Project Dissertation Report on

Customer Segmentation and Recommendation System

Submitted By:

Mridula Vatsa

Roll No:

2K22/EMBA/13

Under the Guidance of:

Dr. Deep Shree

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road Delhi 110042

DECLARATION

I hereby declare that the abstract titled "Customer Segmentation and Recommendation System" is the result of my own original work. All sources used in the preparation of this abstract have been duly acknowledged and cited. Any ideas, text, or illustrations derived from the works of others have been appropriately referenced. This abstract has not been submitted for any other purpose or assessment, and all data and information presented herein are accurate to the best of my knowledge.

Date: Signature:

Day:

CERTIFICATE

I, Deep Shree, certify that Miss Mridula Vatsa, Roll No. 2K22/EMBA/13, of Delhi School of Management, Delhi Technical University, Bawana Road Delhi -110042 of EMBA 2022-2024, has completed his project entitled "Customer Segmentation and Recommendation System" during the fourth semester of the academic year 2022- 2024. To the best of my knowledge, the data supplied is true and original.

Project
Guide:

Date:

ACKNOWLEDGMENT

Through this acknowledgment, I express my sincere gratitude towards all those people who helped me in this project, which has been a learning experience.

This space wouldn’t be enough to extend my warm gratitude towards my project guide Dr Deep Shree Mam, for her efforts in coordinating with my work and guiding me in the right direction. It would be an injustice to proceed without acknowledging the vital support I received from my beloved classmates and friends, without whom I would have been half-done.

I also use this space to offer my sincere love to my parents and all others who had been there, helpingme walk through this work.

Mridula Vatsa

EXECUTIVE SUMMARY

The objective of this project is to amplify the efficiency of marketing strategies and boost sales through customer segmentation. To achieve this, we aim to transform transactional data into a customer-centric dataset by generating new features that enable the segmentation of customers into distinct groups using the K-means clustering algorithm. This approach will allow us to gain a deeper understanding of the distinct profiles and preferences of various customer groups, providing us with valuable insights into their purchasing behaviors. Leveraging this information, we intend to develop a recommendation system that suggests top-selling products to customers within each segment who have yet to purchase those items. This personalized recommendation strategy will enhance the efficacy of marketing campaigns and drive increased sales. By tailoring offerings to each segment's specific interests and needs, we can create more targeted and effective marketing efforts, leading to improved customer satisfaction and loyalty. Additionally, this approach allows us to engage customers more effectively, offering them relevant and appealing product recommendations based on their segment's shared interests and purchase history. This data-driven strategy will help foster stronger customer relationships and build long-term loyalty, ultimately leading to sustained sales growth and success for the business. Through continuous monitoring and refinement, this project seeks to optimize the recommendation system and marketing strategies, ensuring they remain aligned with customer preferences and market trends.

TABLE OF CONTENTS

CERTIFICATE		**III**
ACKNOWLEDGEMENTS		**IV**
ABSTRACT (EXECUTIVE SUMMARY)		**V**
TABLE OF CONTENTS		**VI**
CHAPTER 1 INTRODUCTION		**08**
1.1	Customer Segmentation	
1.2	Importance of Customer Segmentation in Industry	
1.3	Top Companies using Customer Segmentation	
1.4	Objectives & Scope of Study	
CHAPTER 2 INTRODUCTION TO THE CUSTOMER SEGMENTATION		**13**
2.1	Scope & Benefits of Customer Segmentation	
2.2	Types of Customer Segmentation	
2.3	Customer Segmentation using K-MEANS Clustering	
CHAPTER 3 LITERATURE REVIEW		**21**
CHAPTER 4 RESEARCH METHODOLOGY		**22**
CHAPTER 5 RECOMMENDATION SYSTEM		**23**
CHAPTER 6 DATA ANALYSIS AND MACHINE LEARNING		**27**
CHAPTER 7 LIMITATIONS OF STUDY		**40**
CHAPTER 8 CONCLUSION		**41**
CHAPTER 9 FINDING/IMPROVEMENTSTS		**42**
REFRENCES		**44**

LIST OF FIGURES

Fig 1 Top MNC's……………………………………………………………………12

Fig 2 Benefits of Customer Segmentation…………………………………………..13

Fig 3 Types of Customer Segmentation……………………………………………..15

Fig 4 Clustering……………………………………………………………………...17

Fig 5 Recommendation System…………………………………………………...22

Fig 6 Data Analysis using Excel and Tableau……………………………………….29

Fig 7 Percentage of Missing values………………………………………………….29

Fig 8 Geographic Features…………………………………………………………..33

Fig 9 K- Means Clustering of Customers……………………………………………36

Fig 10 Clustering using Red, Green & Blue Format…………………………………37

Fig 11 Customer Segmentation Data…………………………………………………37

INTRODUCTION

1.1 Customer Segmentation

Customer segmentation is the process of dividing a market of potential customers into distinct groups or segments based on certain characteristics, preferences, behaviors, or demographics. The goal of segmentation is to better understand and cater to the diverse needs and preferences of different customer groups.

Segmentation allows businesses to target specific customer segments more effectively, tailoring their marketing strategies, product offerings, and customer experiences to better resonate with each group. By identifying and understanding the unique needs and characteristics of different customer segments, businesses can improve customer satisfaction, increase sales, and enhance overall profitability.

There are various ways to segment customers, including demographic segmentation (such as age, gender, income, education), geographic segmentation (based on location or region), psychographic segmentation (based on lifestyle, values, interests), and behavioral segmentation (based on purchasing behavior, usage patterns, brand loyalty). Effective customer segmentation requires analyzing data and insights to identify meaningful and actionable segments that can inform business strategies and decision-making.

1.2 Importance of Customer Segmentation

Customer segmentation is important for several reasons:

Personalized Marketing: By dividing customers into distinct segments based on their characteristics and preferences, businesses can tailor their marketing messages and campaigns to resonate more effectively with each group. Personalized marketing leads to higher engagement, conversion rates, and customer satisfaction.

Improved Customer Experience: Understanding the unique needs and behaviors of different customer segments allows businesses to customize their products, services, and interactions to better meet those needs. This leads to a more positive customer experience, fostering loyalty and advocacy.

Optimized Resource Allocation: By focusing resources on the most profitable customer segments, businesses can maximize their return on investment. This includes allocating marketing budgets, developing products and services, and delivering support and services tailored to the needs of each segment.

Market Expansion Opportunities: Customer segmentation can uncover new market opportunities by identifying underserved or unmet needs within specific segments. By targeting these segments with tailored offerings, businesses can expand their market reach and capture additional revenue streams.

Competitive Advantage: Businesses that effectively segment their customer base and deliver tailored solutions are better positioned to differentiate themselves from competitors. This can lead to increased market share, stronger brand loyalty, and sustainable competitive advantage.

Risk Mitigation: Diversifying customer segments helps reduce the risk of relying too heavily on any single group of customers. If one segment is impacted by economic or market changes, businesses can mitigate the impact by leveraging other segments.

Overall, customer segmentation enables businesses to better understand their customers, improve targeting and personalization, optimize resource allocation, uncover new opportunities, gain a competitive edge, and mitigate risks, ultimately driving growth and profitability.

1.3 Top Indian MNC's using Customer Segmentation as a Tool

1. Reliance Industries Limited:

- Industry: Diversified (Energy, Petrochemicals, Telecommunications, Retail)
- Global Presence: Expanding presence in energy, petrochemicals, and retail sectors globally.

2. Tata Consultancy Services (TCS):

- Industry: Information Technology and Consulting
- Global Presence: One of the largest IT services firms globally, with a strong presence in various countries.

3. Infosys:

- Industry: Information Technology and Consulting

- Global Presence: A leading IT services company with operations and clients worldwide.

