CAPSTONE PROJECT ON

ONLINE RETAIL CUSTOMER SEGMENTATION

TEAM:

1. Narra Sai Keerthana Reddy
2. Bhanu Shahi

CONTENTS

CHAPTER 1: INTRODUCTION TO CUSTOMER SEGMENTATION

1.1 Introduction to Customer Segmentation:

Customer segmentation is the process of categorizing a diverse customer base into distinct groups or segments based on certain shared characteristics, behaviors, or attributes. The aim is to identify patterns within the customer data that allow businesses to tailor their marketing strategies, improve customer experiences, and optimize their overall approach to different customer groups. This division enables companies to target their efforts more effectively, addressing the specific needs and preferences of each segment.

1.2. Importance of Customer Segmentation:

- Personalized Marketing: Customer segmentation enables businesses to create personalized marketing campaigns for each segment. By understanding the unique preferences and behaviors of different customer groups, companies can tailor their messages and promotions, increasing the likelihood of engagement and conversion.
- Enhanced Customer Experience: Segmentation helps in providing a more personalized and relevant experience to customers. By tailoring product recommendations, content, and communication based on segment characteristics, businesses can enhance customer satisfaction and loyalty.
- Optimized Resource Allocation: With segmented data, companies can allocate their resources more efficiently. They can focus marketing efforts, promotional activities, and product development on areas that are most likely to generate positive results for each specific customer segment.
- Improved Product Development: Understanding the diverse needs of different customer segments allows businesses to develop products or services that better cater to those specific requirements. This can lead to increased customer satisfaction and loyalty.
- Customer Retention: By identifying high-value customer segments and understanding their needs, companies can implement strategies to retain these customers. This is crucial for long-term business success as retaining existing customers is often more cost-effective than acquiring new ones.

1.3. Fields in Which Customer Segmentation is Used - Applications:

- Retail: In online retail, customer segmentation helps businesses understand the shopping behavior of different customer groups. This information can be used for personalized product recommendations, targeted promotions, and optimizing the layout of e-commerce websites.
- Marketing: Customer segmentation is extensively used in marketing to create targeted and personalized campaigns. It allows marketers to tailor their messages to specific audiences, improving the effectiveness of advertising and promotional efforts.
- Finance and Banking: In the finance sector, customer segmentation is utilized to understand the financial behavior of different customer segments. This information can be crucial for offering personalized financial products, such as loans, credit cards, and investment options.
- Healthcare: In healthcare, patient segmentation can help in tailoring healthcare services to different patient groups. This includes personalized treatment plans, communication strategies, and preventive care initiatives.
- Telecommunications: Customer segmentation is employed in the telecommunications industry to understand the usage patterns of different customer groups. This information can be used for targeted promotions, service customization, and customer retention strategies.
- Hospitality and Travel: In the hospitality sector, customer segmentation is valuable for tailoring travel packages, loyalty programs, and promotional offers to different traveler profiles, enhancing the overall customer experience.

Customer segmentation is a versatile and powerful tool that can be applied across various industries to gain insights into customer behavior, optimize business strategies, and ultimately improve customer satisfaction and loyalty.

1.4. Customer Segmentation in Retail - Specific Importance:

In the retail sector, customer segmentation holds particular significance due to the dynamic and diverse nature of consumer behaviors. Here are some specific reasons why customer segmentation is crucial in retail:

- Targeted Marketing Campaigns: Retailers can create targeted marketing campaigns based on the preferences, buying patterns, and demographics of different customer segments. This ensures that promotional efforts resonate more effectively with specific groups, leading to higher engagement and conversion rates.
- Inventory Management and Stocking Strategies: Segmentation aids in optimizing inventory management by identifying which products are popular among specific customer segments. This information helps retailers adjust stocking levels, plan for seasonal demand, and reduce excess inventory of items that may not be in high demand within certain segments.
- Tailored Pricing Strategies: Different customer segments may respond differently to pricing strategies. Retailers can use segmentation to implement tailored pricing models, discounts, or loyalty programs that cater to the specific preferences and price sensitivity of each segment.
- Customer Retention Strategies: Identifying high-value customer segments allows retailers to implement targeted retention strategies. By understanding the needs and preferences of valuable customers, businesses can offer personalized incentives, loyalty programs, or exclusive promotions to encourage repeat purchases and foster long-term customer relationships.
- Optimized Website and User Experience: E-commerce retailers can use customer segmentation to optimize their website design, navigation, and user experience. Tailoring the online shopping journey based on segment characteristics ensures that customers find relevant products more easily, leading to improved satisfaction and conversion rates.
- Market Basket Analysis: Understanding the purchasing behavior of different segments through market basket analysis helps retailers identify product affinities. This information is valuable for cross-selling and bundling strategies, allowing retailers to suggest complementary products and increase the average transaction value.
- Seasonal and Trend Analysis: Retailers can leverage segmentation to analyze how different customer segments respond to seasonal trends and changes in fashion or preferences. This information is valuable for adjusting inventory, marketing strategies, and product offerings to align with the evolving tastes of each segment.

In summary, customer segmentation in the retail sector is instrumental for tailoring marketing efforts, optimizing inventory and pricing strategies, improving customer retention, and providing a personalized and seamless shopping experience. It empowers retailers to understand and respond to the diverse needs of their customer base, ultimately driving business success in a competitive market.

CHAPTER 2: INTRODUCTION TO CLUSTERING

Segmentation techniques involve methods for grouping data into meaningful clusters or segments. In the context of customer segmentation in retail, clustering techniques are commonly used. Here are some popular clustering techniques and methods for evaluating their performance:

2.1 Clustering Techniques:

2.1.1 K-Means Clustering:

- Description: K-Means is a centroid-based clustering algorithm that partitions data into K clusters. It minimizes the sum of squared distances between data points and their assigned cluster centroids.
- Application: It's widely used for its simplicity and efficiency in various domains, including customer segmentation.
- Considerations: The choice of K (number of clusters) is critical, and the algorithm may converge to local optima.

2.1.2 Hierarchical Clustering:

- Description: Hierarchical clustering builds a tree of clusters, where each node represents a cluster. It can be agglomerative (bottom-up) or divisive (top-down).
- Application: Suitable for datasets with a hierarchical structure or when exploring different levels of granularity in segmentation.
- Considerations: The dendrogram needs to be cut at an appropriate level to obtain meaningful clusters.

2.1.3 DBSCAN (Density-Based Spatial Clustering of Applications with Noise):

- Description: DBSCAN identifies clusters based on dense regions separated by sparser areas. It is effective at discovering clusters of arbitrary shapes.
- Application: Suitable for datasets with irregularly shaped clusters and noise.
- Considerations: Sensitivity to parameter settings, such as epsilon (radius) and minimum points.

2.2 Evaluation Metrics for Clustering:

2.2.1 Silhouette Score:

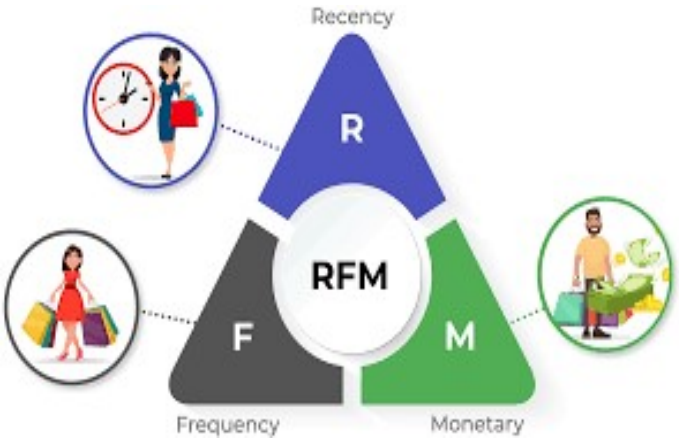
- Measures how similar an object is to its own cluster compared to other clusters. Ranges from -1 to 1, with higher values indicating better-defined clusters.

