Market basket analysis report

Introduction

Market basket analysis - analysis of custumer behavior, if you buy produsts A and B, you are more likely to buy product C.

Our tutorial: https://www.kaggle.com/code/khusheekapoor/market-basket-analysis-in-python/notebook

Our data: https://www.kaggle.com/datasets/mittalvasu95/the-bread-basket

Data

```
# Head of the data
data.head()
```
✓ 0.0s   🖩 Open 'data' in Data Wrangler

	Transaction	Item	date_time	period_day	weekday_weekend
0	1	Bread	30-10-2016 09:58	morning	weekend
1	2	Scandinavian	30-10-2016 10:05	morning	weekend
2	2	Scandinavian	30-10-2016 10:05	morning	weekend
3	3	Hot chocolate	30-10-2016 10:07	morning	weekend
4	3	Jam	30-10-2016 10:07	morning	weekend

The data is a collection of 20.000+ transactions from a bakery.

Transactions are connected to a unique transaction ID, so we can see what products were bought together.

Analysis

We used the Apriori algorithm to find association rules between products.

```
# finding top 5 single items
frequent_itemsets[(frequent_itemsets['length'] == 1)].sort_values(by='support', ascending=False).head()
```
✓ 0.0s

	support	itemsets	length
0	0.478394	(Coffee)	1
1	0.327205	(Bread)	1
2	0.142631	(Tea)	1
3	0.103856	(Cake)	1
5	0.086107	(Pastry)	1

5 most common products, bought as a single item, are coffee, bread, tea, cake and pastry. 47% of all customers bought coffee, and 37% bought bread.

```
frequent_itemsets[(frequent_itemsets['length'] > 1)].sort_values(by='support', ascending=False).head()
```
✓ 0.0s

	support	itemsets	length
4	0.090016	(Coffee, Bread)	2
9	0.054728	(Coffee, Cake)	2
11	0.049868	(Coffee, Tea)	2
12	0.047544	(Pastry, Coffee)	2
17	0.038246	(Coffee, Sandwich)	2

2 / 5

The most common pairs of products are coffee and bread, bread and tea, tea and cake, cake and bread, and coffee and cake.

```
# finding association rules with minimum support of 2% and having lift more than 1
rules[(rules['support'] >= 0.02) &
      (rules['lift'] > 1.0)]
```
[29] ✓ 0.0s

...

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
2	(Coffee)	(Cake)	0.478394	0.103856	0.054728	0.114399	1.101515
3	(Cake)	(Coffee)	0.103856	0.478394	0.054728	0.526958	1.101515
6	(Pastry)	(Coffee)	0.086107	0.478394	0.047544	0.552147	1.154168
7	(Coffee)	(Pastry)	0.478394	0.086107	0.047544	0.099382	1.154168
8	(Coffee)	(Sandwich)	0.478394	0.071844	0.038246	0.079947	1.112792
9	(Sandwich)	(Coffee)	0.071844	0.478394	0.038246	0.532353	1.112792
10	(Coffee)	(Medialuna)	0.478394	0.061807	0.035182	0.073542	1.189878
11	(Medialuna)	(Coffee)	0.061807	0.478394	0.035182	0.569231	1.189878
12	(Coffee)	(Hot chocolate)	0.478394	0.058320	0.029583	0.061837	1.060311
13	(Hot chocolate)	(Coffee)	0.058320	0.478394	0.029583	0.507246	1.060311
14	(Pastry)	(Bread)	0.086107	0.327205	0.029160	0.338650	1.034977
15	(Bread)	(Pastry)	0.327205	0.086107	0.029160	0.089119	1.034977
16	(Coffee)	(Cookies)	0.478394	0.054411	0.028209	0.058966	1.083723
17	(Cookies)	(Coffee)	0.054411	0.478394	0.028209	0.518447	1.083723
20	(Tea)	(Cake)	0.142631	0.103856	0.023772	0.166667	1.604781
21	(Cake)	(Tea)	0.103856	0.142631	0.023772	0.228891	1.604781
22	(Coffee)	(Toast)	0.478394	0.033597	0.023666	0.049470	1.472431
23	(Toast)	(Coffee)	0.033597	0.478394	0.023666	0.704403	1.472431
26	(Coffee)	(Juice)	0.478394	0.038563	0.020602	0.043065	1.116750
27	(Juice)	(Coffee)	0.038563	0.478394	0.020602	0.534247	1.116750

The biggest association rule is between coffee and cake, with a support of 0.055, followed by coffee and pastry, coffee and sandwich, and coffee and medialuna (argentine croissant).

