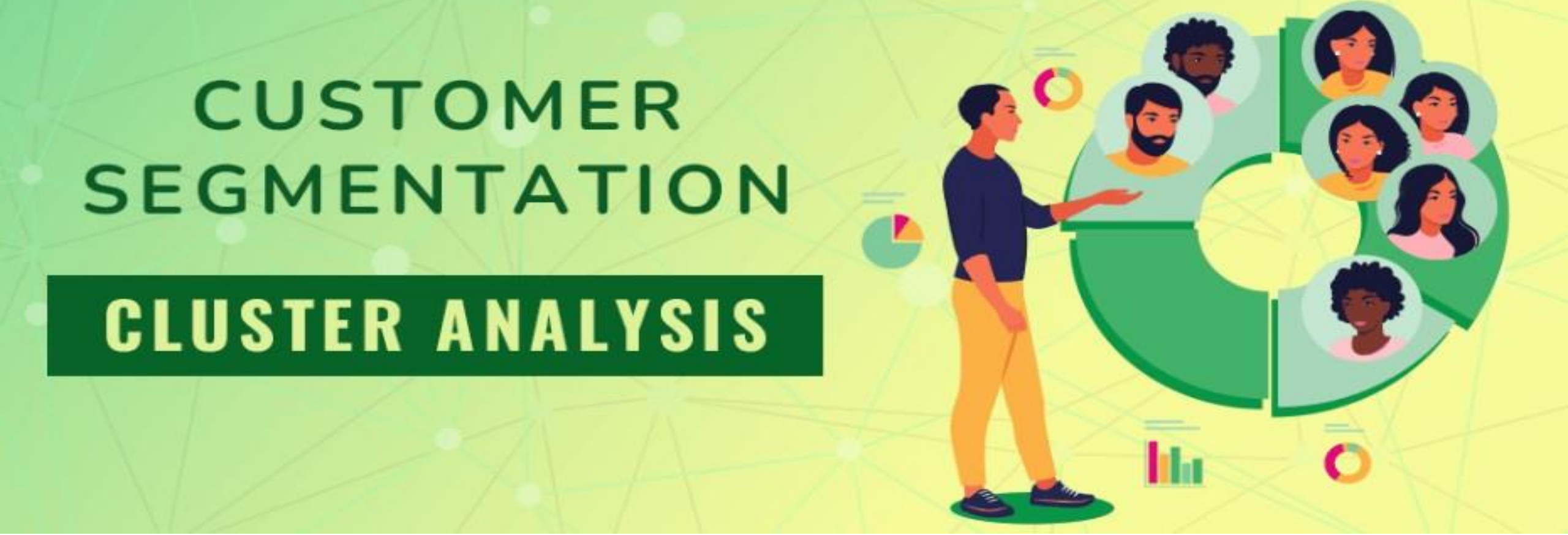


Customer Segmentation Behaviour Analysis using Unsupervised Machine Learning

Presented by : Kehinde Ogundana
(Data Scientist)

Business Introduction

Company: E-Shop Pro
Industry: E-Commerce

E-Shop Pro is a leading online retailer offering a wide range of products, including electronics, apparel, and home goods. With millions of customers globally, the business prides itself on its fast delivery, customer-centric policies, and personalized shopping experience. Some major achievements highlighted by the company include:

- Growth in user base by 30% annually over the last five years.
- Implementation of a personalized recommendation system, improving customer satisfaction.
- Achieved 80% retention rate among VIP customers in the first two years of loyalty program implementation.

Despite its numerous successes, the company has encountered a formidable challenge that has left its leadership team determined to find a solution: an alarmingly high shopping cart abandonment rate.

Business Problem

The company is facing a significant challenge in retaining customers over time. While the company experiences high customer acquisition rates, they are seeing a drop-off in returning customers after the first purchase. Specific challenges include:

- **Declining repeat purchase rate**: New customers are not returning after their initial purchase.
- **Customer churn**: A noticeable percentage of customers leave after a brief period of activity, increasing marketing and acquisition costs.
- **Low engagement**: Customers show decreased interaction with the platform after their first few visits, indicating a failure to maintain long-term interest.

