SALES ANALYSIS IN A BAKERY SHOP: A DATA MINING APPROACH

Prepared By -
Vikas Kashyap

ABOUT DATASET

Dataset Source: The dataset was obtained from Kaggle

Source: https://www.kaggle.com/datasets/matthieugimbert/french-bakery-daily-sales/data

Content: The dataset contains sales transactions from a bakery, including various attributes such as date, time, ticket number, product, quantity, and unit price.

PROBLEM STATEMENT

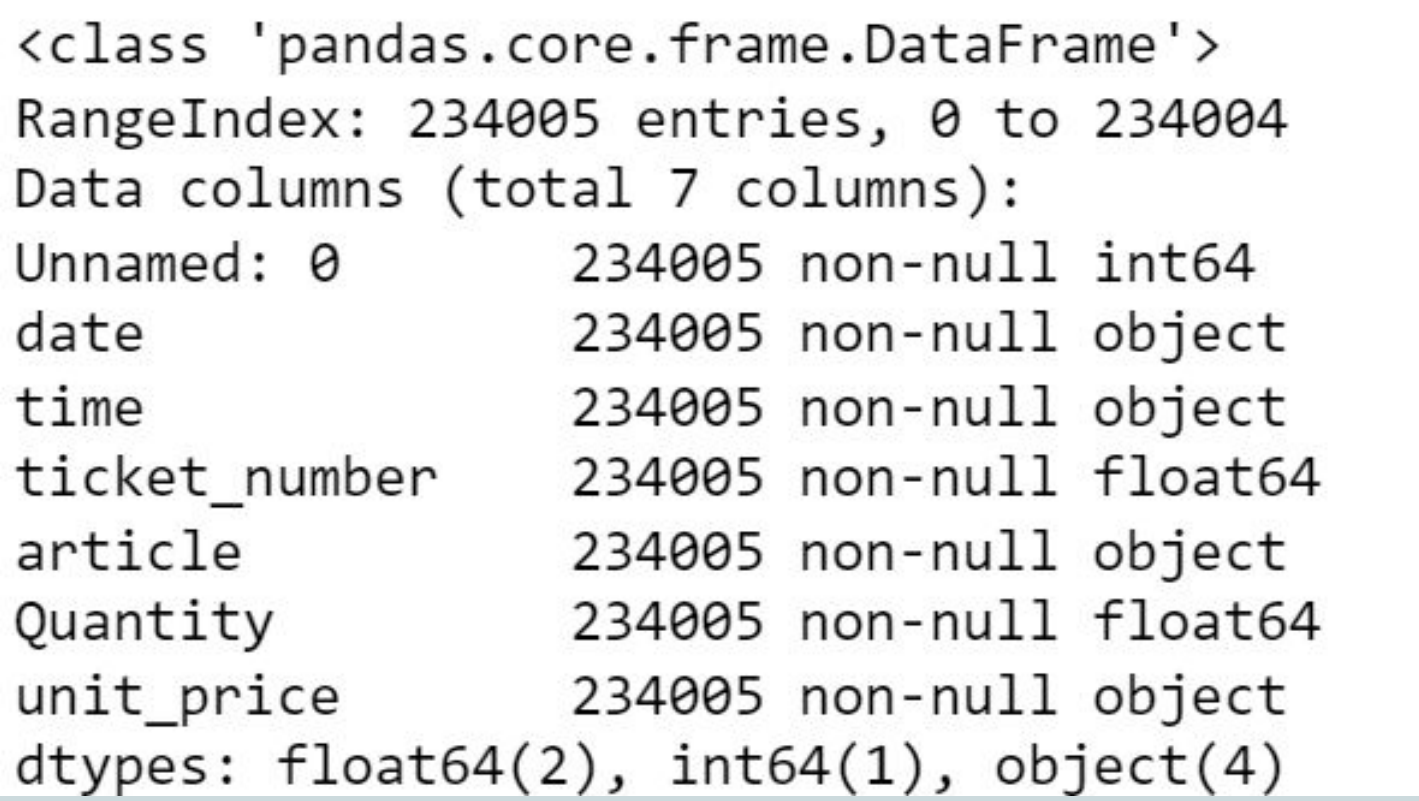
- The primary goal is to uncover patterns, trends, and insights from the French bakery shop sales dataset.
- This analysis will assist in understanding customer buying behaviors and eventually aid in strategic decision-making such as inventory management, promotional bundles, and product placement.