Visualization

We used a network graph to visualize the association rules.
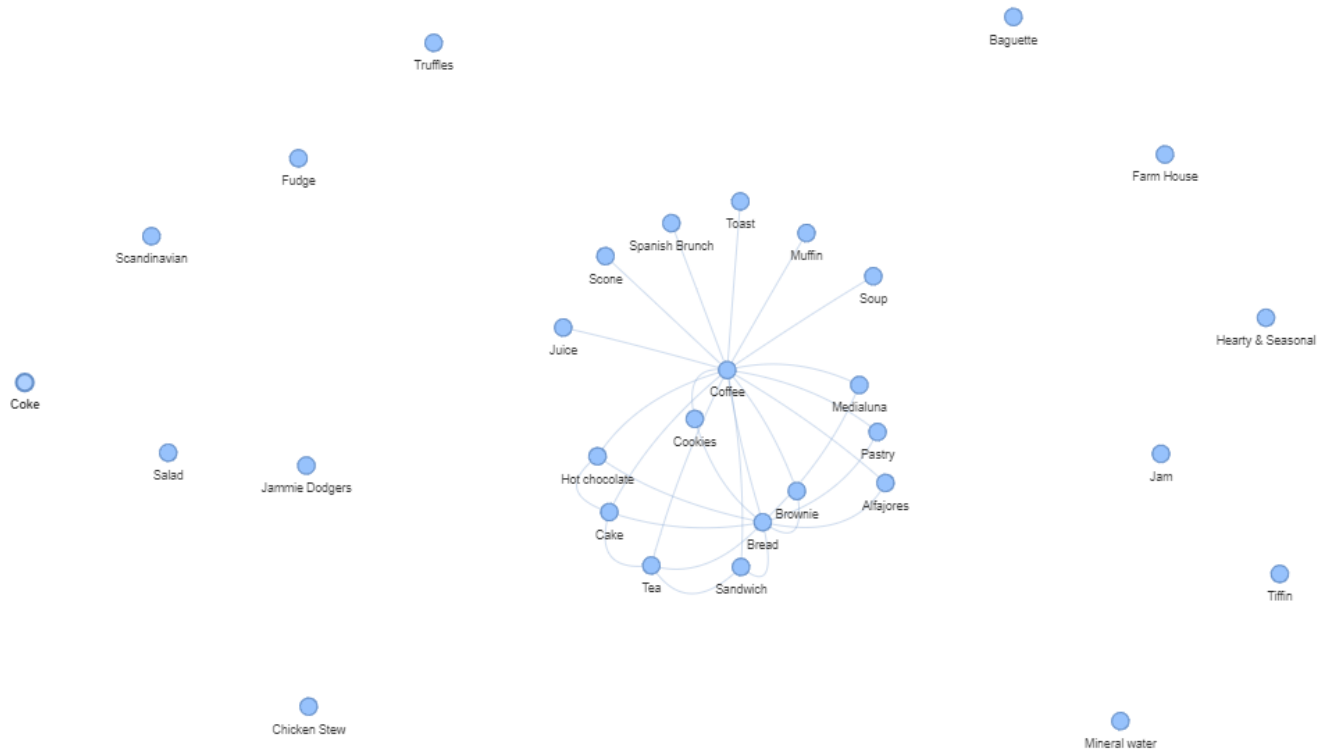
(for the interactive version, see the 'rules_graph.html' file)

Conclusion

Key takeaways from our analysis are:

- Coffee is the most popular product, bought by 47% of all customers, and therefore most connected to other products.
- Top 5 products bought with coffee are bread, cake, tea, pastry and sandwich.