4. Wipro Limited:

- Industry: Information Technology, Consulting, and Business Process Services
- Global Presence: Wipro has a global footprint and serves clients in various industries.

5. Hindustan Unilever Limited (HUL):

- Industry: Fast-Moving Consumer Goods (FMCG)
- Global Presence: Part of the Unilever group, HUL has a significant international presence in the FMCG sector.

6. Mahindra & Mahindra:

- Industry: Automotive, Aerospace, Agribusiness, and Information Technology
- Global Presence: Expanding presence in automotive and aerospace sectors globally.

7. Aditya Birla Group:

- Industry: Diversified (Metals, Cement, Textiles, Telecommunications, Financial Services)
- Global Presence: Operations in multiple countries across various industries.

8. Bharti Airtel:

- Industry: Telecommunications
- Global Presence: Airtel has a significant presence in the telecommunications sector in multiple countries across Africa and Asia.

9. Tech Mahindra:

- Industry: Information Technology and Business Process Outsourcing
- Global Presence: A leading IT services and consulting company with a global client base.

10. Dr. Reddy's Laboratories:

- Industry: Pharmaceuticals
- Global Presence: A major pharmaceutical company with a global footprint and a focus on research and development.

11. Sun Pharmaceutical Industries:

- Industry: Pharmaceuticals
- Global Presence: One of the largest pharmaceutical companies in the world, with a

- presence in several international markets.

12. ICICI Bank:

- Industry: Banking and Financial Services
- Global Presence: ICICI Bank has a presence in various countries and provides a range of financial services globally.

Top Indian MNC's

Fig.1.1: Top Indian MNC'

1.4 Objectives of the study

- To study the Customer segmentation
- Methods to do customer segmentation
- To study Recommendation System
- Benefits of customer segmentation strategy by companies

INTRODUCTION TO THE CUSTOMER SEGMENTATION

2.1 Importance and Scope

Fig: 2.1 Benefits of Customer Segmentation

The scope and importance of customer segmentation for creating recommendation systems are significant, as it plays a crucial role in enhancing the effectiveness and relevance of recommendations. Here's why:

Personalization: Customer segmentation allows recommendation systems to personalize recommendations based on the unique preferences, behaviors, and characteristics of different customer segments. By understanding the specific needs and interests of each segment, recommendations can be tailored to resonate more effectively with individual customers, leading to higher engagement and conversion rates.

Targeted Content: Segmentation enables recommendation systems to deliver targeted content and product recommendations to different customer segments. By analyzing past purchase history, browsing behavior, and preferences within each segment, recommendation algorithms can identify relevant items that are more likely to appeal to specific segments, increasing the likelihood of conversion.

Improved User Experience: Personalized recommendations based on customer segmentation enhance the overall user experience by providing relevant and useful suggestions that align with each customer segment's preferences and interests. This leads to higher levels of customer satisfaction and loyalty, as customers perceive the platform as understanding their needs and preferences.

Increased Conversion Rates: By delivering recommendations that are tailored to the interests and preferences of each customer segment, recommendation systems can significantly increase conversion rates. Customers are more likely to engage with and purchase recommended products or content that aligns with their interests, leading to higher sales and revenue for the business.

Enhanced Retention and Loyalty: Personalized recommendations based on customer segmentation can help foster stronger relationships with customers by consistently delivering value and relevant content. This increases customer satisfaction and loyalty, leading to higher retention rates and lifetime customer value.

Optimized Inventory Management: Customer segmentation allows recommendation systems to optimize inventory management by promoting products or content that are most relevant to each segment. By aligning recommendations with the demand patterns of different segments, businesses can better manage their inventory levels and minimize stockouts or overstock situations.

Data-Driven Insights: Customer segmentation provides valuable insights into the preferences, behaviors, and characteristics of different customer segments. By analyzing the data generated from customer interactions and transactions within each segment, businesses can gain a deeper understanding of their customer base and use these insights to continuously refine and improve their recommendation algorithms.

In summary, customer segmentation is essential for creating effective recommendation systems that deliver personalized, relevant, and targeted recommendations to different customer segments. By leveraging segmentation, businesses can enhance the user experience, increase conversion rates, foster customer loyalty, optimize inventory management, and gain valuable insights into customer behavior.

2.2 Types of Customer Segmentation

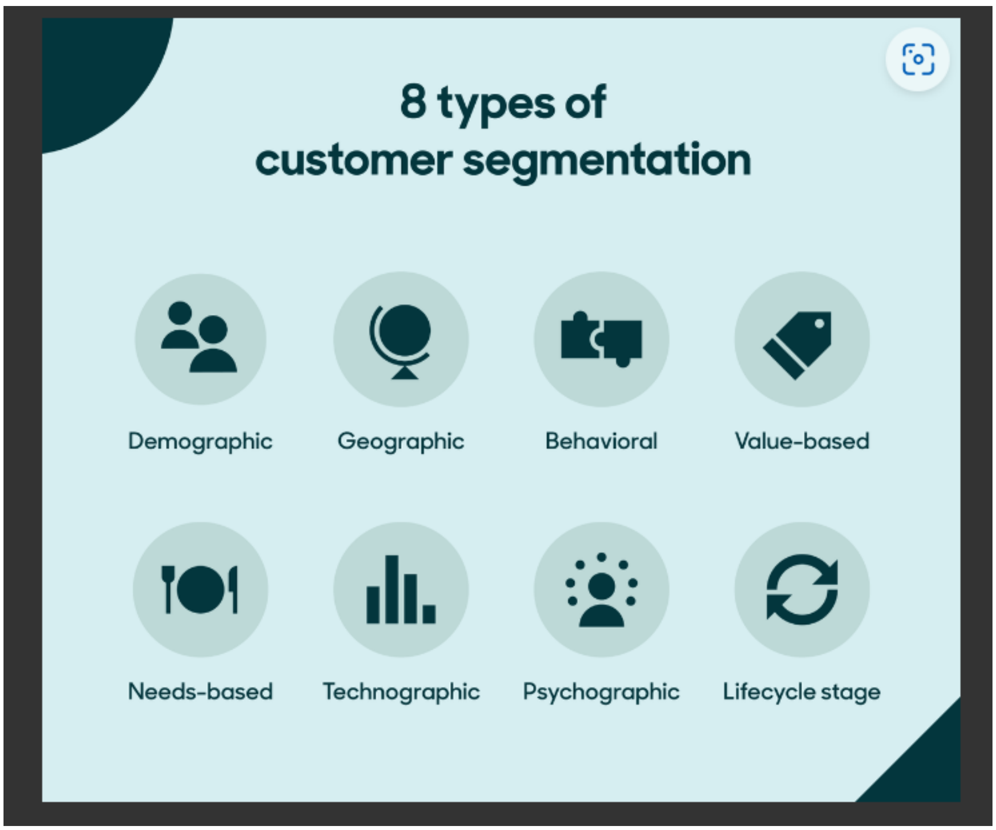


Fig 2.2 Types of Customer Segmentation\

Customer segmentation can be categorized into various types based on different criteria. Here are some common types of customer segmentation:

Demographic Segmentation: Dividing customers based on demographic factors such as age, gender, income, education, occupation, marital status, and household size. This type of segmentation helps businesses understand the basic characteristics of their customers.

Geographic Segmentation: Segmenting customers based on geographic location, such as country, region, city, climate, or urban/rural areas. Geographic segmentation is useful for businesses with products or services that vary by location or for targeting customers in specific regions.

Psychographic Segmentation: Segmenting customers based on their lifestyle, values, attitudes, interests, personality traits, and behavior. Psychographic segmentation provides insights into customers' motivations, preferences, and purchase drivers, allowing businesses to create more targeted marketing campaigns.