DATA PREPROCESSING

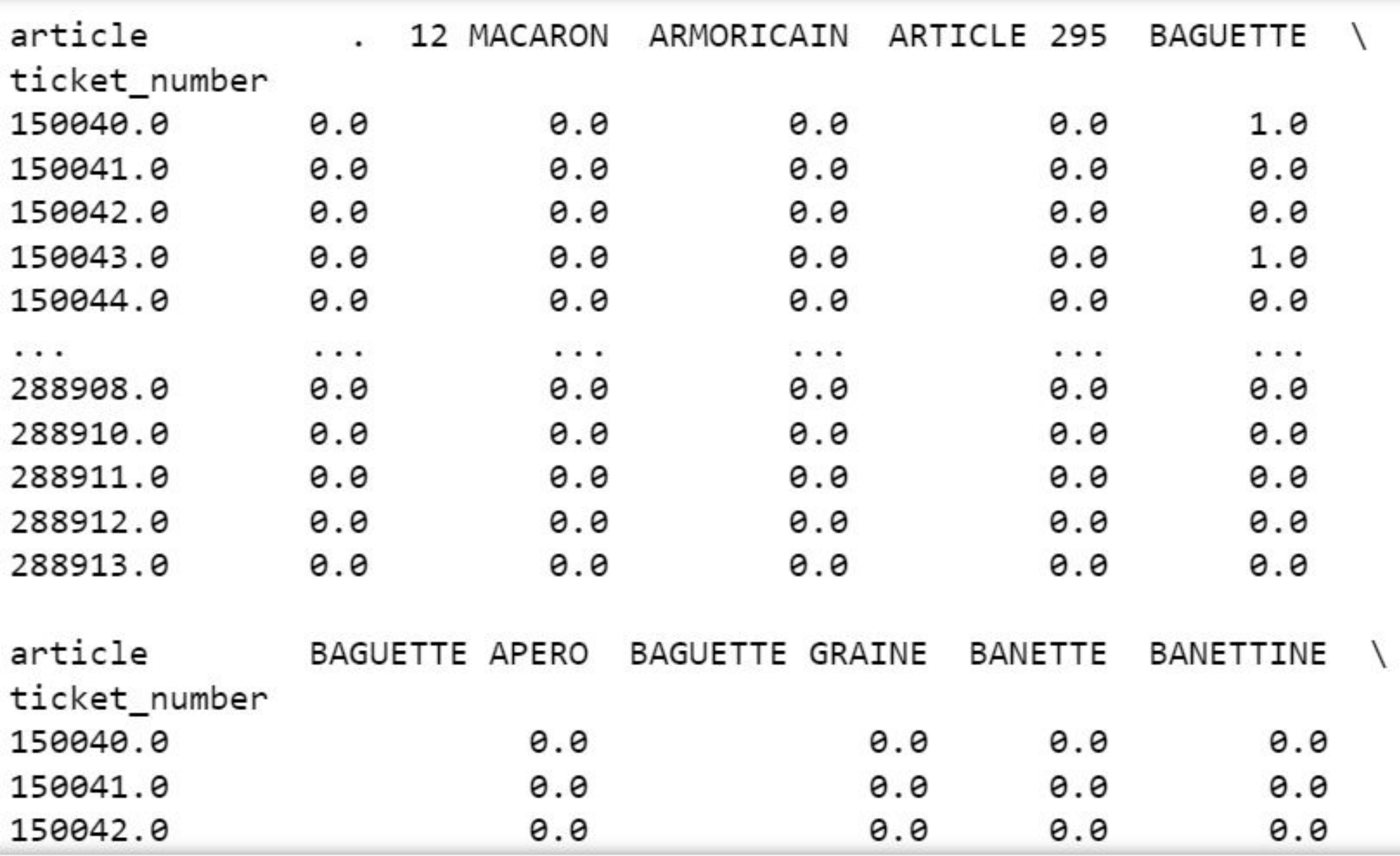
Handling Missing Values: During the preprocessing stage, missing values were checked for in the dataset. Fortunately, no missing values were found, eliminating the need for further handling or imputation techniques.

Checking Mistyped values: We thoroughly examined the article names for any mistyped values such as "None" or "N/A". Fortunately, we found no such occurrences among our 149 unique articles, affirming the dataset's quality and completeness.

One-Hot Encoding: Each unique item is represented as a separate binary column. For each transaction, if an item is present in that transaction, the corresponding column is set to 1; otherwise, it is set to 0.