2.2.2 Davies-Bouldin Index:

- Quantifies the "compactness" and "separation" between clusters. Lower values indicate better clustering.

CHAPTER 3: RFM ANALYSIS

3.1 RFM Analysis



RFM Analysis is a strategic marketing technique commonly used in the retail industry to segment and analyze customer behavior based on three key dimensions: Recency, Frequency, and Monetary Value. This method helps businesses identify and categorize their customers to tailor marketing and sales strategies more effectively. Each component of RFM represents a distinct aspect of customer engagement:

3.1.1 Recency (R):

- Measures how recently a customer has made a purchase. It emphasizes the idea that customers who have recently interacted with the business are more likely to respond positively to marketing efforts.
- Calculation: Determine the time since the customer's last purchase or interaction with the business. This is often represented in terms of days, weeks, or months.
- Segmentation:
    - High Recency: Customers who made a purchase or engaged recently.
    - Medium Recency: Customers with moderate time since their last interaction.
    - Low Recency: Customers who haven't engaged for a longer period.

3.1.2 Frequency (F):

- Measures how often a customer makes a purchase or engages with the business. It indicates the level of customer loyalty and the likelihood of repeat business.

- Calculation: Count the number of transactions or interactions within a specified period.
- Segmentation:
    - High Frequency: Customers who make frequent purchases.
    - Medium Frequency: Customers with moderate purchase frequency.
    - Low Frequency: Infrequent buyers or those who make purchases sporadically.

3.1.3 Monetary Value (M):

- Measures the total monetary value of a customer's purchases or contributions to the business. It reflects the customer's overall value in terms of revenue generation.
- Calculation: Sum the total monetary value of all transactions for each customer.
- Segmentation:
    - High Monetary Value: Customers who contribute significantly to the business's revenue.
    - Medium Monetary Value: Customers with moderate spending.
    - Low Monetary Value: Customers who make smaller or infrequent purchases.

CHAPTER 4: PROJECT

SECTION 1 : Project Overview
SECTION 2: Project Objective
SECTION 3: Data Preprocessing
SECTION 4: Exploratory Data Analysis
SECTION 5: Feature Engineering
SECTION 6: Model Training
SECTION 7: Results
SECTION 8: Conclusion

4.1 Project Overview:

The "Online Retail Customer Segmentation" project aims to enhance the marketing and customer relationship strategies of an online retail business by categorizing its diverse customer base into distinct segments. By leveraging customer segmentation techniques, the project seeks to uncover meaningful patterns, behaviors, and preferences among the customers. This understanding will enable the business to tailor its marketing efforts, optimize resource allocation, and provide personalized experiences to different customer segments.

4.2 Project Objectives:

Segmentation for Targeted Marketing:

- *Objective:* Identify and create meaningful customer segments based on their purchasing behaviors, preferences, and engagement history.
- *Rationale:* This will enable the business to craft targeted marketing campaigns, promotions, and communication strategies for each specific segment.

4.3 Data Preprocessing Steps :

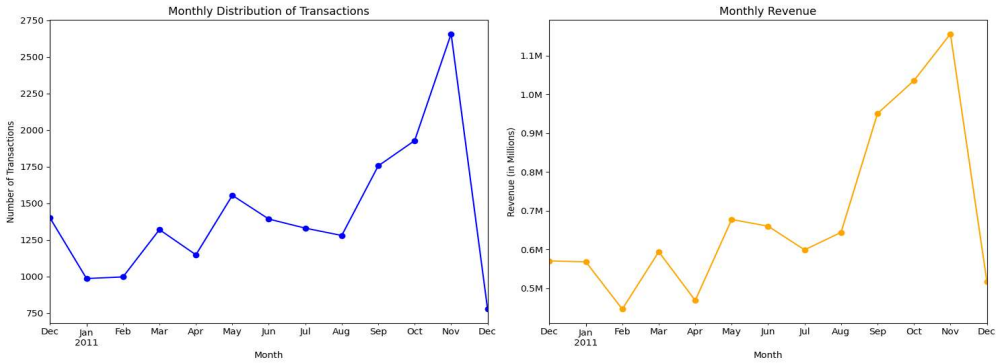
1. Remove duplicates in the dataset to ensure data integrity.
2. Drop rows with null values in identifiers, specifically CustomerID, for accurate customer analysis.
3. Convert data types, ensuring consistency and compatibility for subsequent analyses.
4. Exclude rows corresponding to cancelled transactions as they are irrelevant for segmentation.
5. Handle missing values through appropriate methods to maintain data quality.
6. Drop unnecessary columns that do not contribute to the customer segmentation process.

4.4 Exploratory Data Analysis

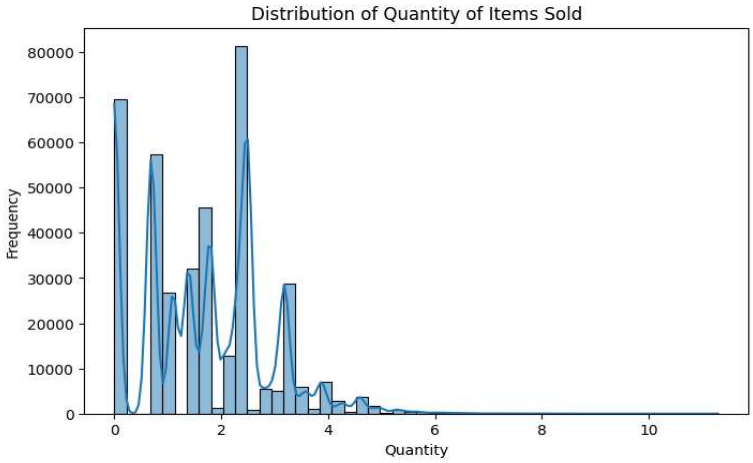
Exploratory Data Analysis (EDA) is a crucial phase in data analysis where the primary focus is on visually and statistically exploring the dataset. By examining patterns, relationships, and anomalies, EDA provides valuable insights into the data's structure and characteristics. It helps in identifying outliers, understanding variable distributions, and assessing data quality. EDA plays a vital role in shaping hypotheses, guiding preprocessing steps, and informing subsequent analyses, contributing to a more informed and effective data-driven decision-making process.

Graphs and Insights

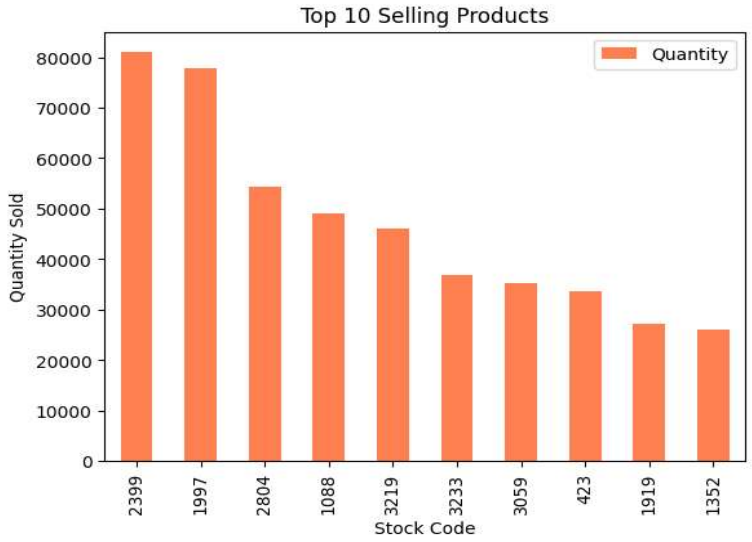
1. Monthly Distribution of Transactions & Monthly Distribution of Revenue



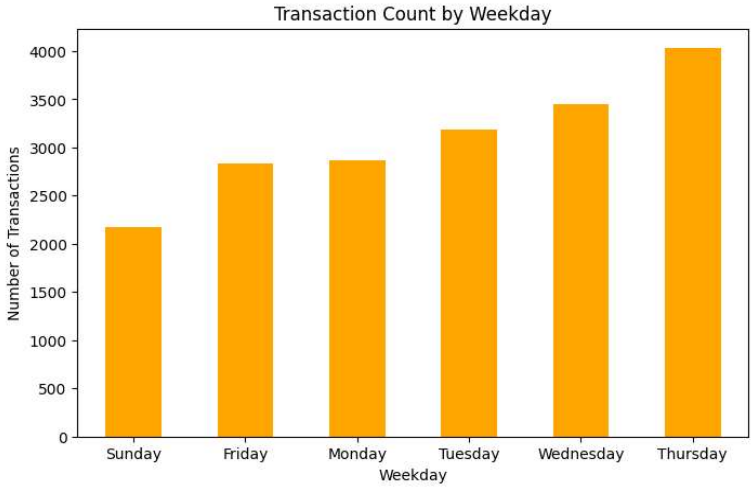
2. Distribution of quantity of item sold in a transaction



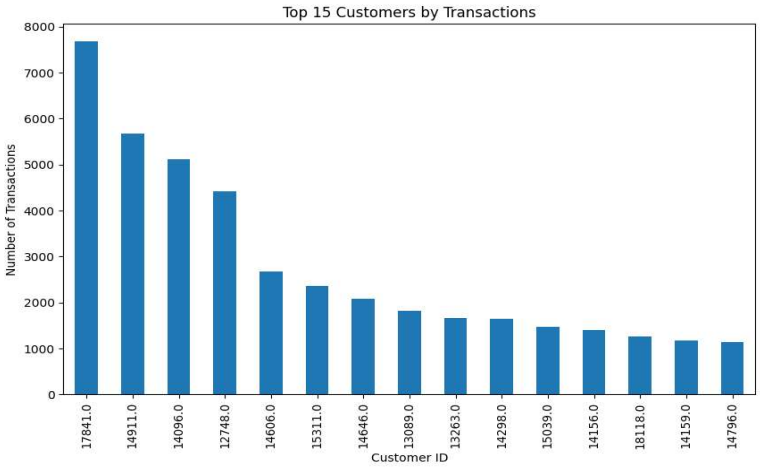
3. Top 10 Selling Products and Quantity Sold



4. Weekday popularity



5. Top 15 customers based on number of transactions

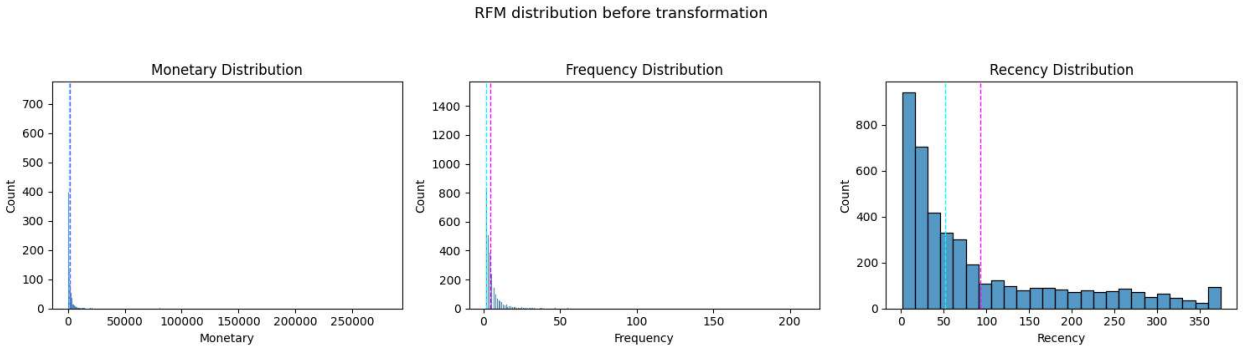


Below are the insights after performing EDA on the dataset taken

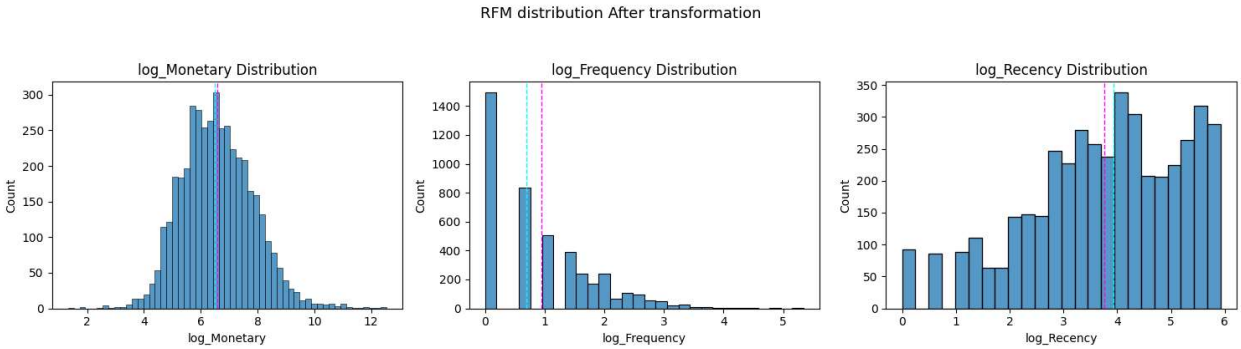
1. Sep, Oct, Nov are where customers are very active. In Sep, Oct customers tendency to buy high priced items compared to in Nov
2. Top 5 countries from where customers are expected from UK, France, Germany, Spain, EIRE.
3. Customer Segmentation based on quantity of items bought is performable as the distribution is too wide.
4. Customer Frequency can be another segmentation
5. No transactions recorded on Saturday. Thursday seems popular to purchases.
6. Top 15 Customers, Top 10 products by transaction count, Top 10 products by revenue are seen.
7. UK purchased mostly in November, France purchased mostly in October, EIRE purchased mostly in March

4.5 Feature Enginerring for RFM Analysis

1. Create Revenue Feature:Calculate revenue by multiplying Quantity and Unit Price.
2. GroupBy Customer ID for RFM Metrics:Obtain Recency, Frequency, and Monetary values by grouping data based on Customer ID.
3. Calculate Recency (R):Extract the number of days from the last purchase date to the maximum date available.
4. Quantile-Based Scoring (1 to 4): Assign quantile scores (1 to 4) to Recency, Frequency, and Monetary values individually.
5. Calculate Total RFM Score: Sum the quantile scores to get the total RFM score.
6. Segmentation: Segment customers based on total RFM scores into Low Risk, High Value; Medium Risk, Medium Value; and High Risk, Low Value segments.

