About the Dataset

The UCI Online Retail dataset is a transactional dataset from an online retail store in the UK, containing a comprehensive record of sales transactions over a period from 2010 to 2011. It is widely used for analyzing customer behavior and segmentation.

Key Features:

- ○ InvoiceNo: Invoice number-Unique identifier for each transaction.
- ○ StockCode: Unique product code - Unique identifier for each product.
- ○ Description: Product description - Description of the product sold.
- ○ Quantity: Number of items purchased - Quantity of the product sold in the transaction.
- ○ InvoiceDate: Purchase date - Date and time of the transaction.
- ○ UnitPrice: Price of each unit
- ○ CustomerID: Unique identifier for each customer
- ○ Country: Customer's country

Size: The dataset contains 541,909 rows and 8 columns.

Data Quality:

- The dataset contains missing values, especially in the CustomerID column, which may affect customer-level analyses.
- Some transactions may include negative quantities, indicating returns or adjustments that should be handled appropriately during analysis.

Purpose:

The dataset serves various analytical purposes, including:

- **Customer Segmentation:** Identifying distinct customer groups based on purchasing behavior using techniques like Self-Organizing Maps (SOM).
- **Market Basket Analysis:** Understanding product combinations frequently purchased together.
- **Sales Trend Analysis:** Analyzing sales over time to identify peak periods and seasonal trends.
- **Product Performance Evaluation:** Evaluating the performance of individual products or categories to inform inventory management and marketing strategies.

Conclusion:

The UCI Online Retail dataset is a rich source of information for exploring customer behaviors and sales patterns, making it a valuable asset for businesses aiming to optimize their marketing strategies and improve customer experiences.