

How to Improve the Bakery Sales with Market Basket Analysis

By: Mega Restu Utari



Outline



Introduction



About the
Data



EDA



Modeling



Conclusion



Business
Strategy



Introduction

A bakery is a business that makes and sells baked goods such as bread, cookies, cakes, and pastries. Some retail bakeries double as cafés, selling coffee and tea to customers who want to eat their baked goods on the spot.

Businesses are constantly seeking methods to improve their setup and boost sales. Bakeries are no different; one technique for increasing sales is to discover the association between items and consumers' mutual interests.



Market basket analysis is one of the methods that can be used. When a consumer buys product A, market basket analysis looks at previous transaction data **to uncover relationships or patterns so that we can recommend product B when they buy product A.**



About the Data



According to the dataset explanation (found at <https://www.kaggle.com/sulmansarwar/transactions-from-a-bakery>) this data belongs to a bakery called "The Bread Basket", located in Edinburgh, Scotland. This bakery presents a refreshing offer of Argentine and Spanish products.

The dataset provides transaction details of all items purchased between 2016 and 2017 from the bakery online. The dataset has **21293** entries, and 4 columns.

Data columns (total 4 columns):				
#	Column	Non-Null Count		Dtype
0	Date	21293	non-null	object
1	Time	21293	non-null	object
2	Transaction	21293	non-null	int64
3	Item	21293	non-null	object



EDA

1

What was **the 10**
best seller items?

2

What was **peak**
hours of sales?

3

How was **the**
monthly sales?

4

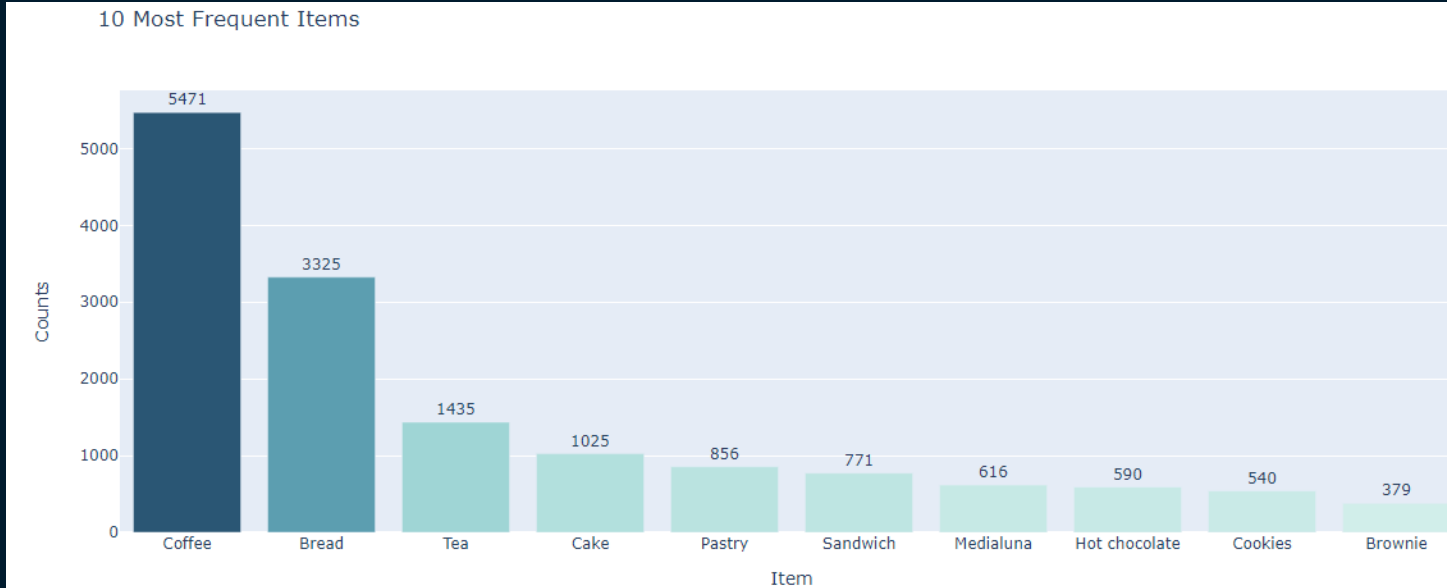
How was **the**
weekly sales?



1

What was
the 10 best
seller
items?

