جامعة أم القرى
UMM AL-QURA UNIVERSITY

DATA ANALYSIS 2

TASK2 :

Market Basket Analysis

Dataset :

Instacart Market Basket Analysis Dataset

SUBMITTED BY:	STUDENT ID:
Fatima Osama Altayyeb	443000772
Asayel Sloom Alharbi	444006773

2024

Introduction:

this analysis, we aimed to explore product and order data to gain insights into purchasing behavior and identify the most popular products among customers. By examining various aspects of the data, we can enhance inventory management, optimize marketing strategies, and improve overall customer satisfaction.

1. Importing Required Libraries:
   - The pandas library was used for data manipulation, along with visualization libraries like matplotlib and seaborn for plotting.

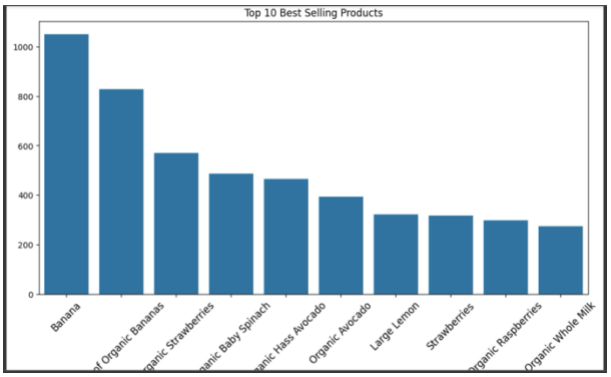
2. Loading Data: - Multiple CSV files containing product, aisle, department, order, and prior order data were loaded into separate DataFrames.

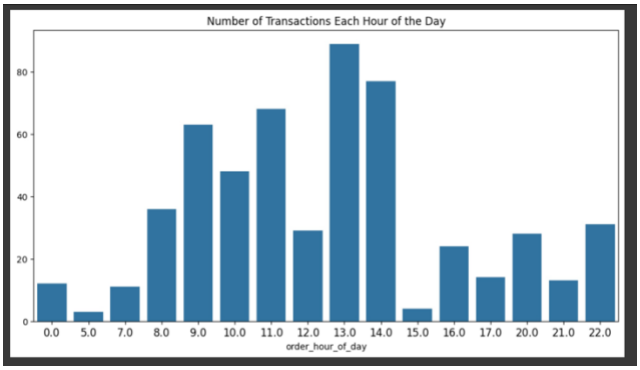
3. Dropping Duplicates:
   - Duplicates were removed from the orders DataFrame to ensure data integrity.

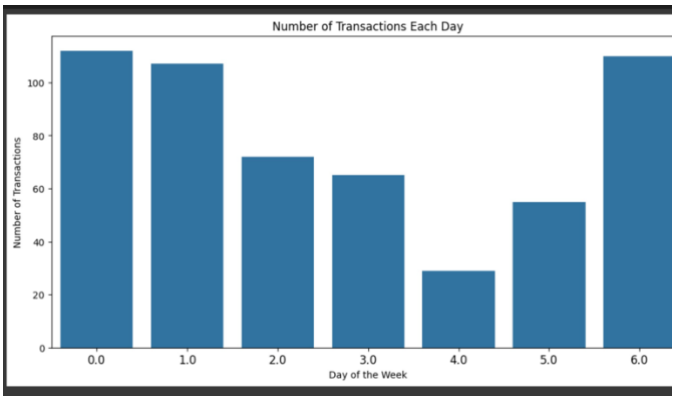
4. Merging Data: - The prior order data was merged with product data, and subsequently with the original orders DataFrame to consolidate information about each order.

5. Calculating Purchase Counts:
   - We calculated the number of purchases for each product and identified the top 10 best-selling products.

6. Charts :







1. Top 10 Best Selling Products:
- "Banana" is the top-selling product.
- Followed by "Bag of Organic Bananas" and "Organic Strawberries."

2. Number of Transactions Each Hour of the Day:
- Peak at 13:00.
- Most activity between 9:00 and 15:00.

3. Number of Transactions Each Day of the Week:
- Sunday and Monday have the highest number of transactions.
- Thursday shows the least activity.

