

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 522,065 rows. The 'Date' column was parsed as datetime during loading.

1.2 Data Cleaning

- Replaced commas with dots in the 'Price' column and converted it to float64 data type.
- Checked and handled missing values: No missing values in 'Date' and 'Price' columns. Dropped rows where 'Itemname' was missing.
- Checked for unusual values: Removed rows with negative quantity and zero price.
- Checked for inconsistent data: Removed rows with non-digit values in the 'BillNo' column.
- Removed rows with 'Itemname' containing "Adjust bad debt" as it was not relevant 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

- The dataset spans from December 2010 to December 2011.
- The number of unique countries in the dataset is 30, with the majority of transactions 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

- Identified and plotted the top 10 most sold products by quantity and count.
- Identified and handled misleading data: Rows with zero price were removed to ensure accurate analysis.

3. Conclusion

The data preprocessing phase involved cleaning the dataset, handling missing values, removing inconsistent and misleading data, and transforming features for analysis. The analysis provided insights into the monthly income trends and the top-selling products, which are valuable for market basket analysis and business decision-making.