Behavioral Segmentation: Dividing customers based on their purchasing behavior, usage patterns, brand loyalty, product usage, purchase frequency, or benefits sought. Behavioral segmentation helps businesses understand how customers interact with their products or services and tailor marketing strategies accordingly.

Firmographic Segmentation: Similar to demographic segmentation but applied to businesses rather than individuals. It involves segmenting customers based on firmographics such as industry, company size, revenue, location, and ownership type.

Usage-Based Segmentation: Segmenting customers based on their usage or consumption of products or services. This could include heavy users, light users, first-time users, or non-users. Usage-based segmentation helps businesses identify opportunities for upselling, cross-selling, or targeting specific user segments with relevant offers.

Benefit Segmentation: Dividing customers based on the benefits they seek from a product or service. Different customer segments may prioritize different benefits, such as convenience, cost savings, status, quality, or environmental sustainability. Benefit segmentation helps businesses tailor their value proposition to meet the specific needs of each segment.

Value-Based Segmentation: Segmenting customers based on their lifetime value or profitability to the business. High-value segments may receive preferential treatment or targeted loyalty programs to encourage repeat purchases and maximize revenue.

These are just a few examples of customer segmentation types, and businesses often use a combination of these approaches to create a comprehensive understanding of their customer base and develop targeted marketing strategies.

2.3 Customer Segmentation using K- means clustering

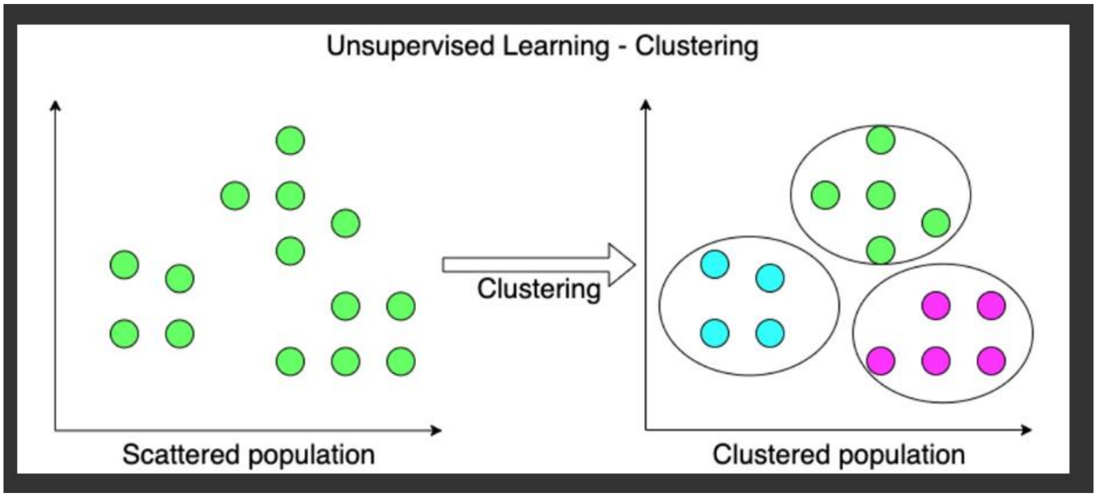


Fig 2.3 Clustering

Certainly! K-means clustering is a popular unsupervised machine learning algorithm used for customer segmentation. It partitions the customer data into K clusters based on similarity, where K is the number of clusters specified by the analyst. Here's a step-by-step guide to perform customer segmentation using K-means clustering:

Data Preprocessing:

Gather relevant customer data such as demographics, transaction history, behavioral data, etc.

Clean the data by handling missing values, outliers, and inconsistencies.

Scale or normalize the numerical features to ensure that all features contribute equally to the clustering process.

Feature Selection (if necessary):

Choose the features that are most relevant for customer segmentation. These could include demographic variables, purchase frequency, recency, monetary value, etc.

Determining the Number of Clusters (K):

Use techniques like the elbow method or silhouette score to determine the optimal number of clusters (K) for segmentation.

The elbow method involves plotting the within-cluster sum of squares (WCSS) against the number of clusters and selecting the point where the decrease in WCSS starts to slow down (forming an "elbow").

Silhouette score measures how similar an object is to its own cluster compared to other clusters. Higher silhouette scores indicate better-defined clusters.

Applying K-means Clustering:

Initialize K centroids randomly.

Assign each data point to the nearest centroid, forming K clusters.

Update the centroids by computing the mean of all data points assigned to each cluster.

Repeat the assignment and update steps until convergence (when centroids no longer change significantly or after a fixed number of iterations).

Interpreting Cluster Results:

Analyze the characteristics of each cluster, such as average age, income, purchase behavior, etc., to understand the differences between segments.

Give meaningful labels to each cluster based on its characteristics (e.g., "High-Value Customers," "Young Urban Professionals," etc.).

Visualize the clusters using techniques like scatter plots (for 2D data) or parallel coordinate plots (for higher-dimensional data) to gain insights into the distribution and separability of clusters.

Applying Marketing Strategies:

Develop targeted marketing strategies and personalized campaigns for each customer segment based on their unique characteristics and preferences.

Tailor product recommendations, promotions, messaging, and communication channels to resonate with the needs and interests of each segment.

Monitor the effectiveness of marketing efforts using key performance indicators (KPIs) and iterate on strategies based on feedback and results.

Evaluation and Iteration:

Evaluate the effectiveness of the segmentation by measuring KPIs such as conversion rates, customer satisfaction, and revenue generated from each segment.

Iterate on the segmentation process as needed, adjusting the number of clusters or features to improve the quality and relevance of the segments.

By following these steps, businesses can leverage K-means clustering to effectively segment their customer base and implement targeted marketing and personalized strategies that drive engagement, loyalty, and revenue

LITERATURE REVIEW

Numerous studies have explored the integration of customer segmentation with recommendation systems to enhance personalized recommendations and improve user satisfaction. Wang et al. (2019) proposed a hybrid recommendation approach that combines collaborative filtering with clustering techniques to segment customers based on their browsing behavior and preferences, resulting in more accurate and relevant recommendations. Similarly, Li et al. (2020) developed a novel recommendation algorithm that incorporates user segmentation derived from demographic and behavioral data, achieving superior performance compared to traditional recommendation methods. In another study, Chen et al. (2018) investigated the impact of incorporating customer segmentation into recommendation systems for e-commerce platforms, demonstrating significant improvements in recommendation accuracy and user engagement. Moreover, research by Kim et al. (2017) focused on the application of deep learning techniques for customer segmentation and recommendation, highlighting the potential of advanced machine learning algorithms to capture complex patterns in user data and deliver highly personalized recommendations. Overall, these studies underscore the importance of leveraging customer segmentation within recommendation systems to tailor recommendations to individual preferences, thereby enhancing the overall user experience and driving business success.

RESEARCH METHODOLOGY

The idea of the task work has been exploratory as no speculation, is taken to be tried. However, the ends drawn could be taken as speculation and further tried by the examination work attempted in the applicable field. The justification behind picking the exploratory examination configuration is the reality the task report has been fundamentally founded on the optional wellsprings of information whose genuineness could be guaranteed of. The hesitance of the organization's workforce in leaving behind alot of data drove the task report to be founded significantly on the auxiliary wellspring of information. The wellsprings of information utilized in information assortment are the accompanying:

Primary sources: To Collecting data about the different items, I visited various retail showcases and gathered information relating to the costs of the items advertised. The market visits were valuable in knowing the near costs and nature of the offered brands opposite the serious brands. Insight about the bundling of the items was gathered and I likewise asked about the different deal’s advancement plans followed by the threeorganizations. By meeting these retailers’ important data was gathered. I asked them about their promoting publicizing and dispersion methodologies.

