

## **Project: Customer Segmentation Analysis**

**Tools Used:** Python (with relevant libraries such as scikit-learn, pandas, and matplotlib)

### **Industry Context:**

#### **1. The Growing Need for Personalization:**

- In today's highly competitive market, understanding customer behaviour and preferences is crucial for businesses to offer personalized services and products.
- Effective customer segmentation allows companies to tailor marketing strategies, enhance customer experience, and ultimately increase revenue.

#### **2. Challenges Addressed by Customer Segmentation:**

- Identifying distinct customer groups from large datasets.
- Enhancing marketing strategies based on customer segments.
- Improving product recommendations and customer retention.

### **Project Objective:**

#### **1. Understand Customer Behaviour:**

- The primary goal was to segment customers based on their purchasing behaviour and preferences.
- Utilize unsupervised learning models to identify natural groupings within the customer base.

#### **2. Actionable Insights for Business Strategy:**

- Provide recommendations for targeted marketing campaigns.
- Identify potential high-value customer segments for loyalty programs.

### **About Data:**

#### **1. Customer Dataset:**

- In this case study, our task was to identify major customer segments on a transactional data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.
- The dataset includes information such as purchase history, demographics, customer feedback, and interaction with marketing channels.

#### **2. Key Features:**

- **Invoice No:** Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation.
- **Stock Code:** Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.
- **Description:** Product (item) name. Nominal.
- **Quantity:** The quantities of each product (item) per transaction. Numeric.
- **Invoice Date:** Invoice Date and time. Numeric, the day and time when each transaction was generated.
- **Unit Price:** Unit price. Numeric, Product price per unit in sterling.
- **Customer ID:** Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.
- **Country:** Country name. Nominal, the name of the country where each customer resides.

### Data Preprocessing:

#### 1. Data Cleaning:

- Handled missing values using techniques like mean imputation or by removing records with excessive missing data.
- Normalized and standardized numerical features to ensure uniformity in model input.

#### 2. Feature Engineering:

- Created new features such as customer lifetime value (CLV), recency-frequency-monetary (RFM) scores, and customer engagement scores.
- Binned age groups to better understand demographic segmentation.

### Modelling Approach:

#### 1. Unsupervised Learning Models:

- **K-Means Clustering:** Used to segment customers into distinct groups based on purchasing behavior.
- **Hierarchical Clustering:** Employed to validate and explore the hierarchical relationships among customer segments.

#### 2. Model Selection and Evaluation:

- **Elbow Method:** Used to determine the optimal number of clusters.
- **Cluster Profiling:** Interpreted the characteristics of each cluster to derive actionable insights.

### Key Results:

#### 1. Customer Segmentation:

- Identified three primary customer segments:
  - **High-Value Loyal Customers:** High frequency of purchases, high engagement, and high CLV.
  - **Price-Sensitive Occasional Buyers:** Less frequent purchases, high sensitivity to promotions.
  - **Potential High-Value Customers:** Moderate engagement, potential to be converted into loyal customers through targeted campaigns.

## 2. Cluster Analysis:

- Detailed analysis of each cluster's demographic and behavioral traits.
- Recommended personalized marketing strategies for each segment.

## Business Impact:

### 1. Enhanced Marketing Efficiency:

- Targeted campaigns based on customer segments, leading to improved conversion rates.

### 2. Customer Retention:

- Loyalty programs tailored to high-value customers, increasing retention rates.

### 3. Revenue Growth:

- Optimized promotional strategies for price-sensitive customers, driving incremental sales.

## Expected Interview Questions:

1. How did you determine the optimal number of clusters?
2. Can you explain the significance of the Silhouette Score in evaluating cluster quality?
3. What were the main challenges you faced in customer segmentation and how did you address them?
4. How do your segmentation results influence marketing strategies and business decisions?
5. What continuous improvement strategies would you recommend based on your findings?
6. Can you describe the main objectives of your project in the Marketing and sales domain?
7. What specific challenges do retail companies face that data analytics can help address?

8. Could you walk us through the steps you took in preprocessing the customer datasets for analysis?
9. How did you handle missing values in the dataset? Why did you choose those specific methods (mean, median, mode)?
10. What types of feature engineering did you perform on the dataset, and how did it enrich the analysis?
11. Can you explain some of the actionable insights you derived from the EDA you developed?
12. How did you identify peak times for purchases and most productive period which causes using your analysis?