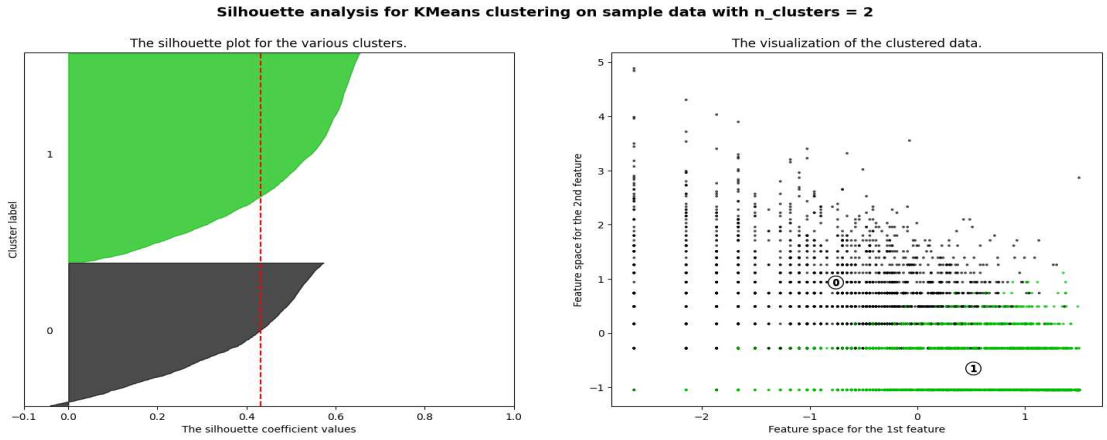


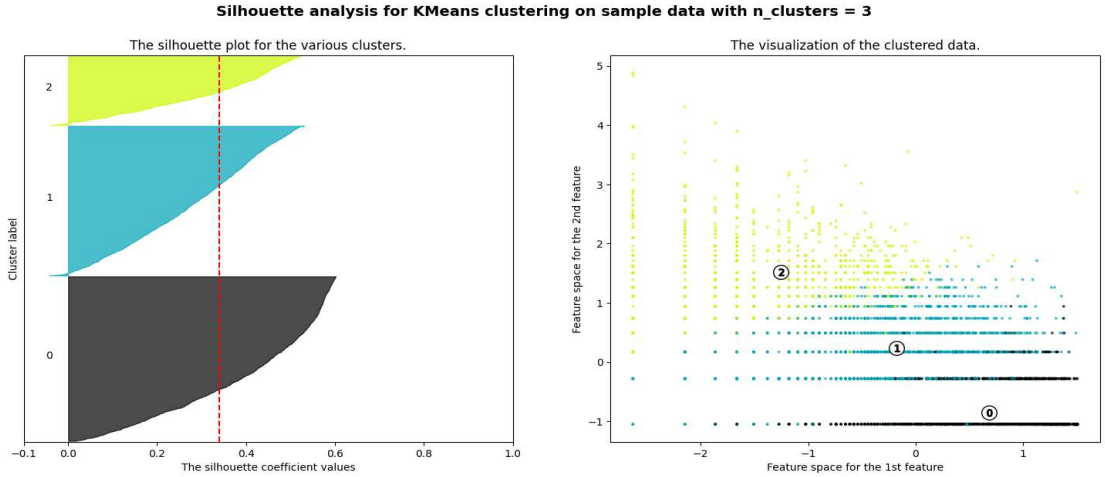
Data Transformation using Log Transformation:Reduce skewness by applying log transformation to Recency, Frequency, and Monetary values.