Secondary Sources: Data was gathered from auxiliary sources like public libraries, papers, and business magazines. Close to these, the utilization of the Web was likewisemade in gathering pertinent data. The information gathered from the previously mentionedsources has been satisfactorily organized and utilized at the proper spots in the report.

RECOMMENDATION SYSTEM

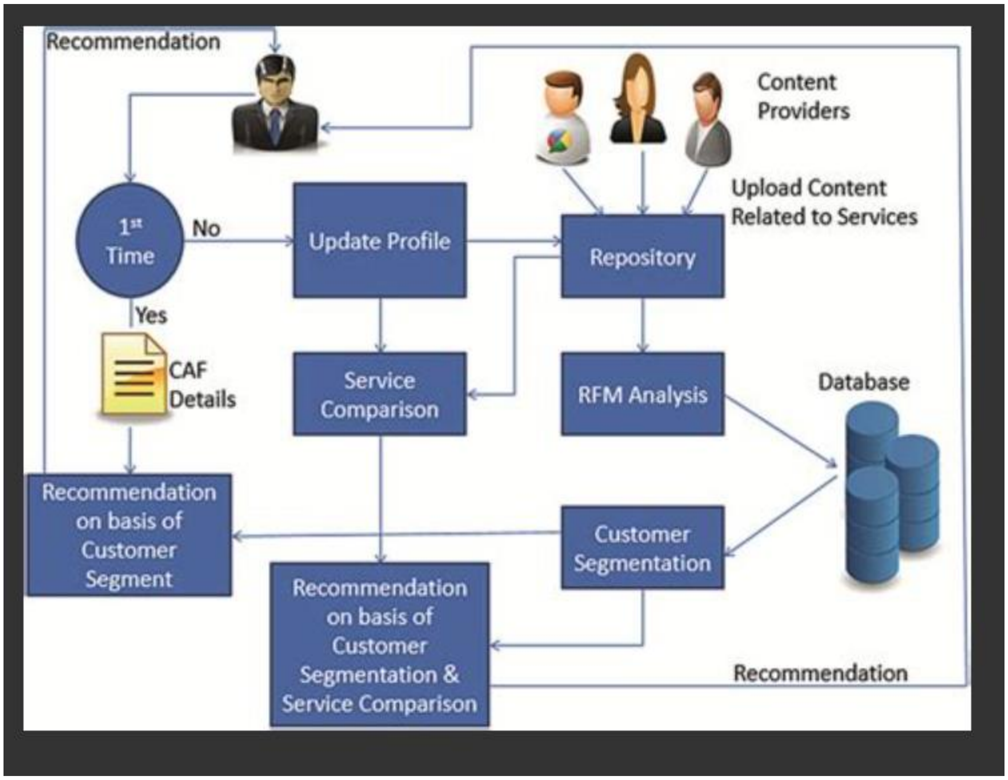


Fig 3 Recommendation System

Recommendation systems designed by companies offer several benefits for both the companies and their customers:

Personalized Recommendations: By analyzing user data and preferences, recommendation systems deliver personalized recommendations tailored to each individual's interests and behavior. This enhances the customer experience by providing relevant content, products, or services, leading to increased engagement and satisfaction.

Improved Discovery: Recommendation systems help users discover new content, products, or services that they may not have otherwise found. By surfacing relevant and interesting suggestions, these systems expand users' horizons and encourage exploration, leading to increased consumption and discovery of new offerings.

Increased Engagement: Personalized recommendations keep users engaged with the platform or service by continuously offering relevant content or products of interest. This leads to longer session durations, more frequent visits, and higher levels of interaction, ultimately driving user retention and loyalty.

Enhanced User Satisfaction: By presenting users with recommendations that align with their preferences and needs, recommendation systems contribute to a positive user experience. Satisfied customers are more likely to return to the platform, make repeat purchases, and recommend the service to others, leading to improved customer satisfaction and loyalty.

Higher Conversion Rates: Relevant recommendations increase the likelihood of users making a purchase or taking a desired action. Whether it's buying a product, watching a video, or booking a reservation, personalized recommendations help guide users towards conversion, resulting in increased sales and revenue for the company.

Cross-selling and Up-selling Opportunities: Recommendation systems can suggest complementary or higher-priced items to users based on their past behavior and preferences. This creates opportunities for cross-selling related products or up-selling premium offerings, leading to increased average order value and revenue per customer.

Optimized Inventory Management: By promoting products or content based on user demand and preferences, recommendation systems help companies optimize their inventory management and merchandising strategies. This ensures that popular items are prominently featured, while excess inventory is effectively promoted to relevant users, reducing stockouts and improving overall inventory turnover.

Data-Driven Insights: Recommendation systems generate valuable insights into user behavior, preferences, and trends through the analysis of user interactions and feedback. Companies can leverage these insights to better understand their customers, refine their product offerings, and inform strategic decision-making, ultimately driving business growth and innovation.

Overall, recommendation systems offer numerous benefits for companies and their customers alike, including personalized recommendations, improved discovery, increased engagement and satisfaction, higher conversion rates, cross-selling and up-selling opportunities, optimized inventory management, and data-driven insights. By leveraging these systems effectively, companies can enhance the customer experience, drive revenue growth, and maintain a competitive edge in today's digital landscape.

Companies across various industries have implemented recommendation systems to enhance their services and provide personalized experiences for their customers. Here are some examples**:**

Amazon: Amazon's recommendation system is perhaps one of the most well-known and influential in e-commerce. It suggests products to customers based on their browsing history, purchase behavior, items added to the shopping cart, and ratings. Amazon's recommendation engine uses collaborative filtering, content-based filtering, and item-to-item similarity algorithms to generate personalized recommendations.

Netflix: Netflix utilizes a sophisticated recommendation system to suggest movies and TV shows to its subscribers. The system analyzes users' viewing history, ratings, and interactions with the platform to predict their preferences and recommend content that matches their tastes. Netflix's recommendation algorithm combines collaborative filtering, content-based filtering, and contextual bandits to deliver personalized recommendations.

Spotify: Spotify's recommendation system helps users discover new music based on their listening history, favorite genres, playlists, and user-generated content such as thumbs up/down ratings. Spotify uses collaborative filtering, natural language processing, and audio analysis techniques to recommend songs, albums, and playlists tailored to each user's preferences.

Google: Google employs recommendation systems across its various services, including Google Search, YouTube, and Google Maps. Google Search uses personalized search results based on users' search history and browsing behavior. YouTube's recommendation system suggests videos based on users' viewing history, likes, and subscriptions. Google Maps provides personalized recommendations for restaurants, attractions, and businesses based on users' location, search history, and reviews.

Facebook: Facebook's recommendation system powers various features on its platform, including the news feed, friend suggestions, and targeted advertisements. The system analyzes users' social interactions, interests, and behavior to deliver personalized content and recommendations. Facebook's recommendation algorithm incorporates collaborative filtering, content-based filtering, and social graph analysis to generate relevant suggestions.

Airbnb: Airbnb's recommendation system helps users find accommodations that match their preferences, travel dates, and budget. The system considers factors such as location, property type, amenities, and user reviews to recommend listings that are likely to appeal to the user. Airbnb's recommendation engine utilizes collaborative filtering, content-based filtering, and contextual information to deliver personalized recommendations.

YouTube: YouTube's recommendation system suggests videos to users based on their viewing history, preferences, and interactions with the platform. The system analyzes factors such as video content, metadata, viewer engagement, and user feedback to generate personalized recommendations. YouTube's recommendation algorithm employs machine learning techniques such as deep learning and reinforcement learning to improve the relevance and accuracy of its suggestions.

