**MARKET BASKET INSIGHTS**

**DATA PREPROCESSING AND ANALYSIS**

# 1. Data Preprocessing

## 1.1 Data Loading and Inspection

The dataset was loaded from a CSV file with 7 attributes and 522,060 rows. The ‘Date’ column was parsed as datetime during loading.

## 1.2 Data Cleaning

• Changed the 'Price' column's commas to dots and changed its data type to float64.

• Verified and dealt with missing values: 'Date' and 'Price' columns have no missing values. Rows with missing 'Itemname' were removed.

• Looked out for odd values: Rows with a negative quantity and a zero price were removed.

• Examined the data for discrepancies: Removed rows where the 'BillNo' field included non-digit values.

• Rows including "Itemname" and "Adjust bad debt" were removed because they were irrelevant for market basket analysis.

**1.3 Feature Engineering**

* Created a new column ‘Total\_Price’ by multiplying ‘Quantity’ and ‘Price’.

## 1.4 Data Transformation

• Converted the ‘Date’ column to datetime format using the format “%d.%m.%Y %H:%M”.

# 2. Data Analysis

## 2.1 Overview

• December 2010 through December 2011 are included in the dataset.

•The dataset contains 30 distinct countries, with the majority of transactions coming from the United Kingdom (93.4%).

## 2.2 Income Analysis

• Plotted monthly income for the years 2010 and 2011. Income shows a steady increase over time.

## 2.3 Product Analysis

• Recognized and displayed the top 10 items by number and count that were sold.

• Recognized and dealt with deceptive data: To ensure correct analysis, rows with zero prices were eliminated.

# 3. Conclusion

Cleaning the dataset, addressing missing values, deleting inconsistent and misleading data, and transforming features for analysis were all part of the data preprocessing procedure. Insights into the top-selling goods and monthly income trends were offered by the analysis, which is useful for analyzing the market basket and making strategic business decisions.