- Bread is the second most popular product, bought by 37% of all customers, and most connected to coffee, pastry and tea.
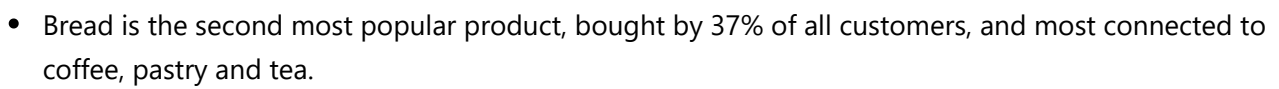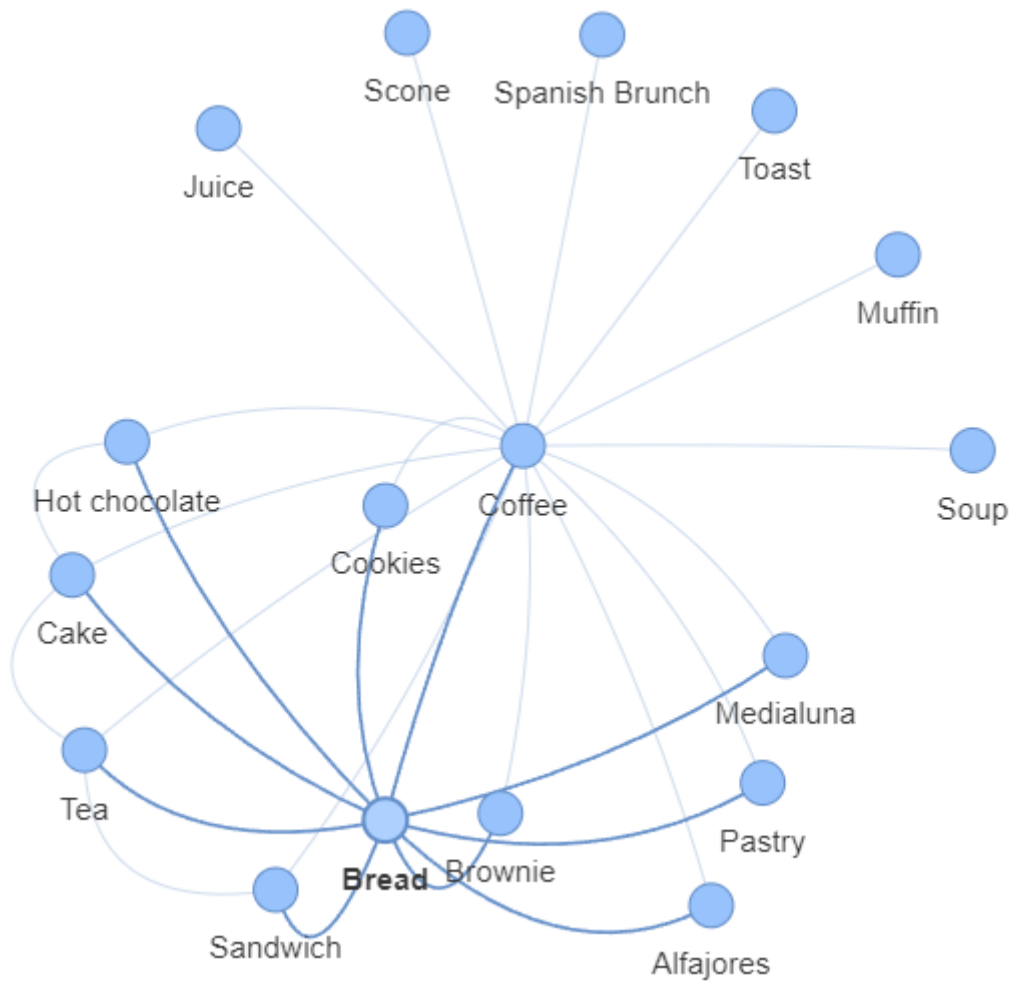
- All other products are connected to coffee and/or bread, but not to each other.
- There are some products, that have no connection to other products, like salad, baguette or coke.