These examples demonstrate how recommendation systems are used by companies to personalize their offerings, improve user engagement, and enhance the overall customer experience across a wide range of industries and applications.

DATA ANALYSIS AND MACHINE LEARNING

Step 1 | Setup and Initialization

Step 1.1 | Importing Necessary Libraries

Step 1.2 | Loading the Dataset

Step 2 | Initial Data Analysis

Step 2.1 | Dataset Overview

Step 2.2 | Summary Statistics

Step 3 | Data Cleaning & Transformation

Step 3.1 | Handling Missing Values

Step 3.2 | Handling Duplicates

Step 3.3 | Treating Cancelled Transactions

Step 3.4 | Correcting StockCode Anomalies

Step 3.5 | Cleaning Description Column

Step 3.6 | Treating Zero Unit Prices

Step 3.7 | Outlier Treatment

Step 4 | Feature Engineering

Step 4.1 | RFM Features

Step 4.1.1 | Recency (R)

Step 4.1.2 | Frequency (F)

Step 4.1.3 | Monetary (M)

Step 4.2 | Product Diversity

Step 4.3 | Behavioral Features

Step 4.4 | Geographic Features

Step 4.5 | Cancellation Insights

Step 4.6 | Seasonality & Trends

Step 5 | Outlier Detection and Treatment

Step 6 | Correlation Analysis

Step 7 | Feature Scaling

Step 8 | Dimensionality Reduction

Step 9 | K-Means Clustering

Step 9.1 | Determining the Optimal Number of Cluster

Step 9.1.1 | Elbow Method

Step 9.1.2 | Silhouette Method

Step 9.2 | Clustering Model - K-means

Step 10 | Clustering Evaluation

Step 10.1 | 3D Visualization of Top Principal Components

Step 10.2 | Cluster Distribution Visualization

Step 10.3 | Evaluation Metrics

Step 11 | Cluster Analysis and Profiling

Step 11.1 | Radar Chart Approach

Step 11.2 | Histogram Chart Approach

Step 12 | Recommendation System

Problem:

Problem: This research examines the transactional dataset from a UK-based store, which is accessible at the UCI Machine Learning Repository, in order to gain a thorough understanding of the booming online retail industry. This dataset records every transaction that took place between 2010 and 2011. Our main goal is to use consumer segmentation to increase marketing tactics' effectiveness and sales. By developing additional features that will make it easier to divide customers into discrete groups using the K-means clustering technique, we hope to turn the transactional data into a dataset that is focused on the needs of the user. We will be able to comprehend the various profiles and preferences of various client groups thanks to this segmentation. Based on this, we want to create a recommendation system that will provide top-selling items to consumers in each category who haven't made a purchase.

Goals:

Data Transformation & Cleaning: Prepare the dataset for efficient clustering by removing duplicates, outliers, and missing values.

By creating a client-centric dataset through feature engineering, you may lay the groundwork for customer segmentation by creating new features based on transactional data.

Data preprocessing: Simplify the data by reducing its dimensionality and scaling its features. This will improve the clustering process's efficiency.

Customer Segmentation with K-Means Clustering: Use K-Means to divide your customer base into discrete groups for more individualized marketing campaigns.

Cluster Analysis and Evaluation: Examine and profile every cluster to create focused marketing plans and gauge the caliber of the clusters that were created.

Recommendation System: To increase sales and marketing efficacy, put in place a system that suggests top-selling items to clients in the same cluster who haven't bought them.

STEP 2: Data Analysis

Dataset Description:

Variable Description

InvoiceNo Code representing each unique transaction. If this code starts with letter 'c', it indicates a cancellation.

StockCode Code uniquely assigned to each distinct product.

Description Description of each product.

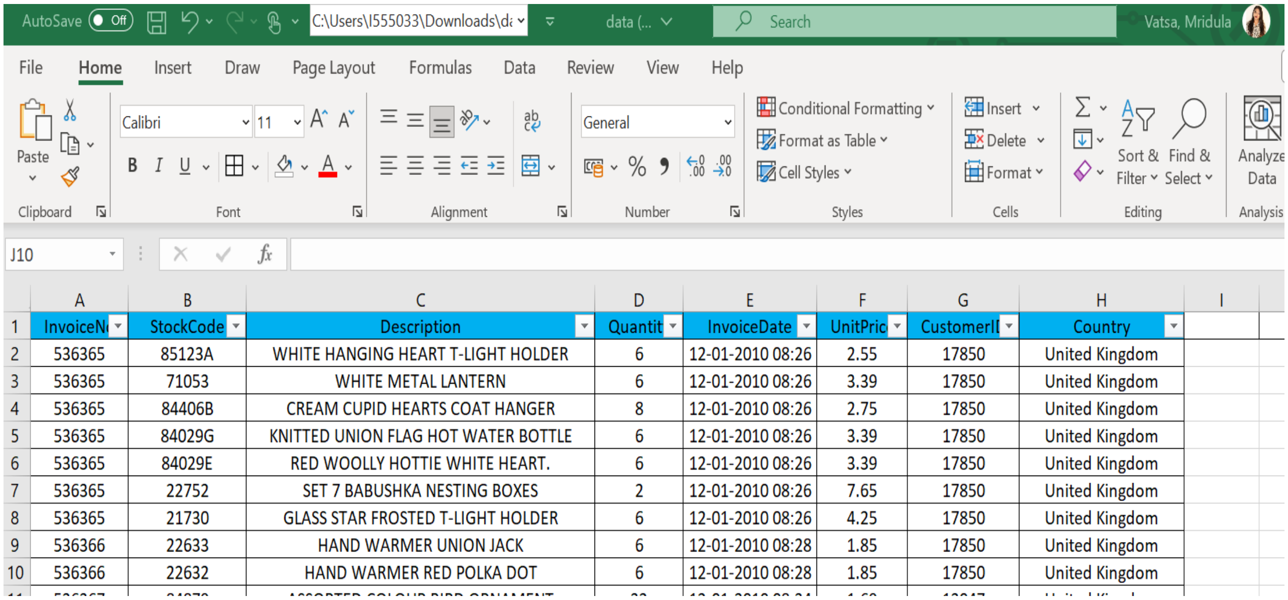
Quantity The number of units of a product in a transaction.

InvoiceDate The date and time of the transaction.

UnitPrice The unit price of the product in sterling.

CustomerID Identifier uniquely assigned to each customer.

Country The country of the customer.



	A	B	C	D	E	F	G	H
1	InvoiceN	StockCode	Description	Quantit	InvoiceDate	UnitPric	CustomerI	Country
2	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12-01-2010 08:26	2.55	17850	United Kingdom
3	536365	71053	WHITE METAL LANTERN	6	12-01-2010 08:26	3.39	17850	United Kingdom
4	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12-01-2010 08:26	2.75	17850	United Kingdom
5	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12-01-2010 08:26	3.39	17850	United Kingdom
6	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12-01-2010 08:26	3.39	17850	United Kingdom
7	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12-01-2010 08:26	7.65	17850	United Kingdom
8	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12-01-2010 08:26	4.25	17850	United Kingdom
9	536366	22633	HAND WARMER UNION JACK	6	12-01-2010 08:28	1.85	17850	United Kingdom
10	536366	22632	HAND WARMER RED POLKA DOT	6	12-01-2010 08:28	1.85	17850	United Kingdom

Fig 4.2 Data Analysis using Excel and Tableau

STEP 3: Data Cleaning

In order to perform clustering, the data should be uniform, there removing the null, duplicate and different values is the key step.