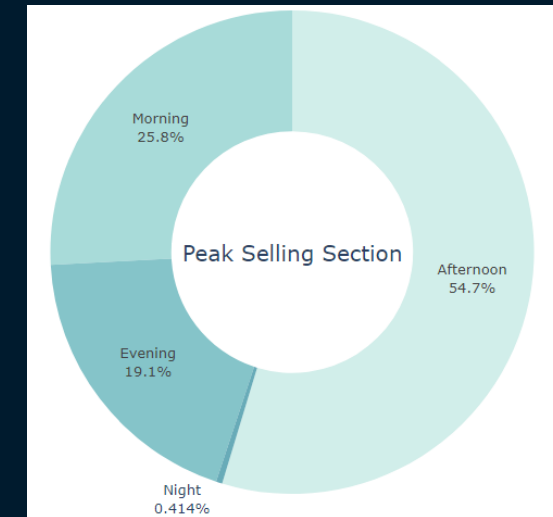
Coffee is the best-selling item by far, followed by bread and tea.



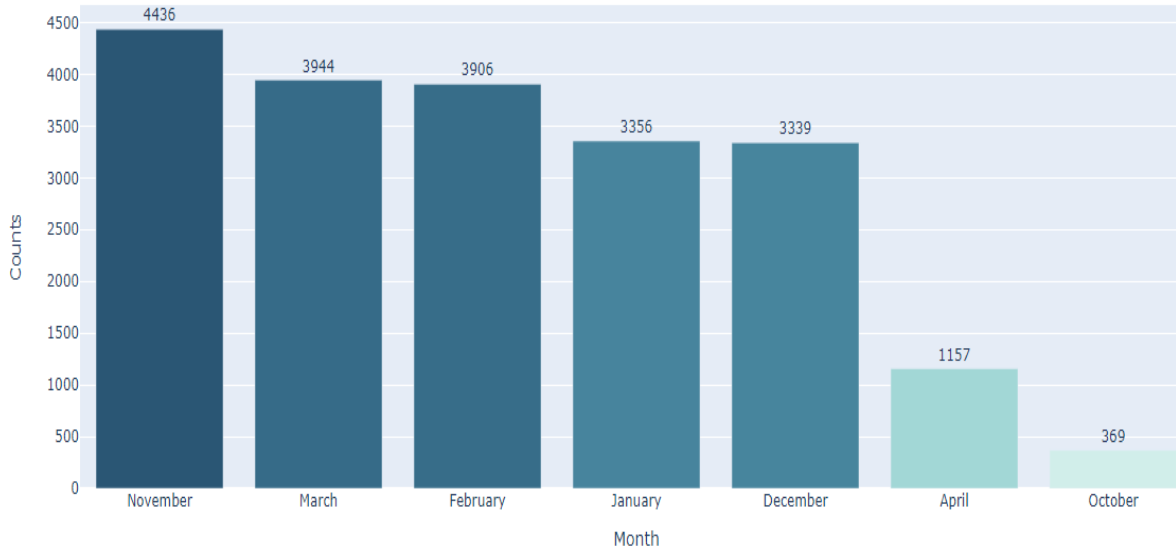
2

What was
peak hours
of sales
?

The bakery seems to be making most of its sales in the afternoon everyday with over 56% of the sales. Sales fall sharply after that. However, the bakery makes a decent amount of sales in the morning as well.