Resolving these issues is critical for improving customer lifetime value (CLV) and overall business growth.

Project Objectives:

- **Retention Rate Analysis:** Measure and evaluate customer retention rates across different cohorts over time.
- **Customer Segmentation**: Group customers into distinct clusters based on shared characteristics or behaviors, utilizing methods like RFM analysis.
- **Recommendation Strategies:** Develop data-driven recommendations to enhance customer retention, leveraging insights from cohort analysis.

Data Description

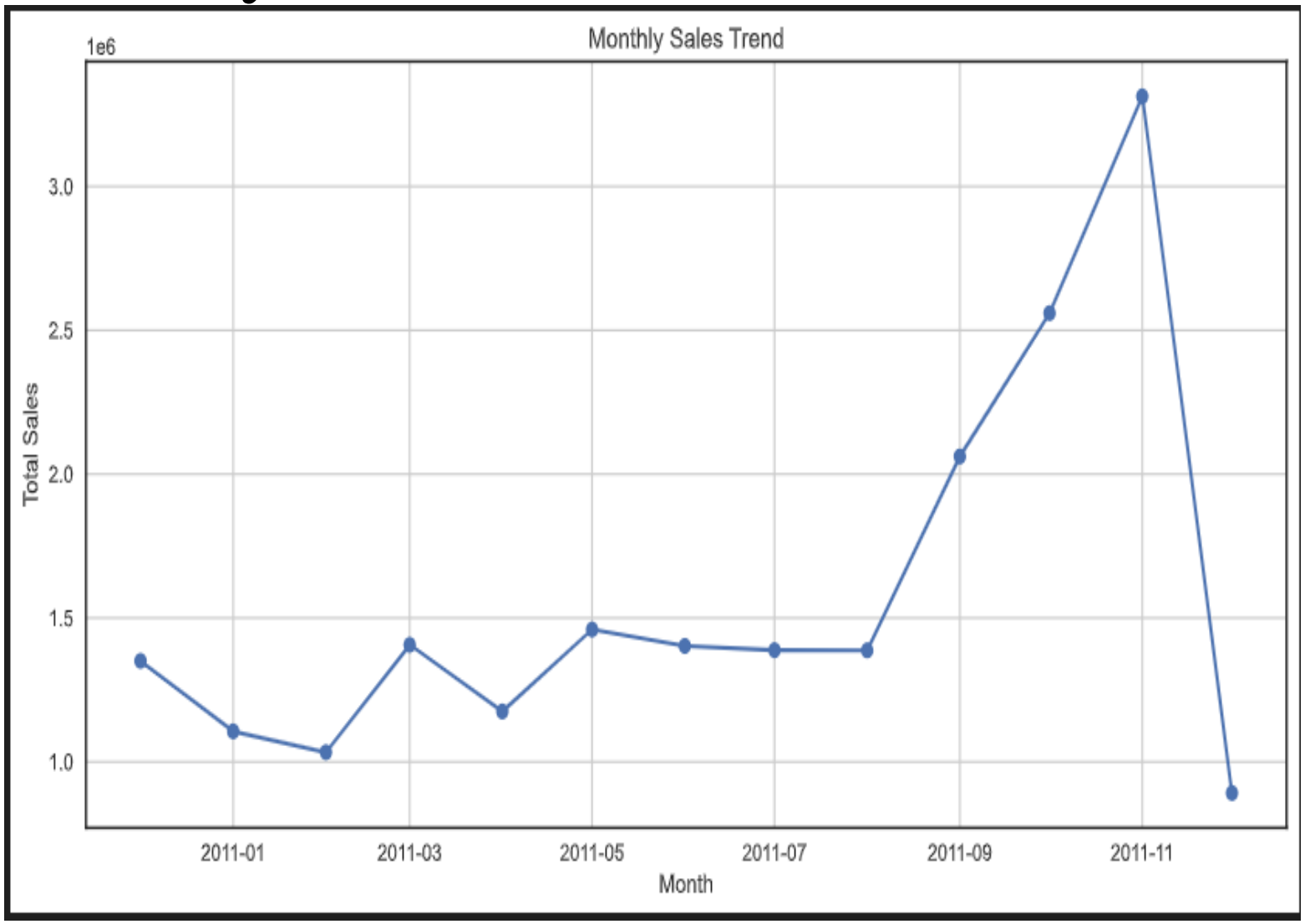
- The dataset contains 541,909 rows and 8 features, including purchase details such as Invoice, Invoice Date, Customer ID, Unit Price, Quantity, and Country.