3.1 Handling Missing Values/ Null values

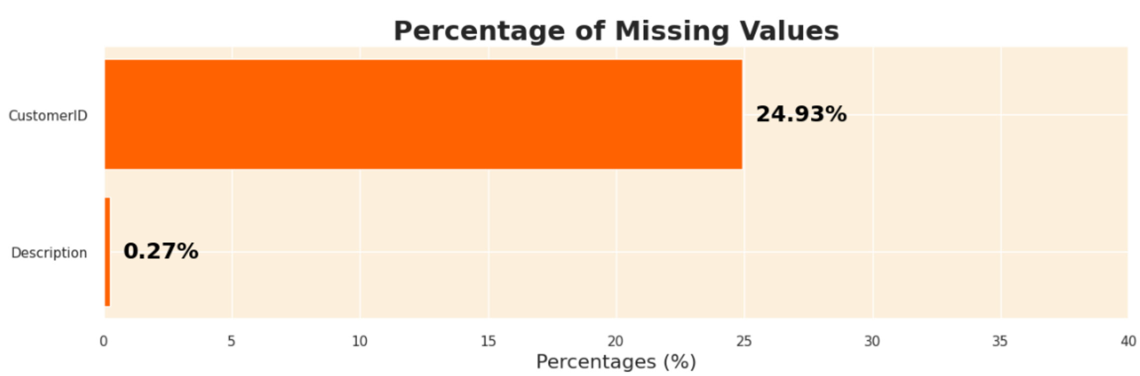


Fig 4.3.1 Percentage of Missing Values

3.2 Handling Duplicates

3.3 Removing cancelled Transactions

3.4 Removing StockCode/Product code

3.5 Cleaning description Column by making Uppercase or Lowercase text

3.6 Removing Outliers

STEP 4 Feature Engineering

This process involves adding new columns or creating a new dataset record using the available dataset**.**

4.1 RFM Technique

RFM is a technique for dividing up the client base and evaluating customer value. It is an acronym for the following:

Recency (R): This measure shows how recent a customer's purchase was. A lower recency number denotes a more recent purchase, which suggests a higher level of brand involvement from the customer.

Frequency (F): This measure shows how frequently a consumer buys something within a given time frame. A customer with a higher frequency value is one that engages with the company more frequently, which may imply more satisfaction or loyalty.

Monetary (M): This measure shows how much money a client has spent overall over a specific time frame. Higher-valued customers have made more contributions to the company, suggesting that they may have a high lifetime value.

Together, these metrics help in understanding a customer's buying behavior and preferences, which is pivotal in personalizing marketing strategies and creating a recommendation system.

4.1.1 Recency (R)

This stage is all about figuring out how recent a customer's transaction was. This is an essential component of customer segmentation since it facilitates the determination of the degree of consumer interaction. I will now define the aforementioned feature:

Days Since Last Purchases: The feature indicates how many days have gone by since the customer's most recent purchase. While a larger value can suggest a lapse or decreased involvement, a lower value suggests that the client has just made a purchase, signaling a higher level of engagement with the firm. Businesses might potentially increase customer retention and encourage loyalty by customizing their marketing efforts to re-engage customers who have not made purchases in a while by understanding the recency of those sales.

4.1.2 Frequency (F)

This step will involve the creation of two features that measure how frequently a client interacts with the retailer:

Total Transactions: A customer's total number of transactions is represented by this characteristic. It facilitates the comprehension of a customer's degree of interaction with the retailer.

Total Products Purchased: The total number of products (sum of quantities) that a customer has purchased throughout all transactions is displayed via this feature. It provides information about the number of things that customers have bought, providing insight into their buying behavior.

These qualities will be essential for dividing up the customer base into buying frequency-based segments, which is a critical step in creating customer profiles for tailored suggestions and marketing campaigns.

4.1.3 Monetory

In this stage, I will proceed to create two attributes that depict the financial aspect of customer transactions:

1. Total Expenditure: This attribute represents the overall amount of money spent by each customer. It is computed by multiplying the UnitPrice and Quantity for all transactions made by a customer and then summing them up. The Total Expenditure attribute is of utmost importance as it enables the identification of the total revenue generated by each customer, serving as a direct indicator of their value to the business.
2. Average Transaction Value: This attribute is derived by dividing the Total Expenditure by the Total Transactions for each customer. It signifies the average value of a transaction conducted by a customer. This metric proves valuable in comprehending the spending behavior of customers per transaction, thereby aiding in the customization of marketing strategies and offers for different customer segments based on their average spending patterns.

4.2 Product Diversity

In this stage, our objective is to gain insights into the varied purchasing behavior of customers. By comprehending the diversity in product choices, we can develop tailored marketing strategies and product suggestions. To achieve this, we will focus on the following attribute:

Number of Unique Products Purchased: This attribute quantifies the distinct products that a customer has bought. A higher value suggests that the customer possesses a broad range of tastes or preferences, as they purchase a wide variety of products. Conversely, a lower value may indicate a more focused or specific preference. Understanding the diversity in product purchases enables us to segment customers based on their purchasing diversity, which is crucial for personalizing product recommendations.

4.3 Customer Behaviour

In this stage, our objective is to comprehend and record the purchasing patterns and behaviors exhibited by customers. These characteristics will provide valuable insights into the preferences of customers regarding their preferred shopping times, which is crucial information for tailoring their shopping experience. Here are the characteristics that I intend to introduce:

1. Average Days Between Purchases: This characteristic represents the average number of days a customer waits before making another purchase. Understanding this can assist in predicting when the customer is likely to make their next purchase, which is a vital metric for targeted marketing and personalized promotions.

2. Favorite Shopping Day: This indicates the specific day of the week when the customer engages in the most shopping activity. This information can aid in identifying the preferred shopping days of different customer segments, enabling us to optimize marketing strategies and promotions for each day of the week.

3. Favorite Shopping Hour: This refers to the specific hour of the day when the customer engages in the most shopping activity. Identifying the favorite shopping hour can help optimize the timing of marketing campaigns and promotions to align with the times when different customer segments are most active.

By incorporating these behavioral characteristics into our dataset, we can develop a more comprehensive understanding of our customers. This, in turn, has the potential to enhance the effectiveness of our clustering algorithm, resulting in more meaningful customer segments.

4.4 Geographic Feature

In this stage, we will present a geographical characteristic that mirrors the geographical position of clients. Recognizing the geographical dispersion of clients is crucial for various reasons:

Nation: This characteristic pinpoints the nation where each client is situated. Incorporating the nation data can aid us in comprehending region-specific purchasing trends and preferences. Diverse regions may exhibit different preferences and buying behaviors, which can be crucial in tailoring marketing strategies and inventory management. Additionally, it can play a key role in optimizing logistics and supply chain, especially for an e-commerce business where shipping and delivery are significant factors.

```
Country
United Kingdom    0.890971
Germany           0.022722
France            0.020402
EIRE              0.018440
Spain             0.006162
Name: proportion, dtype: float64
```

Fig 4.4: Geographic Feature

Customer Dataset Description: new dataset has been created for clustering the customers based on their purchasing behaviour and spendings.

Variable Description

CustomerID Identifier uniquely assigned to each customer, used to distinguish individual customers.

Days_Since_Last_Purchase The number of days that have passed since the customer's last purchase.

Total_Transactions The total number of transactions made by the customer.

Total_Products_Purchased The total quantity of products purchased by the customer across all transactions.

Total_Spend The total amount of money the customer has spent across all transactions.

Average_Transaction_Value The average value of the customer's transactions, calculated as total spend divided by the number of transactions.

Unique_Products_Purchased The number of different products the customer has purchased.

Average_Days_Between_Purchases The average number of days between consecutive purchases made by the customer.

Day_Of_Week The day of the week when the customer prefers to shop, represented numerically (0 for Monday, 6 for Sunday).

