# Customer Sentiment and Trends: Analyzing Online Retail Data

Aditya Gambhir (Student ID: 862466997)
Omkar Kadam (Student ID: 862467471)

## Problem Statement

The goal of this project is to analyze customer behavior patterns and trends in online retail transactions, with a focus on understanding customer segmentation and product association. By leveraging data mining techniques such as clustering and association rule learning, we aim to uncover insights into customer purchase behavior, identify popular products, and discover frequently purchased product bundles. This analysis can provide valuable insights for designing better marketing strategies and improving customer satisfaction.

## Why is this Problem Important?

Understanding customer behavior and segmenting customers based on their purchasing patterns is crucial for businesses to optimize marketing strategies and enhance customer satisfaction. Insights from customer segmentation can help retailers design targeted marketing campaigns, improve customer retention, and increase sales. Additionally, identifying frequently purchased items together can improve cross-selling strategies and product recommendations, directly impacting a retailer's profitability.

## Related Work

- **Chen et al. (2012)**, "Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining"
  This case study presents the application of data mining techniques like clustering and decision tree induction to create customer segments based on Recency, Frequency, and Monetary (RFM) values. The study highlights how businesses can use data mining to understand customer behavior better and target their marketing efforts accordingly [**?**].

- **Kumar & Reinartz (2006)**, "Customer Relationship Management: A Databased Approach"
  This book discusses how businesses can leverage CRM data to improve their marketing strategies. It emphasizes the importance of understanding customer profitability through segmentation techniques.

- **Collica (2007)**, "CRM Segmentation and Clustering Using SAS Enterprise Miner"
  The book outlines the use of clustering techniques to segment customers for more personalized marketing strategies. The insights gained from such analysis can guide the development of more effective business practices.

## Plan Outline

**Phase 2 Tasks:**

- **Data Cleaning:**

  - Clean the dataset to handle missing values (e.g., missing CustomerID or UnitPrice).

  - Remove duplicate transactions and filter out any canceled orders (indicated by 'c' in InvoiceNo).

- **Exploratory Data Analysis (EDA):**

- Conduct initial analysis to explore transaction frequency, top-selling products, and customer demographics.
- Visualize patterns such as seasonal product demand, popular products, and customer distribution by country.

- **Feature Engineering:**

  - Engineer new features such as total spending per customer, average purchase value, and transaction frequency per customer.

**Phase 3 Tasks:**

- **Clustering (K-Means):**

  - Use K-Means clustering to segment customers based on purchasing behavior (e.g., frequency of transactions, total expenditure, average basket size).
  - Analyze the customer segments to identify high-value and low-value customers.

- **Association Rule Learning (Apriori Algorithm):**

  - Apply the Apriori algorithm to discover association rules between products that are frequently purchased together.
  - Identify the most common product bundles to suggest cross-selling opportunities.

- **Visualization:**

  - Create visualizations for customer segments, purchasing patterns, and frequently bought products to make the results easy to understand.

# Workload Distribution

- **Aditya:**

  - Focus on data cleaning and handling missing values.
  - Prepare feature engineering tasks, such as calculating customer lifetime value and average transaction values.
  - Apply the Apriori algorithm and handle the final visualization of product associations.

- **Omkar:**

  - Lead exploratory data analysis (EDA) and visualizations for customer behavior.
  - Assist with clustering tasks and ensure data quality before modeling.
  - Implement K-Means clustering and interpret the results to identify distinct customer segments.

# Dataset

- **Dataset Name:** UCI Online Retail Dataset

  - **Link:** UCI Repository
  - **Description:** The dataset contains over 541,909 transactions from an online retail store based in the UK, with data ranging from 01/12/2010 to 09/12/2011. The fields include invoice numbers, stock codes, product descriptions, quantities, invoice dates, unit prices, customer IDs, and country information.
  - **Size:** 541,909 rows and 8 columns.
  - **Why it's interesting:** This dataset captures real-world customer transaction behavior, providing rich multivariate and time-series data suitable for clustering, association rule learning, and pattern recognition.

# Evaluation Plan

- **Evaluation Metrics:**
  - **Clustering:** Evaluate clustering quality using silhouette score and intra-cluster variance to ensure well-separated customer segments.
  - **Association Rules:** Measure the strength of the discovered association rules using support, confidence, and lift to identify meaningful product bundles.
  - **Visualization:** Assess the clarity and effectiveness of visualizations in conveying insights about customer segmentation and product trends.

# References

[1] Liu, Bing. *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers, 2012.

[2] Das, Amitava, and Björn Gambäck. "Customer Sentiment Analysis Using Social Media Data." arXiv preprint arXiv:2007.06544, 2020.

[3] Heydari, Abbas, and Mohammad Reza Mohammadi. "Data Mining Techniques in E-commerce: A Study on Customer Sentiment Analysis." IEEE, 2018.

[4] Russell, Matthew A. *Mining the Social Web.* O'Reilly Media, Inc., 2013.

[5] "Sentiment Analysis of E-commerce Product Reviews." Towards Data Science, 2020.