhangs_metric
0       -0.277778
1        0.625000
2        0.769231
3       -0.277778
4        0.538462
..            …
87       1.000000
88       0.882353
89       1.000000
90       1.000000
91       1.000000

[92 rows x 10 columns]
```

[70]:

```
asr=association_rules(freq,metric="lift",min_threshold=0.3)
```

[71]:

```
asr
```

[71]:

```
          antecedents                      consequents  antecedent support  \
0             (BREAD)                        (BISCUIT)            0.631579
1           (BISCUIT)                          (BREAD)            0.315789
2              (COCK)                        (BISCUIT)            0.157895
3           (BISCUIT)                           (COCK)            0.315789
4            (COFFEE)                        (BISCUIT)            0.421053
..                …                               …                  …
```

```
109  (COCK, BISCUIT)          (CORNFLAKES, COFFEE)            0.105263
110     (CORNFLAKES)        (COCK, COFFEE, BISCUIT)            0.315789
111         (COFFEE)    (CORNFLAKES, COCK, BISCUIT)            0.421053
112           (COCK)  (CORNFLAKES, COFFEE, BISCUIT)            0.157895
113        (BISCUIT)     (COCK, CORNFLAKES, COFFEE)            0.315789

     consequent support   support  confidence      lift  leverage  conviction  \
0              0.315789  0.157895    0.250000  0.791667 -0.041551    0.912281
1              0.631579  0.157895    0.500000  0.791667 -0.041551    0.736842
2              0.315789  0.105263    0.666667  2.111111  0.055402    2.052632
3              0.157895  0.105263    0.333333  2.111111  0.055402    1.263158
4              0.315789  0.105263    0.250000  0.791667 -0.027701    0.912281
..                  …         …          …        …        …          …
109            0.210526  0.105263    1.000000  4.750000  0.083102         inf
110            0.105263  0.105263    0.333333  3.166667  0.072022    1.342105
111            0.105263  0.105263    0.250000  2.375000  0.060942    1.192982
112            0.105263  0.105263    0.666667  6.333333  0.088643    2.684211
113            0.105263  0.105263    0.333333  3.166667  0.072022    1.342105

     zhangs_metric
0        -0.416667
1        -0.277778
2         0.625000
3         0.769231
4        -0.312500
..             …
109       0.882353
110       1.000000
111       1.000000
112       1.000000
113       1.000000

[114 rows x 10 columns]
```

[ ]: