

# **Market Basket Analysis Report**

## **On E-Commerce Data**

**Umm Al-Qura University**

**College of Computer Science**

**Data Science Department**

## Data Description

The Ecommerce Dataset is a comprehensive collection of online retail transaction data, primarily focused on sales made by a UK-based online retailer from 2010 to 2011. This dataset is commonly used for customer segmentation, purchase behavior analysis, and sales trend forecasting in e-commerce.

### Column Data:

**InvoiceNo:** Unique identifier for each invoice

**StockCode:** Identifier for the product

**Description:** Description of the product

**Quantity:** Number of items purchased

**InvoiceDate:** Date of the transaction

**UnitPrice:** Price per unit of the product

**CustomerID:** Unique identifier for customers

**PCountry:** Country of the customer.

## 1. Introduction

- **Purpose:** This analysis aims to uncover relationships between products frequently purchased together using association rule learning techniques.
- **Problem Solved:** It provides valuable insights for marketing strategies and inventory management by analyzing retail transaction data.

## 2. Theoretical Background

- **Basic Concepts:** Market Basket Analysis focuses on identifying co-occurring items in transactions, commonly using association rules.
- **Key Libraries/Tools:** Commonly utilized libraries include pandas for data manipulation, matplotlib for visualization, and mlxtend for implementing the Apriori algorithm.

### 3. Data Processing

- 1. Remove Duplicates:** Identify and remove any duplicate records to ensure data integrity.
- 2. Calculate Non-Empty Values:** Count the number of non-empty (non-null) entries in each relevant column to understand data completeness.
- 3. Convert Date to Time Format:** Convert the **InvoiceDate** column to a standard datetime format for easier analysis.
- 4. Grouping Dataset:** Group the dataset by **CustomerID** and **InvoiceDate** to create a list of products bought by each customer on the same date.

### 4. Modeling Used

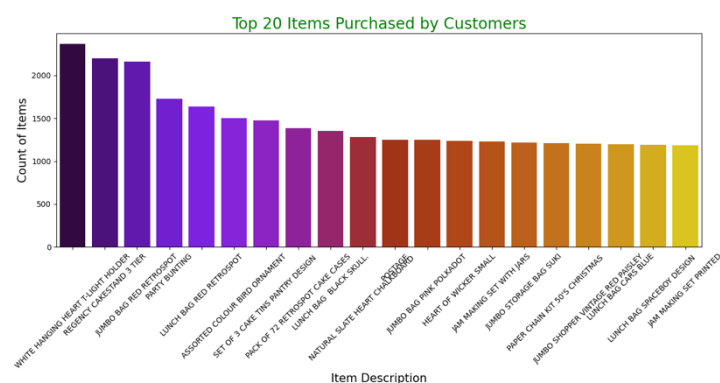
**Apriori Algorithm:** This algorithm is utilized to find frequent itemsets and association rules in the dataset.

### 5. Evaluation Metrics

- **Lift:** Measures how much more likely two items are purchased together compared to being purchased independently.
- **Confidence:** Indicates the likelihood of purchasing an item given that another item has been purchased.
- **Support:** Represents the proportion of transactions that include a particular item or itemset.

### 6. Visualization and Results

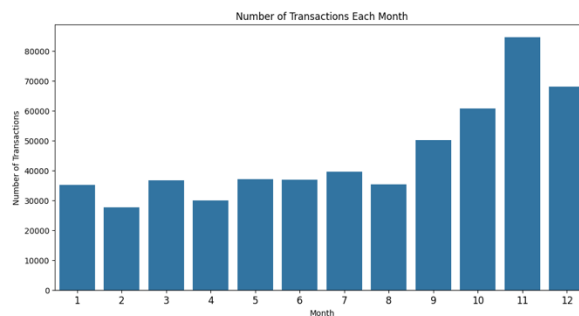
- **Chart:** Use visualizations to showcase the results of the analysis, particularly focusing on frequent itemsets and their relationships.



- **Displaying Results:** Show the first five and first ten results of the association rules generated by the Apriori algorithm.

	antecedents	consequents	support	confidence	\
35	[rolls/buns, soda, whole milk]	[rolls/buns]	0.2	1.0	
33	[rolls/buns, soda, whole milk]	[whole milk]	0.2	1.0	
6	[other vegetables, whole milk]	[whole milk]	0.4	1.0	
15	[rolls/buns, whole milk]	[whole milk]	0.4	1.0	
27	[other vegetables, yogurt, whole milk]	[whole milk]	0.2	1.0	
	lift				
35				2.50	
33				1.25	
6				1.25	
15				1.25	
27				1.25	

- **Chart:** The chart shows monthly sales. Sales increase throughout the year, peaking in December. This information can help businesses plan for busy seasons, stock inventory, and predict revenue.



## 7. Potential Issues

- **Limitations:** Possible limitations include handling large datasets and varying transaction formats.
- **Suggestions for Improvement:** Consider optimizing data loading and analysis processes, and implement error handling for unexpected input formats. And use smaller data and different algorithm like FP-Growth.

## 8. Real-World Applications

**Cross-Selling Strategies:** By identifying frequently co-purchased items, businesses can create bundled offers or promotional discounts, enhancing customer value.

**Store Layout Optimization:** Retailers can place frequently bought together items in proximity, improving customer convenience and potentially increasing sales.

**Inventory Management:** Understanding item relationships helps in stocking decisions, ensuring high-demand items are available together.

## 9. Conclusion

Market Basket Analysis using the Apriori algorithm provides valuable insights into customer purchasing behavior, facilitating informed marketing strategies and operational improvements. This analysis can guide retailers in enhancing customer satisfaction and increasing revenue through targeted offers and efficient inventory management.

## 10. References

E-Commerce Data: <https://www.kaggle.com/datasets/carrie1/ecommerce-data>