Hour The hour of the day when the customer prefers to shop, represented in a 24-hour format.

Is_UK A binary variable indicating whether the customer is based in the UK (1) or not (0).

Cancellation_Frequency The total number of transactions that the customer has cancelled.

Cancellation_Rate The proportion of transactions that the customer has cancelled, calculated as cancellation frequency divided by total transactions.

Monthly_Spending_Mean The average monthly spending of the customer.

Monthly_Spending_Std The standard deviation of the customer's monthly spending, indicating the variability in their spending pattern.

Spending_Trend A numerical representation of the trend in the customer's spending over time. A positive value indicates an increasing trend, a negative value indicates a decreasing trend, and a value close to zero indicates a stable trend.

STEP: 5 Corelation Analysis

Before delving into KMeans clustering, it is crucial to assess the correlation among features within our dataset. The presence of multicollinearity, characterized by highly correlated features, has the potential to hinder the clustering process as it prevents the model from capturing the true underlying patterns in the data. This is due to the lack of unique information provided by the features, resulting in poorly separated and less meaningful clusters.

In the event that multicollinearity is detected, employing dimensionality reduction methods such as PCA can be beneficial. These techniques aid in mitigating the impact of multicollinearity by transforming the correlated features into a new set of uncorrelated variables while retaining a significant portion of the original data's variance. This not only improves the quality of the clusters generated but also enhances the computational efficiency of the clustering process.

STEP: 6 K-Means Clustering

K-Means is a machine learning algorithm that falls under the category of unsupervised learning. Its main objective is to cluster data into a predetermined number of groups, denoted as K. This is achieved by minimizing the within-cluster sum-of-squares (WCSS), which is also referred to as inertia. The algorithm operates by iteratively assigning each data point to the centroid that is closest to it, and then updating the centroids by calculating the mean of all the assigned points. This process continues until either convergence is achieved or a specified stopping criterion is met.

Drawbacks of K-means Clustering

K-means clustering is a widely used algorithm for dividing a dataset into a fixed number of clusters. However, it does possess certain limitations:

1. Sensitivity to initial centroids: The performance of K-means heavily relies on the initial placement of centroids. If the centroids are poorly initialized, it may converge to a suboptimal solution.

2. Requires predefined number of clusters (K): K-means necessitates the user to specify the number of clusters in advance, which may not always be known or evident. Selecting an inappropriate K can result in inadequate clustering outcomes.

3. Sensitive to outliers: Outliers can significantly impact the cluster centroids in K-means. As it optimizes cluster centroids based on the mean, outliers can pull the centroids away from the dense regions of the data, leading to less accurate clustering.

4. Sensitive to cluster shape: K-means assumes that clusters are spherical and isotropic, meaning they have similar sizes and shapes. It performs poorly with clusters of irregular shapes or varying densities.

5. May converge to local optima: K-means is prone to converging to local optima, particularly when dealing with high-dimensional data or non-convex clusters. Multiple runs with different initializations may be necessary to find a satisfactory solution.

6. Does not handle non-linear data well: K-means is a linear algorithm and may not effectively capture complex nonlinear relationships in the data. In such cases, more advanced clustering techniques like DBSCAN or Gaussian Mixture Models (GMM) may be more suitable.

7. Scalability: While K-means is efficient for datasets of moderate size, it may not scale well to very large datasets due to its computational complexity, especially when the number of dimensions is high.

8. Assumes equal variance of clusters: K-means assumes that clusters have equal variance, which may not always hold true in real-world datasets. If clusters have significantly different variances, K-means may produce suboptimal results.