Monthly Sales

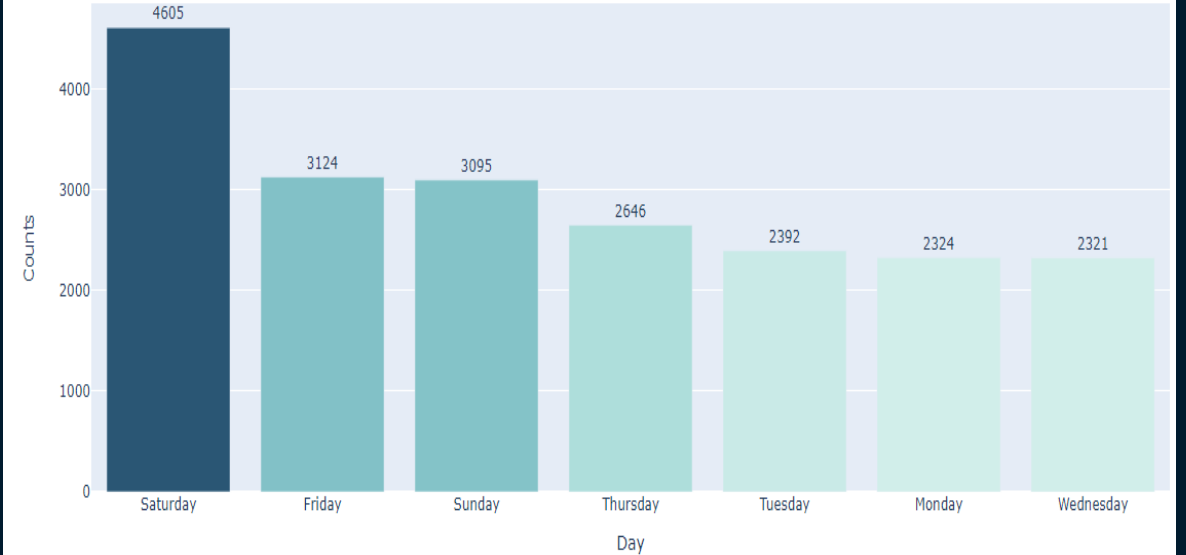


3

How was **the monthly sales?**

The data set includes dates from 30/10/2016 to 09/04/2017, that's why we have so few transactions in October and April.

Most Productive Day



As we can see, Saturday is the busiest day in the bakery. Conversely, Wednesday is the day with fewer transactions.

4

How was **the weekly sales?**



Modeling

Market Basket Analysis uses something called as **Association Rules** which analyze the frequently occurring items in a shoppers cart. The most popular algorithm is called as **Apriori algorithm**. It would come up with rules such as

“If the customer has picked up bread, then there is a good likelihood that the customer would also pick up butter”

To understand Association Rules better, it is important to note and understand the three concepts of **Support, Confidence and Lift**.



Modeling

Support

Support is an indication of how frequently the item set appears in the data set.

$$\text{Support} = \text{Frequency}(\text{Product } A) / \text{Total Number of Transactions}$$

Confidence

Confidence measures the percentage of product B is purchased, whenever the product A is purchased. It's an indication of how often the rule is proven correct.

$$\text{Confidence} = \text{Frequency}(A, B) / \text{Frequency}(A)$$

Lift

Lift indicates whether the two products A & B are dependent on each other or independent.

$$\text{Lift} = \frac{\text{Support}(A, B)}{(\text{Support}(A) * \text{Support}(B))}$$



Modeling

Item_1	Item_2	support	convidence	lift
Toast	Coffee	0.023666	0.704403	1.472431
Spanish Brunch	Coffee	0.010882	0.598837	1.251766
Medialuna	Coffee	0.035182	0.569231	1.189878
Pastry	Coffee	0.047544	0.552147	1.154168
Alfajores	Coffee	0.019651	0.540698	1.130235
Juice	Coffee	0.020602	0.534247	1.116750
Sandwich	Coffee	0.038246	0.532353	1.112792
Cookies	Coffee	0.028209	0.518447	1.083723
Bread	Pastry	0.029160	0.338650	1.034977
Cake	Tea	0.023772	0.228891	1.604781
Hot Chocolate	Cake	0.011410	0.195652	1.883874

How do we interpret these rules?

For example:

Toast -> Coffee

- In around 2% of all the transactions, Toast and Coffee are bought together.
- When Toast is bought, the likelihood that Coffee is then bought is 70%.
- When Toast is sold, it increases the probability of then selling Coffee by 47%.

And so on. From the table we know that coffee is popular consequents. It seems in this bakery there are many coffee lovers!



Conclusion

- Coffee is the bestseller of this bakery and it shows association with **13** other items.
- Whenever the customer purchases toast, they may also purchase coffee for which support is 2%, confidence is 70% and lift is 1.47 which means both of the products have a high probability to occur in the same order.
- Besides coffee, we found more interesting item correlations:
 - Bread -> Pastry
 - Cake -> Tea
 - Hot Chocolate -> Cake



Business Strategy

- The team (leader and staff) can work together to design strategies and training to encourage customers who purchased item 1 to buy item 2, doing cross-selling and increased store's income.
- The bakery might be interested in having a promotion of a free item, given the great chances of another item being sold as a result of it.
- The bakery could place their freshly baked bread near their pastries since customers who purchase pastries seem to also be enticed by bread.



THANK YOU!