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 234005 entries, 0 to 234004
Data columns (total 7 columns):
Unnamed: 0      234005 non-null int64
date            234005 non-null object
time            234005 non-null object
ticket_number   234005 non-null float64
article         234005 non-null object
Quantity        234005 non-null float64
unit_price      234005 non-null object
dtypes: float64(2), int64(1), object(4)
```



```
article          .  12 MACARON  ARMORICAIN  ARTICLE 295  BAGUETTE  \
ticket_number
150040.0       0.0         0.0         0.0          0.0       1.0
150041.0       0.0         0.0         0.0          0.0       0.0
150042.0       0.0         0.0         0.0          0.0       0.0
150043.0       0.0         0.0         0.0          0.0       1.0
150044.0       0.0         0.0         0.0          0.0       0.0
...            ...         ...         ...          ...       ...
288908.0       0.0         0.0         0.0          0.0       0.0
288910.0       0.0         0.0         0.0          0.0       0.0
288911.0       0.0         0.0         0.0          0.0       0.0
288912.0       0.0         0.0         0.0          0.0       0.0
288913.0       0.0         0.0         0.0          0.0       0.0

article        BAGUETTE APERO  BAGUETTE GRAINE  BANETTE  BANETTINE  \
ticket_number
150040.0                  0.0              0.0      0.0        0.0
150041.0                  0.0              0.0      0.0        0.0
150042.0                  0.0              0.0      0.0        0.0
```

Source: https://www.analyticsvidhya.com/blog/2022/07/market-basket-analysis-for-coffee-shop-with-apriori/

METHODOLOGY

Approach: Used the Apriori algorithm to find frequent itemsets in the transaction data.

- **Support(X) = (Number of transactions containing X) / (Total number of transactions)**

