



جامعة أم القرى  
UMM AL-QURA UNIVERSITY

## project of Data Analysis 2

Name	ID
Mawadh Saleem Aloufi	444005971
Reham Hameed Alqurashi	444002142

## Market Basket Analysis

- **Overview of the Dataset:**

The “Ecommerce Data” dataset contains information about electronic transactions from an online retail store. The data was collected from a variety of customers and sales, consisting of approximately 25,000 transactions, with the following key columns:

- **InvoiceNo:** A unique identifier for each transaction.
- **StockCode:** A unique identifier for each product.
- **Quantity:** The number of units purchased in each transaction.
- **CustomerID:** A unique identifier for each customer.
- **Country:** The country where the customer resides.

- **Data Preparation:**

Column Transformation to Text: InvoiceNo and StockCode are converted to strings to ensure proper handling during analysis.

Data Cleaning: Rows with negative quantities are removed, and rows without a CustomerID are deleted. These steps ensure that the data used is accurate and does not contain invalid values.

- **Creating the Basket:**

Pivot Table Creation: After cleaning the data, a pivot table is created where the rows represent invoices and the columns represent products. The values represent the quantities sold for each product, which are then transformed into binary values (1 and 0) to indicate whether the product was purchased or not.

- **Applying the Apriori Algorithm:**

Applying Apriori on a Subset: The Apriori algorithm is applied to the first 10,000 rows of the basket to extract frequent itemsets.

- **Displaying Frequent Itemsets:** The first frequent itemset is displayed to uncover patterns in the purchase data.



- **Generating Association Rules:**

Generating Association Rules: Association rules are generated using the extracted frequent itemsets, with the lift metric used to evaluate the strength of relationships between products.

- **Analyzing Best-Selling Products:**

Sales Analysis: The data is grouped by StockCode to calculate the total quantities sold for each product, and the top 10 products based on sales are extracted.

Displaying Descriptions: The descriptions associated with the best-selling products are displayed for a better understanding of the nature of these products.

- **Filtering Rules:**

Filtering Rules: The extracted rules are filtered based on confidence and lift to retain the most significant rules.

- **Saving Results:**

Exporting Rules: The filtered rules are saved to a CSV file, allowing for further analysis or sharing of the results later.

- **Conclusion:**

Market basket analysis is a powerful tool for understanding customer buying patterns, offering valuable insights that can be used to improve various aspects of a business. From product recommendations to inventory management and marketing strategies, analytics enables businesses to make data-driven decisions that enhance operational efficiency and customer experience. By applying the results of the analysis, retailers can optimize performance and increase revenue, leading to long-term success in the market.