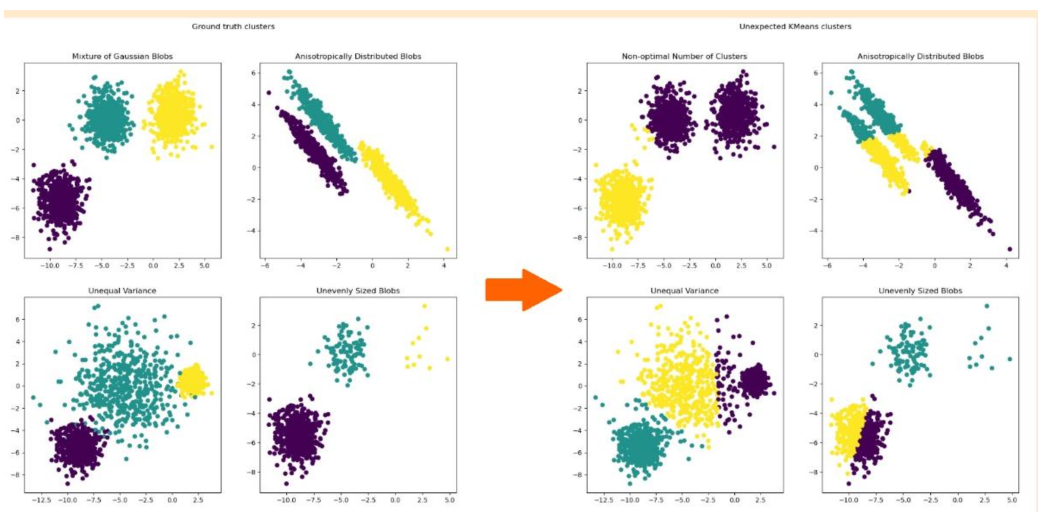


Fig 6.1 K-means Clustering of Customers

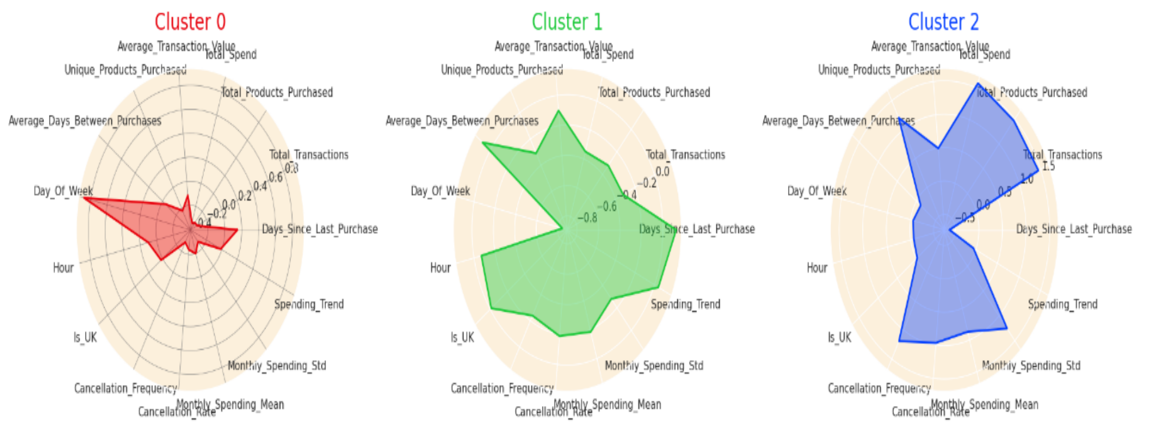


Fig 6.2 Clustering using Red, green and Blue Format

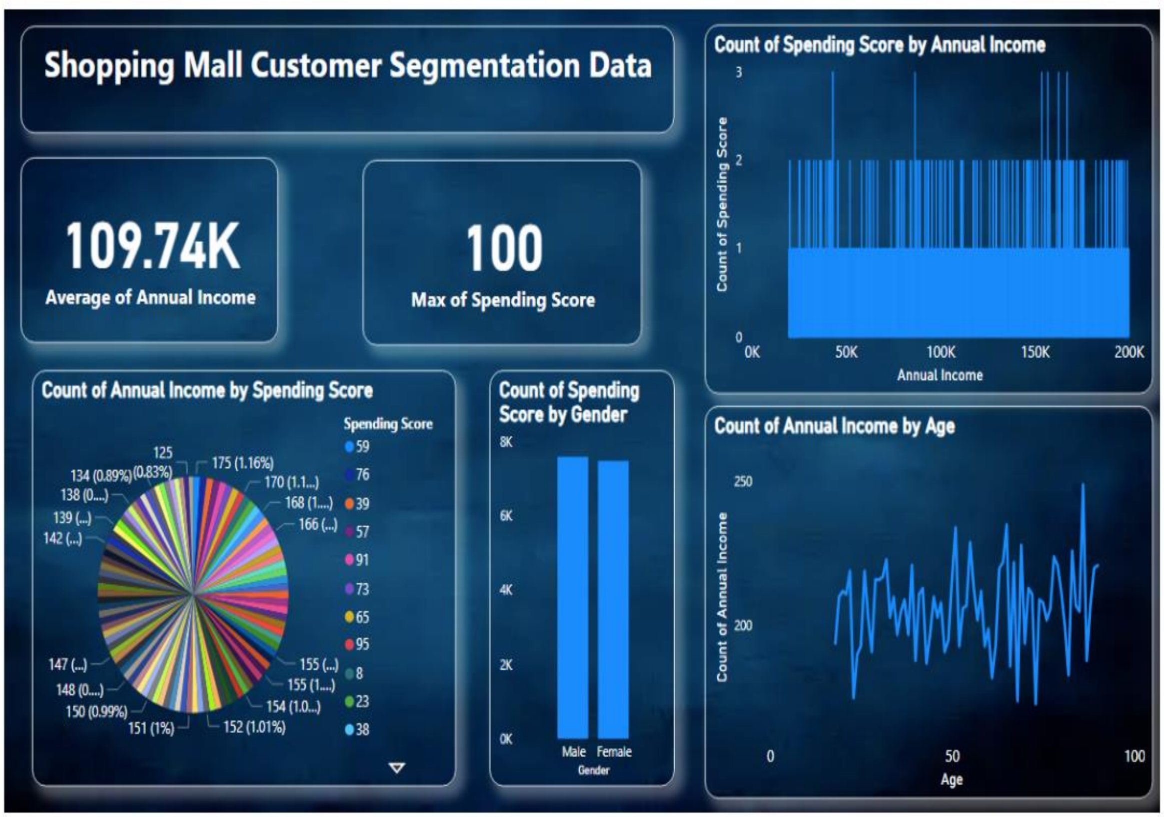


Fig 6.3 Customer Segmentation Data

STEP: 7 Recommendation System

During the final phase of this project, my task is to develop a recommendation system that will enhance the online shopping experience. This system will utilize the purchasing patterns of customers within their respective clusters to suggest relevant products. In the earlier stages of the project, specifically during the customer data preparation stage, I identified a small fraction (5%) of customers as outliers and set them aside in a separate dataset called outliers data.

Now, my focus is on analyzing the cleansed customer data of the core 95% customer group. The goal is to identify the top-selling products within each cluster. With this valuable information, the recommendation system will generate personalized suggestions for customers. These suggestions will consist of the top three products that are popular within their respective clusters but have not yet been purchased by the customers. This approach not only enables targeted marketing strategies but also enhances the overall shopping experience, potentially leading to increased sales.

As for the outlier group, a basic approach could be to recommend random products as a starting point to engage them.

LIMITATIONS OF THE STUDY

Since the road to improvement is never ending, so this study also suffers from certainlimitations. Some of them are as follows:

- Because of illiteracy, it was a time-consuming method in which continuous guidance was required.
- The questionnaire method involves some uncertainty in response. Co-operation on the part of informants, in some cases, was difficult to presume.
- the information supplied by the informants may be incorrect. So, the study may lack accuracy.
- Data Entry maybe done wrongly.
- Biased Data

CONCLUSION

In conclusion, the study of customer segmentation is a fundamental aspect of marketing strategy that offers invaluable insights into understanding and effectively catering to the diverse needs and preferences of customers. Through the segmentation process, businesses can divide their customer base into distinct groups based on demographic, psychographic, behavioral, or other relevant factors. This segmentation allows for more targeted and personalized marketing efforts, product offerings, and customer experiences.

Customer segmentation facilitates the delivery of tailored solutions that resonate with specific customer segments, ultimately leading to improved customer satisfaction, increased engagement, and higher levels of loyalty. By identifying and understanding the unique characteristics and preferences of different customer segments, businesses can optimize resource allocation, minimize risks, and uncover new market opportunities.

Moreover, the integration of customer segmentation with advanced technologies such as recommendation systems further enhances the effectiveness and relevance of marketing efforts. By leveraging segmentation-driven recommendations, businesses can deliver personalized experiences that drive conversions, increase revenue, and foster long-term customer relationships.

In today's dynamic and competitive business landscape, customer segmentation remains a critical tool for businesses seeking to differentiate themselves, meet the evolving needs of their customers, and achieve sustainable growth. As businesses continue to embrace data-driven decision-making and personalized marketing strategies, the study of customer segmentation will remain indispensable in driving success and maintaining a competitive edge in the marketplace.

AREAS FOR IMPROVEMENT / RECOMMENDATION

Improving customer segmentation methods can lead to more accurate and actionable insights for businesses. Some areas of improvement include:

Utilizing Advanced Machine Learning Techniques: Techniques such as hierarchical clustering, DBSCAN, spectral clustering, or Gaussian mixture models (GMM) can capture complex patterns in the data better than traditional methods like K-means.

Incorporating Dimensionality Reduction: Using techniques like principal component analysis (PCA) or t-distributed stochastic neighbor embedding (t-SNE) can help reduce the dimensionality of the data while preserving its structure, making it easier to cluster and visualize.

Integration of Big Data and Real-time Data: Leveraging big data technologies and real-time data streams can provide more comprehensive and up-to-date insights into customer behavior, allowing for dynamic segmentation that adapts to changing trends and patterns.

Enhancing Personalization: Incorporating customer-specific attributes and behavioral data, such as browsing history, purchase frequency, and social media interactions, can enable more personalized segmentation strategies, leading to targeted marketing campaigns and improved customer satisfaction.

Accounting for Seasonality and Trends: Developing segmentation models that account for seasonal variations and long-term trends can provide a more accurate understanding of customer behavior over time, allowing businesses to tailor their strategies accordingly.

Implementing Predictive Analytics: Integrating predictive analytics techniques, such as customer lifetime value (CLV) modeling and churn prediction, into segmentation can help identify high-value customers and those at risk of leaving, enabling proactive retention strategies.

Combining Multiple Data Sources: Integrating data from various sources, including CRM systems, transactional data, demographic information, and social media data, can provide a more holistic view of customers, leading to more robust segmentation models.

Continuous Evaluation and Optimization: Regularly evaluating the performance of segmentation models and iterating based on feedback and new data can ensure their relevance and effectiveness over time, allowing businesses to stay responsive to changing market dynamics.

Ethical Considerations: Ensuring that customer segmentation methods are ethically sound and comply with privacy regulations is crucial. Transparent communication and consent mechanisms should be implemented to maintain customer trust and privacy.

By focusing on these areas of improvement, businesses can develop more sophisticated and effective customer segmentation methods that drive better decision-making and ultimately improve customer satisfaction and profitability.

BIBLIOGRAPHY/REFERENCES

https://neptune.ai/blog/customer-segmentation-using-machine-learningb

https://towardsdatascience.com/customer-segmentation-with-clustering-933caa4c9ea3