The support of an item is the proportion of transactions in which the item appears. Items with a support greater than a specified minimum support threshold are considered frequent items. Support threshold of 0.01 or 1%.

- Confidence(X => Y) = (Number of transactions containing X and Y) / (Number of transactions containing X)

Association rules generated from the frequent itemsets. An association rule consists of an antecedent (premise) and a consequent (conclusion) and is in the form A -> B, where A and B are itemsets. Calculate the confidence of each rule, which is the proportion of transactions containing A that also contain B. Prune rules that do not meet a specified minimum confidence threshold.

Source: https://www.geeksforgeeks.org/apriori-algorithm/

ANALYSIS

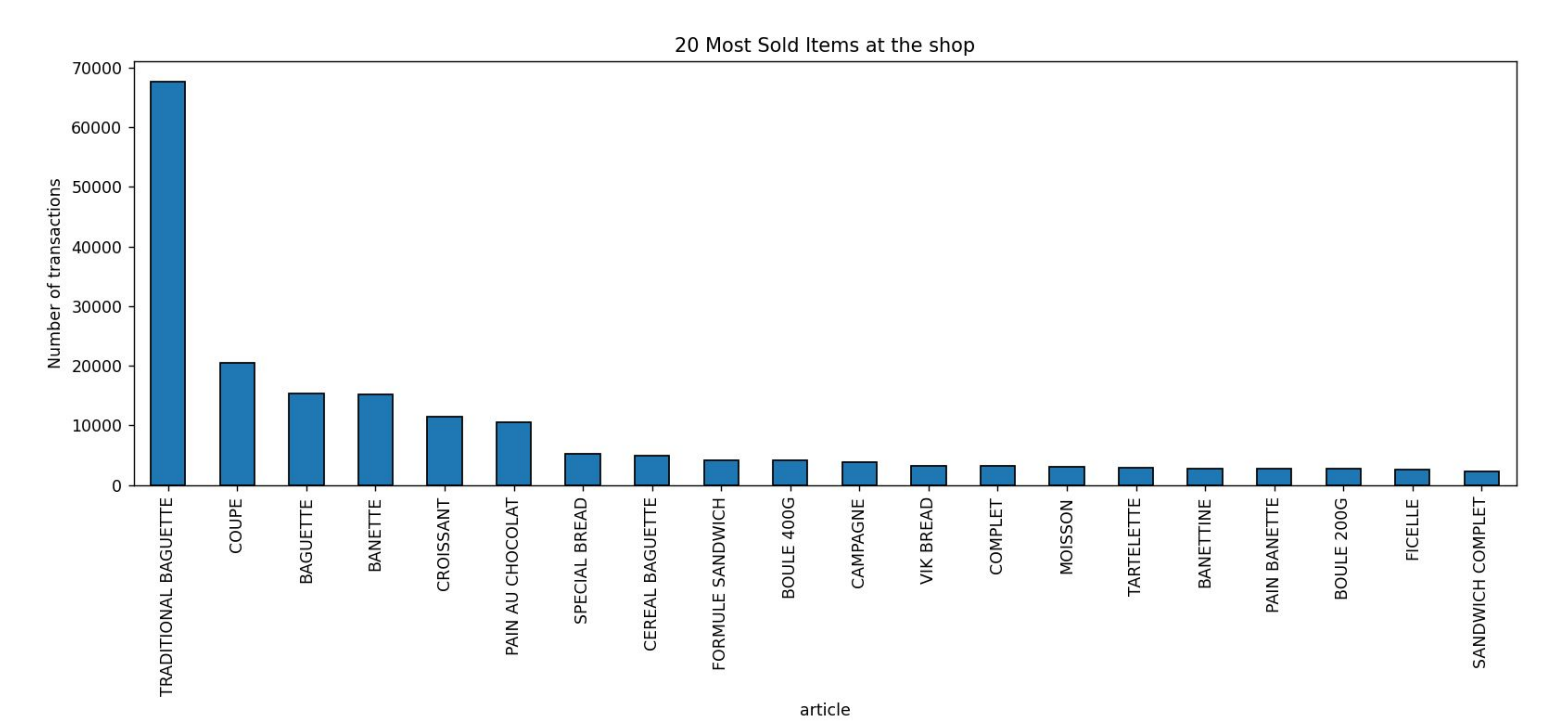


Figure: Top 20 best-selling items for the French bakery shop

ANALYSIS

Frequent item sets were identified using the Apriori algorithm with a support threshold of 0.01 or 1%.

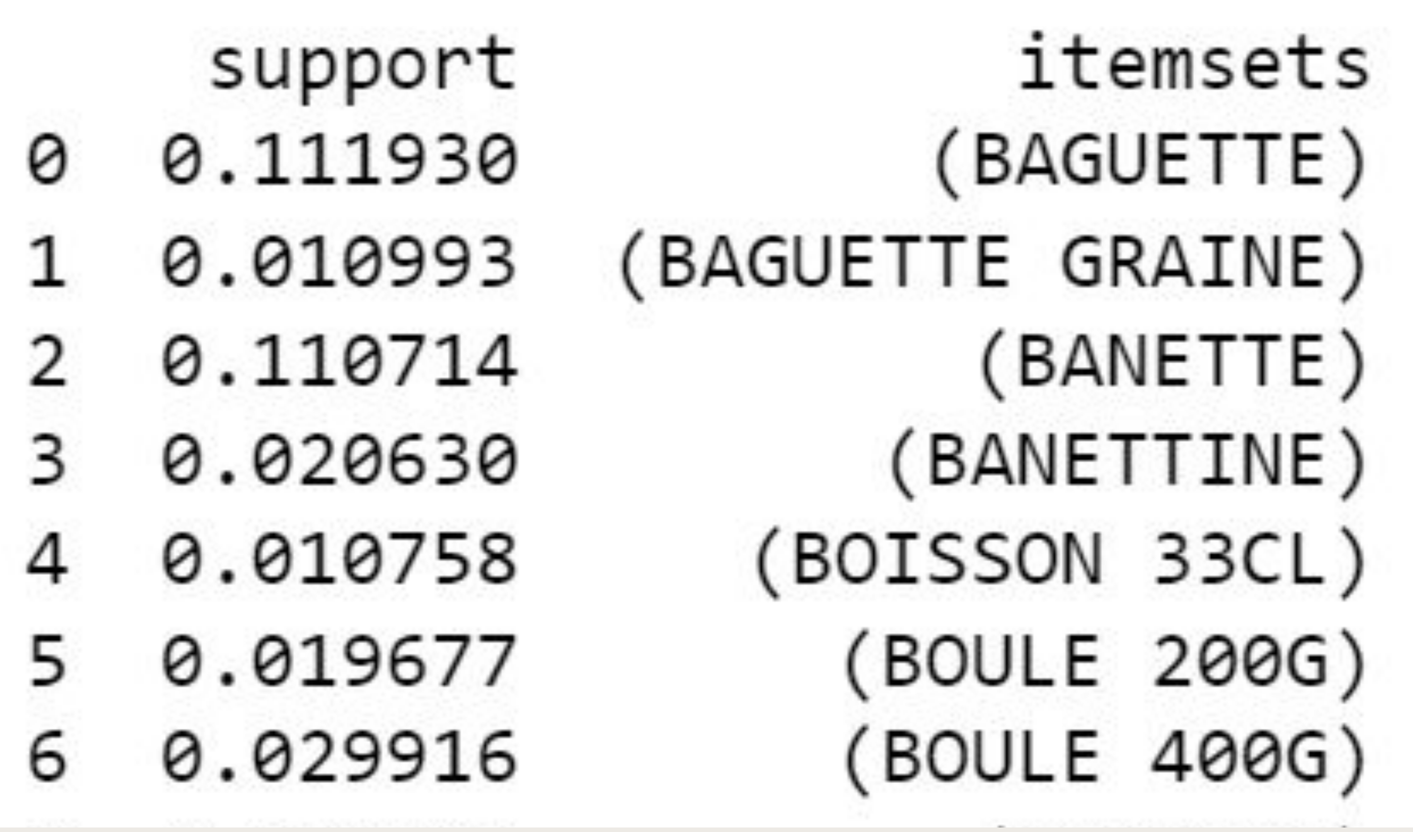
- BAGUETTE) has the highest support of 0.111930, which means that approximately 11.19% of the transactions in the dataset contain the item "BAGUETTE".

- (BANETTE) has the second-highest support of 0.110714, indicating that around 11.07% of the transactions include the item "BANETTE".

- (CAMPAGNE) has a support of 0.028523, meaning that about 2.85% of the transactions contain the item "CAMPAGNE".

- (BOULE 400G) and (BOULE 200G) have support values of 0.029916 and 0.019677, respectively, suggesting that these bread items are also frequently purchased.

- (BANETTINE), (BRIOCHE), (BOISSON 33CL), (BAGUETTE GRAINE), and (CAFE OU EAU) are other frequent itemsets with lower support values, ranging from



