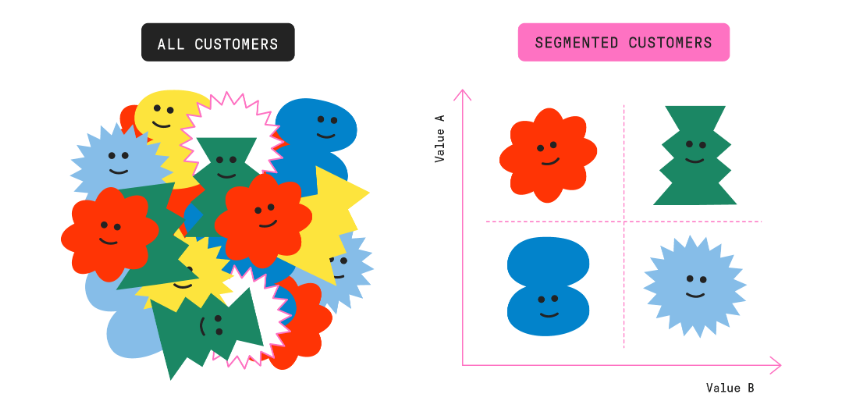
**CUSTOMER SEGMENTATION ANALYSIS REPORT**

**What is Customer Segmentation?**

* Customer segmentation is the process of dividing a company's customers into distinct groups based on shared characteristics.
* This can include demographics, behaviors, purchasing habits, or other criteria.
* The goal is to tailor marketing strategies and improve customer service by understanding and addressing the specific needs and preferences of each segment.
* By doing so, businesses can enhance customer satisfaction, increase loyalty, and drive sales growth.
* It's a powerful strategy for making data-driven decisions and optimizing marketing efforts.



**Table of Contents:**

1. Introduction

* Project Overview
* Objectives

1. Dataset Description
2. Exploratory Data Analysis (EDA)
3. Feature Engineering
4. Customer Segmentation using K-Means Clustering
5. Cluster Profiling
6. Marketing Strategies
7. Conclusion
8. References

**1.Introduction:**

**Project Overview:**

* In today’s highly competitive retail landscape, understanding customer behaviour is paramount.
* This project aims to leverage data-driven techniques for customer segmentation using an online retail dataset.
* The dataset encompasses transactional data including InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country.

**Objectives:**

1. **Data Exploration and Cleaning**: To prepare the data for analysis.
2. **Feature Engineering**: To derive meaningful features such as Recency, Frequency, and Monetary value from the transactional data.
3. **Customer Segmentation**: To classify customers into distinct segments based on their purchasing behavior using clustering algorithms like K-Means.
4. **Cluster Profiling and Strategy Development**: To profile the resultant clusters and propose targeted marketing strategies for each segment.

**2.Dataset Description**

* **Data Source**:  The dataset was obtained from “Kaggle”
* **Description of Features**: Detailed description of each column in the dataset, including InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country.

**3.Exploratory Data Analysis**:

Exploratory Data Analysis (EDA) was conducted to understand the structure, distribution, and relationships within the dataset, leading to these key findings:

1. **Descriptive Statistics**:
   * **Total Transactions**: The dataset contains transactions made by various customers.
   * **Customer Count**: Unique Customer IDs indicate the total number of individual buyers.
   * **Product Diversity**: Various unique product descriptions hint at a wide range of products sold.
2. **Data Cleaning**:
   * **Missing Values**: Identified and handled missing values, especially in critical columns like CustomerID and InvoiceDate.
   * **Duplicates**: Removed duplicate entries to ensure data integrity.
3. **Revenue Analysis**:
   * **Total Revenue**: Summed up the total revenue generated by multiplying Quantity and UnitPrice.
   * **Revenue by Country**: Identified the top-performing countries in terms of revenue, with the UK leading significantly.
4. **Sales Trends:**

* **Monthly Sales**: Observed sales trends over time, highlighting peak periods and seasonal variations.
* **Day of the Week Analysis**: Identified the days with the highest and lowest sales, providing insights into customer shopping behaviour.

1. **Top Products:**

* **Most Popular Products:** Analysed the most frequently bought items to identify

Customer preferences.

* **Revenue-Generating Products:** Highlighted the products that contributed the most to overall revenue.

1. **Customer Insights:**

* **Purchase Frequency**: Distribution of how often customers make purchases.
* **Customer lifetime value (CLV)**: Distribution of the total revenue generated by each customer.

1. **Visualization of Findings:**

* **Bar Graphs**: Used to display total sales by country and top products.
* **Line charts:** Showed sales trends over time.
* **Histograms:** Depicted purchase frequency
* **Pie charts:** Used to display top Customers by Revenue
* **Box plot and Violin plots:** Illustrated the distribution of the key metrics like Monetary, Frequency, and Recency and CLV distribution.

**4.Feature Engineering:**

We derived three key features essential for customer segmentation:

1. **Recency**: The number of days since the customer's last purchase.
2. **Frequency**: The total number of purchases made by each customer.
3. **Monetary**: The total revenue generated by each customer.

**Process**

* Calculated the Recency by finding the difference between the most recent date in the dataset and the customer's last purchase date.
* Determined Frequency by counting the unique InvoiceNos per customer.
* Computed Monetary by summing the total revenue from each customer's purchases.

**5.Customer Segmentation**

We used K-Means clustering to segment customers based on the features derived.

**Steps:**

1. **Standardized the Data**: Used StandardScaler to standardize the Recency, Frequency, and Monetary values.
2. **Optimal Number of Clusters**: Determined using the Elbow Method, which suggested the optimal number of clusters.
3. **Applied K-Means**: Clustered the customers into groups and added the cluster labels to the dataset.

**6**.**Cluster profiling:**

Profiling each cluster to understand its characteristics:

1. **Cluster 0 - High-Value Customers**:
   * These customers have low recency, high frequency, and high monetary values.
2. **Cluster 1 - At-Risk Customers**:
   * High recency, low frequency, and moderate monetary values indicate they haven't purchased recently but have done so in the past.
3. **Cluster 2 - Potential Loyalists**:
   * Low recency, moderate frequency, and monetary values suggest they have recently started purchasing and could become loyal customers.
4. **Cluster 3 - Casual Shoppers**:
   * Moderate recency and low frequency with low monetary values indicate infrequent buyers.

**7.Marketing Strategies:**

Based on the cluster profiles, tailored marketing strategies were developed:

1. **High-Value Customers:** Offer exclusive loyalty programs, early access to new products, and personalized recommendations.
2. **At-Risk Customers**: Use re-engagement campaigns, win back offers, and personalized follow-ups.
3. **Potential Loyalists:** Encourage repeat purchases with targeted promotions, product recommendations, and nurturing campaigns.
4. **Casual Shoppers:** Provide occasional discounts, reminders about sales events, and personalized content.

**8.Conclusion:**

* This customer segmentation analysis provides a robust foundation for tailored marketing strategies, aiming to enhance customer satisfaction, boost sales, and improve overall business performance.
* Future steps could include regular monitoring of customer segments and adjusting strategies based on evolving trends and feedback.

**9.References:**

* **Tools and Libraries Used**: List of tools and libraries used in the analysis, such as pandas Matplotlib, scikit-learn, Seaborn

-THE END-