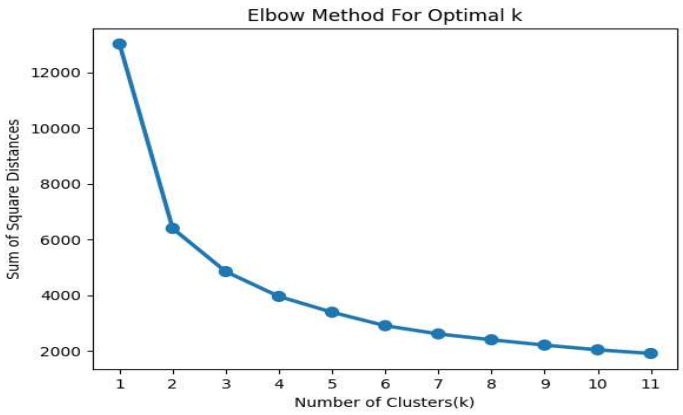
4.6 Model Training and Metrics

4.6.1 K-Means with silhouette score

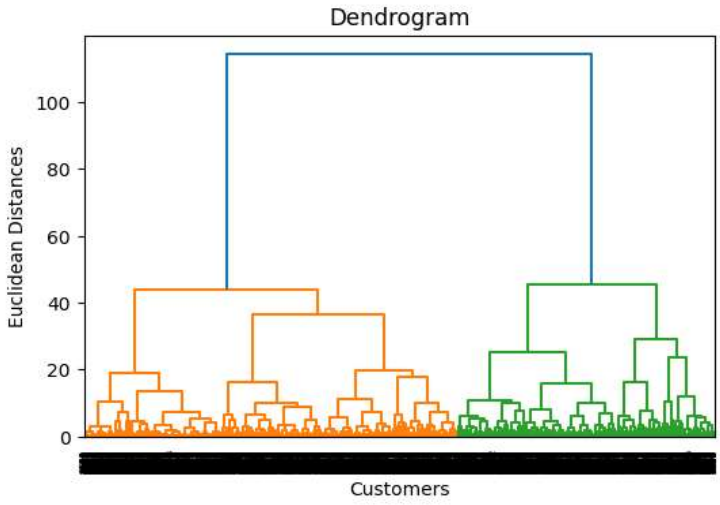




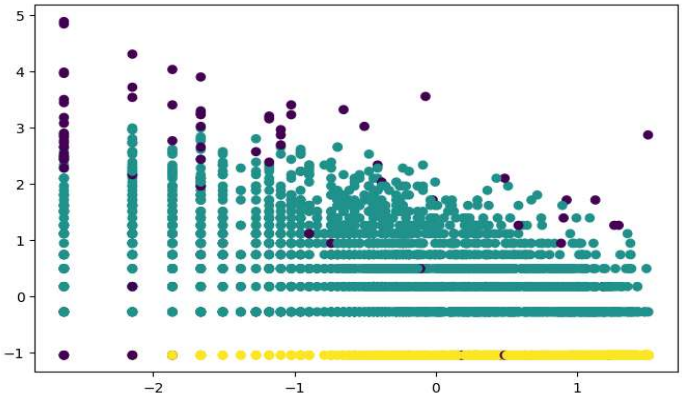
4.6.2 K-Means with Elbow Method



4.6.3 Hirerachial Clustering



4.6.4 DBSCAN



4.7 Results

Results on segmentation on R, F, M log transformed features :

SL No.	Model Name	No.of Clusters
1	K-Means with Silhouette score	2
2	K-Means with Elbow method	2
3	Hierarchical Clustering	2
4	DBSCAN	3

4.8 Customer Segements

Customer Segments from RFM analysis :

1. High-Value Segment:

Objective: Retain and maximize value.

Strategies: Offer personalized promotions or loyalty programs. Provide exclusive access to premium products/services. Gather feedback to enhance their experience.

2. Churn-Prone Segment:

Objective: Prevent churn and re-engage.

Strategies: Implement targeted marketing campaigns to re-engage. Provide incentives or discounts for repeat purchases. Understand reasons for potential churn through surveys.

3. Medium-Value or Average Segment :

Objective: Increase engagement and loyalty.

Strategies: Introduce loyalty programs or tiered rewards. Cross-sell or upsell relevant products/services. Monitor customer feedback and address concerns.

4.9 Conclusion

The "Online Retail Customer Segmentation" project has successfully categorized customers based on Recency, Frequency, and Monetary values, enabling targeted marketing strategies. By identifying high-value segments, the business can optimize resource allocation and enhance customer experiences. The risk and value segmentation framework provides a clear strategy for customer management. Additionally, log transformation addresses data skewness, ensuring robust analyses. This project equips the business with actionable insights for a customer-centric approach, fostering sustained growth and competitiveness in the online retail market.