```
    support            itemsets
0  0.111930          (BAGUETTE)
1  0.010993  (BAGUETTE GRAINE)
2  0.110714           (BANETTE)
3  0.020630         (BANETTINE)
4  0.010758      (BOISSON 33CL)
5  0.019677        (BOULE 200G)
6  0.029916        (BOULE 400G)
```

Association rules were generated based on the identified frequent item sets.

Insights into item associations and purchasing patterns were derived from these association rules.

Here's an interpretation of the association rules:

Rule 1: (BOULE 200G) => (COUPE)

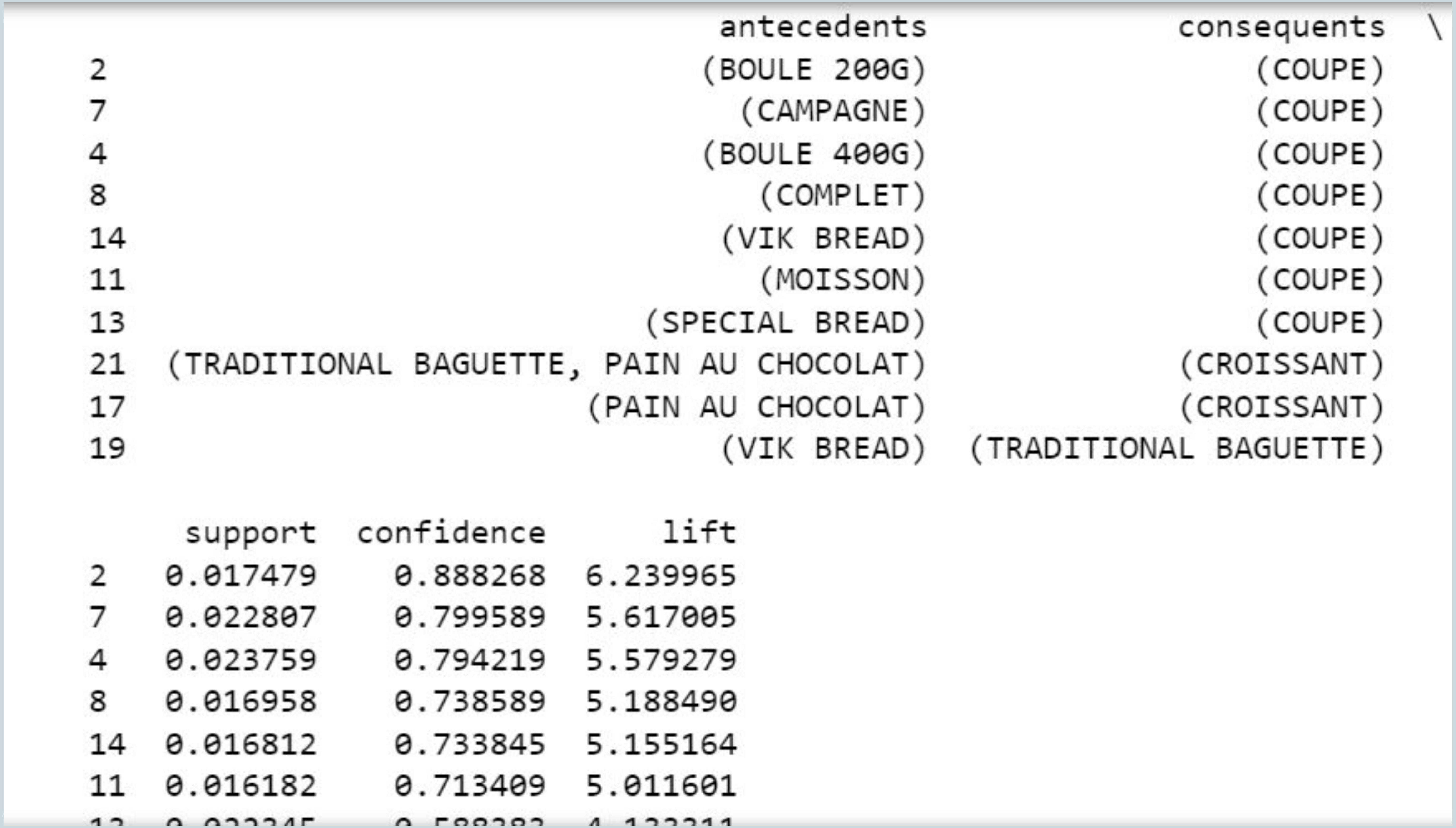
- ○ Support: 0.017479 (1.75% of transactions contain both items)
- ○ Confidence: 0.888268 (88.83% of transactions with "BOULE 200G" also contain "COUPE")
- ○ Lift: 6.239965 (The presence of "BOULE 200G" increases the likelihood of purchasing "COUPE" by 6.24 times)

Rule 2: (CAMPAGNE) => (COUPE)

- ○ Support: 0.022807 (2.28% of transactions contain both items)
- ○ Confidence: 0.799589 (79.96% of transactions with "CAMPAGNE" also contain "COUPE")
- ○ Lift: 5.617005 (The presence of "CAMPAGNE" increases the likelihood of purchasing "COUPE" by 5.62 times)

Rule 3: (BOULE 400G) => (COUPE)

- ○ Support: 0.023759 (2.38% of transactions contain both items)
- ○ Confidence: 0.794219 (79.42% of transactions with "BOULE 400G" also contain "COUPE")
- ○ Lift: 5.579279 (The presence of "BOULE 400G" increases the likelihood of purchasing "COUPE" by 5.58 times)