- Missing Data & Duplicates :
  - No duplicate records were found in the dataset.
  - The Customer ID column had 25% missing values.

Data Preprocessing

- **Handling Missing Data**:
  - Missing Customer ID values were dropped since this column is crucial for customer segmentation and recommendation systems. Imputing such a large percentage of missing data could introduce bias, potentially compromising the integrity of the clustering analysis.
  - After removing missing values, 75% of the data remained available for analysis.

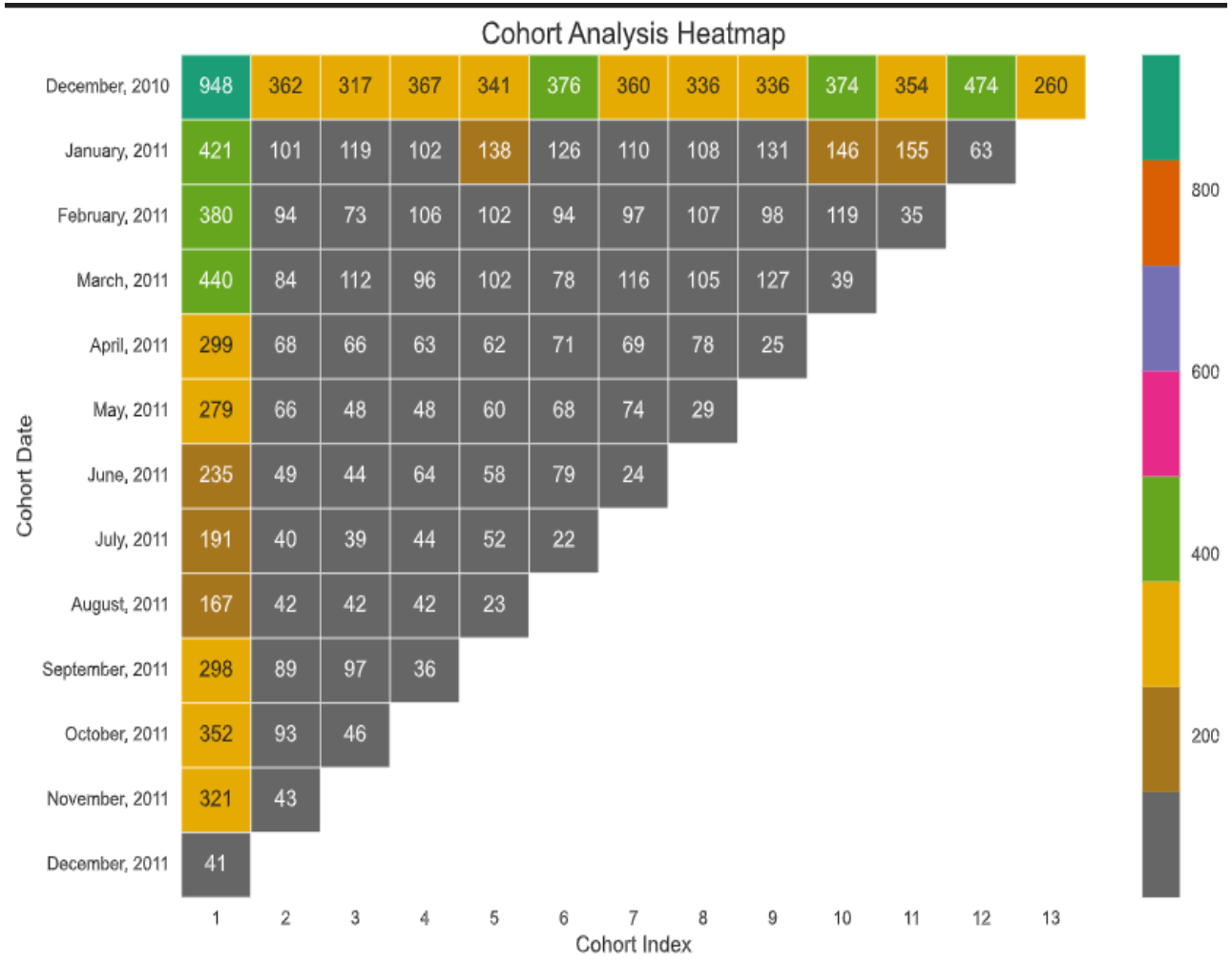
Exploratory Data Analysis (EDA)
Monthly Sales Trend