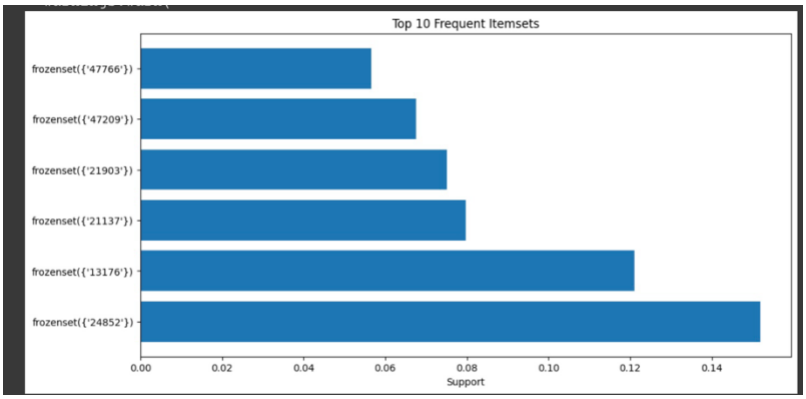
7 . Order analysis and unique transaction identification:

1.Product Count Calculation:
The code groups the data by user ID, order ID, and product name, then calculates how many times each product was purchased, storing this in the Count column.

2.Adding a Transaction Number:
A Transaction column is added, containing a unique number for each order to identify it easily.

Results:
Each customer has the number of products purchased in their orders, with a unique transaction number for each order.

8 . Association Rules :



• Extract recurring patterns from shopping baskets using FP-Growth algorithm to identify products that are frequently purchased together.

• Create association rules between products to identify relationships between items based on lift.

• Display top 10 groups of recurring items using a horizontal chart based on support.

Results Overview:

shows the most popular products that are frequently purchased together, with the top 10 most frequent products in shopping carts shown in the chart. This information helps you understand customer behavior and decide which products to focus on in promotions.

	antecedents	consequents	support	confidence	lift
29	(28204)	(24852)	0.010660	0.381125	2.510126
27	(49683)	(24852)	0.010965	0.352365	2.320711
21	(47766)	(24852)	0.017158	0.303411	1.998294
15	(27966)	(13176)	0.013097	0.297921	2.460694
31	(16797)	(24852)	0.013249	0.288079	1.897318
6	(47209)	(13176)	0.018986	0.280992	2.320862
25	(47626)	(24852)	0.012133	0.258378	1.701704
23	(27845)	(24852)	0.010305	0.248470	1.636446
32	(26209)	(24852)	0.010153	0.232558	1.531649
17	(21137)	(24852)	0.018123	0.227244	1.496648

1. Extracting Frequent Itemsets: The FP-Growth algorithm is used to identify frequent items in the basket data, with a minimum support threshold of 0.01.

2. Generating Association Rules: Association rules are generated from the frequent itemsets, using "lift" as a metric, with a minimum lift value of 0.5.

3. Ranking Results: The results are sorted by "confidence" to show the top 10 association rules, helping to understand the relationships between products better.

Results Overview:
- Understanding Relationships: Shows how products are related to each other.
- Support: Indicates the proportion of transactions that contain both products.
- Confidence: Reflects the strength of the relationship between products.
- Lift: Measures the effect of one product on another, where values above 1 indicate a positive relationship.

Practical Benefits:
- Marketing Strategies: Enhancing campaigns and recommendations.
- Inventory Management: Identifying products that should be stocked together.
- Improving Customer Experience: Providing recommendations to increase sales.

Conclusion:

1. Top-Selling Products: Bananas and bags of organic bananas topped the sales list, reflecting customers' preferences for fresh and healthy options.

2. Transactions by Day and Hour: Graphs indicated an increase in transactions on specific days of the week and at certain hours, which can inform marketing and sales strategies.

3. Purchase Behavior Analysis: By calculating the number of products purchased per customer, we gained a better understanding of repeat purchases and customer preferences.

4. Product Association Exploration: Utilizing FP-Growth techniques, we identified common buying patterns, aiding in targeted promotions and marketing strategies.

These findings are valuable for enhancing sales and marketing strategies, enabling businesses to better meet customer needs and increase revenue.