```
                                      antecedents               consequents  \
2                                    (BOULE 200G)                   (COUPE)
7                                      (CAMPAGNE)                   (COUPE)
4                                    (BOULE 400G)                   (COUPE)
8                                       (COMPLET)                   (COUPE)
14                                    (VIK BREAD)                   (COUPE)
11                                      (MOISSON)                   (COUPE)
13                                (SPECIAL BREAD)                   (COUPE)
21  (TRADITIONAL BAGUETTE, PAIN AU CHOCOLAT)                   (CROISSANT)
17                            (PAIN AU CHOCOLAT)                (CROISSANT)
19                                    (VIK BREAD)  (TRADITIONAL BAGUETTE)

     support  confidence      lift
2   0.017479    0.888268  6.239965
7   0.022807    0.799589  5.617005
4   0.023759    0.794219  5.579279
8   0.016958    0.738589  5.188490
14  0.016812    0.733845  5.155164
11  0.016182    0.713409  5.011601
13  0.022345    0.588383  4.133311
```

Figure: Association rules

KEY LEARNING

- Businesses are always looking to optimize their setup and drive up their sales. This kind of analysis could have been done for any retail store or marketplace.
- Now we know the correlation between items and the common interest of the customers, the business can make decisions based on these findings.
- Market basket analysis can inform **promotional strategies.**
- **Customer Preferences**
- Analysis over time might reveal **seasonal trends.**
- **Cross-Selling Opportunities**

THANK YOU