The sales trend analysis reveals clear seasonal patterns:

- A significant surge in sales occurs in the second half of the year, particularly from September to November 2011, likely driven by holiday shopping or promotional activities.
- A sharp decline follows from December 2011 to February 2011, possibly due to reduced activity after the holiday peak.

Exploratory Data Analysis (EDA)
Cohort Analysis Heatmap



- Acquisition Trends

The acquisition rate started high (44.4% in January 2011) but fluctuated over time. The highest acquisition month was March 2011 (46.4%), while the lowest was December 2011 (4.3%).

- Retention Trends

The highest retention was seen in the second month (38.2% for December 2010 cohort).Retention gradually dropped in later months, indicating that customer engagement decreased over time.

Feature Engineering, Selection, and Dimensionality Reduction

Feature Engineering:

Recency: measures how recently a customer has made a purchase, providing insights into their engagement with the business.

- **Frequency:** Two features were engineered ;
  - **Frequency:** Represents the total number of unique transactions made by a customer. This helps assess the level of customer interaction with the retailer.
  - **Total Products Purchased:** Represents the total quantity of products purchased by a customer across all transactions, offering insights into their buying behavior and purchase volume.

- **Monetary:** Two features were engineered ;
  - **Monetary:** This metric measures the total amount spent by each customer.
  - **Average Purchase Value:** It highlights the average spending per transaction and offers insights into a customer's purchasing behavior, helping to personalize offers and marketing campaigns.

Feature Engineering, Selection, and Dimensionality Reduction(PCA)

Feature Selection :

Correlation Heatmap Matrix : reveals that recency, frequency and Toal products are correlated.

Feature Scaling and PCA:

- Scaling: Here, it's crucial to scale the features, as this is vital for distance-based methods like K-means and PCA and features with larger ranges can dominate, leading to biased clustering.

- PCA: this reveals that three features can account for 84% can explained the variance to perform the clustering.

Model Selection and Building

Optimal Clusters Analysis:

To ascertain the optimal number of clusters (k) for segmenting customers, two renowned methods are used ; Elbow Method and Silhouette Method. Both the Elbow Method and the Silhouette Score suggest that optimal k =3.

Models:

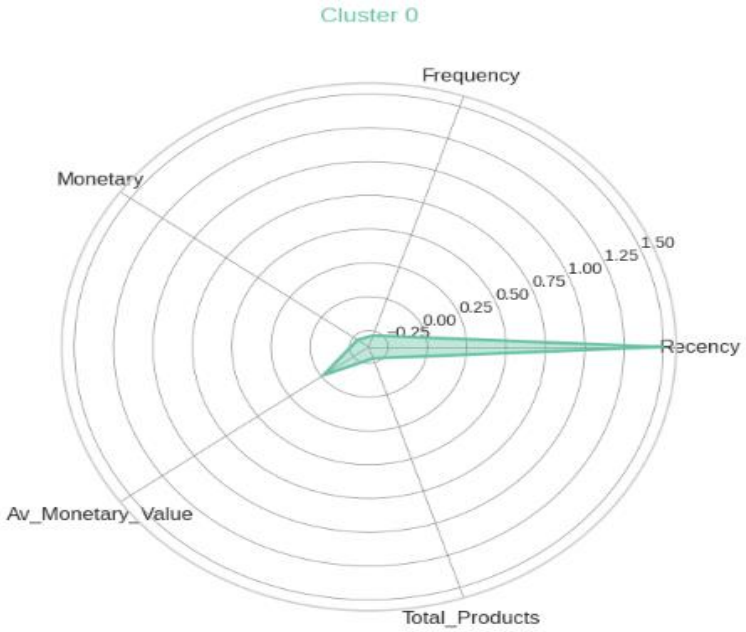
To determine the best clustering models, evaluation was performed using the **Silhouette Score**, which measures how well-separated and compact the clusters are. The results indicate that K-Means is the optimal model, achieving the highest Silhouette Score of 61%, signifying well-defined and distinct clusters.

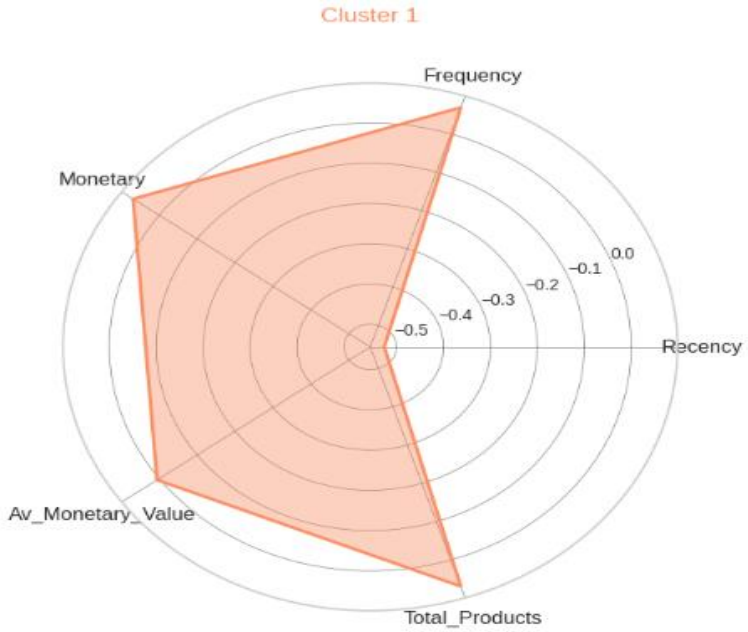
	Model	Silhouette Score
0	KMeans	0.606087
1	Agglomerative	0.596841
2	Hierarchical	0.596841

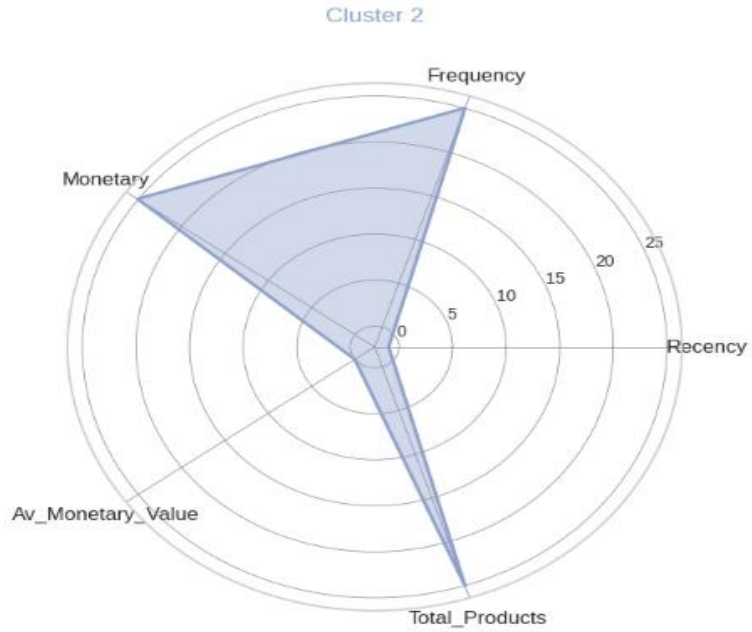
Best Model – Kmeans Clustering

Optimal Cluster Visualization:

cluster	Recency	Frequency	Monetary	Av_Monetary_Value	Total_Products
0	1.492959	-0.283292	-0.283007	-0.015794	-0.283112
1	-0.526058	0.069016	0.068845	0.005543	0.068914
2	-0.964201	25.042252	25.099297	0.028102	25.073298







Clusters Interpretation & Recommendation

- **Cluster 0: Inactive or Churned Customers:**

- Looking at the very high recency ratio(1.49), these customers haven't purchased in a long time. Hence not surprising that the frequency, monetary and total products are below zero as this means they purchase less from us, spend less, and buy fewer products.

➢ **Recommendation to Regain Cluster 0 Customers (Win-Back Strategy)**

- **Reactivation Campaign**: Send personalized offers, discounts, or limited-time promotions to encourage them to return.
- **Email or SMS Reminders**: Highlight new products or special loyalty perks.
- **Loyalty Program Incentives**: Offer points or cashback for their next purchase.
- **Surveys & Feedback Collection**: Understand why they stopped buying and address concerns.

Clusters Interpretation & Recommendation

- **Cluster 1: Average or Regular Customers:**

- The recency ratio(0.526) shows that they have purchased relatively recently, these customers purchase at an average rate, spend a typical amount, and buy a standard number of products.

➢ **Recommendation to Maintain Cluster 1 Customers (Retention Strategy)**

- **Personalized Upselling & Cross-Selling**: Recommend complementary products based on past purchases.
- **Exclusive Early Access**: Provide early access to new products or special deals.
- **Engagement Campaigns:** Keep them interested with newsletters, product recommendations, and seasonal discounts.
- **Loyalty Rewards:** Encourage repeat purchases with discounts for consistent buyers.

Clusters Interpretation & Recommendation

- **Cluster 2: High-Value, Loyal Customers:**

- The very low recency ratio(-0.964) shows that these customers buy frequently and recently, these customers purchase very often, spend significantly more, and buy a lot of products.

- **Recommendation to Retain & Strengthen Cluster 2 (VIP Customers)**

- **Exclusive VIP Programs:** Offer premium services, VIP membership, or free shipping to reward loyalty.
- **Early Access & Priority Support:** Give them first access to new collections, exclusive deals, or dedicated customer support.
- **Premium Upselling**: Introduce high-end or personalized product recommendations.
- **Appreciation & Recognition:** Send thank-you emails, personalized discounts, or even surprise gifts to strengthen their connection with the brand.

By applying these plans, the business can maximize customer retention, encourage repeat purchases, and re-engage lost customers effectively.

Further Recommendation

- **Healthy Retention Rate:** E-commerce platforms typically have a retention rate between 20% to 40%, meaning that this percentage of customers continue purchasing after their first transaction.

- **Factors Driving High Retention (Dec 2010):**
  - Analyze key contributors to the high retention rate for the December 2010 cohort.
  - Identify successful marketing campaigns, product improvements, or customer engagement strategies that worked.
  - Replicate these strategies for other cohorts to boost retention.
- **Investigate December 2011 Drop:**
  - Identify why retention rates were low in December 2011 across all cohorts.
  - Examine customer feedback, product quality, customer service, or business operation changes during that period.
- **Continuous Monitoring & Adaptation:**
- Retention rates fluctuate due to various factors, requiring ongoing monitoring and strategy adjustments.
- Regular cohort analysis and tracking customer behavior help in making data-driven decisions to improve retention.