

Maximizing Customer Value in E-commerce: Leveraging RFM Analysis and Big Data for Effective Segmentation

Abdul Baqi Malik
Masters in Data Science
Dept of Creative Technologies
Air University, Islamabad
222407@students.au.edu.pk

Faisal Mumtaz
Masters in Data Science
Dept of Creative Technologies
Air University, Islamabad
222394@students.au.edu.pk

Dr. Amanullah Yasin
Associate Professor
Dept of Creative Technologies
Air University, Islamabad
amanyasin@mail.au.edu.pk

Abstract—This research project focuses on maximizing customer value in the e-commerce sector through the integration of RFM (Recency, Frequency, Monetary) analysis and big data analytics for efficient customer segmentation. Leveraging transaction data obtained from a prominent UK-based online retail store, the study implements a comprehensive customer segmentation model based on the RFM framework and employs the K-means clustering algorithm. By classifying customers into eight distinct clusters, valuable insights into customer behavior and revenue generation potential are revealed. The findings emphasize the significance of customer segmentation as a robust tool for comprehending customer preferences and facilitating the development of targeted marketing strategies and promotional campaigns. The research also highlights opportunities for further exploration of alternative tools and methodologies to enhance the segmentation process. Overall, this study contributes to the advancement of customer segmentation practices and provides practical implications for optimizing customer value within the dynamic e-commerce industry.

Index Terms—Big data analytics, K-Means Clustering, Customer segmentation, RFM analysis, Cluster analysis, Recommendation Systems, Customer behavior analysis

I. INTRODUCTION

Supply chain management in big data analytics utilizes advanced techniques to analyze and optimize the supply chain, processing large volumes of data for insights and improved decision-making. It enables organizations to identify patterns, forecast demand, optimize transportation and inventory, enhance supplier performance, and improve operational efficiency. By leveraging big data analytics, organizations gain a competitive edge in managing their supply chain effectively.

In the era of online retail, understanding and meeting different customer needs is essential for business success. To achieve this, companies use different methods, one of which is customer segmentation. Customer segmentation involves dividing a large group of customers into smaller, distinct groups based on their characteristics, interests, and behaviors. This allows companies to align their marketing efforts, product offerings, and customer experience in certain areas, thereby improving customer satisfaction and increasing profits. [1]

In recent years, the advent of big data analytics has changed the way businesses approach customers. Big data analytics refers to the use of advanced techniques and tools to extract valuable insights from large amounts of data. This wealth of knowledge, along with the ability to organize and analyze it effectively, allows companies to gain a deeper understanding of their customers' behaviors and preferences. Among the various methods used in customer segmentation, RFM analysis stands out as a powerful method that enables large data analysis to identify and classify customers based on their shopping time (recency), frequency and how much money (monetary value) of transactions.

RFM analysis, an acronym for Recency, Frequency, and Monetary provides a comprehensive method for customer segmentation in online retail. By looking at a customer's last purchase history, the frequency of their transactions, and the financial value they bring to the business, RFM analysis allows customers to assign a numerical score to each customer, reflecting their overall value and contribution. After that, customers can be divided into different groups, such as high loyalty customers, potential customers, or one-time customers. This granular segmentation allows businesses to design targeted marketing campaigns, personalized promotions, and customized customer experiences, leading to customer satisfaction and loyalty and ultimately revenue-high income. [2]

The V's of big data (Volume, Velocity, Variety, Veracity, and Value) are of utmost importance in our research topic. The large Volume of customer data generated by online retailers allows for comprehensive analysis and segmentation. The Velocity of data processing enables real-time insights and timely decision-making. The Variety of data sources, such as transaction records, website interactions, and social media data, provides a holistic view of customer behavior. Ensuring the Veracity of data quality and accuracy is crucial for reliable segmentation results. Ultimately, leveraging these V's enables businesses to derive Value by delivering personalized experiences, targeted marketing, and improved customer satisfaction.

In this research study, we aim to determine the effectiveness of customer segmentation using RFM analysis in the context of online retail, supported by big data analytics. We will dive further into the various steps involved in the RFM analysis process, including data collection, pre-processing, and calculating the RFM score. Additionally, we will study the impact of different segmentation strategies on key business metrics such as customer retention, order value, and conversion rates. By examining real-world data from an online retail platform, we intend to provide insights into how RFM analysis, combined with big data analytics, can drive business growth and customer-centric strategies in the dynamic landscape of online retail.

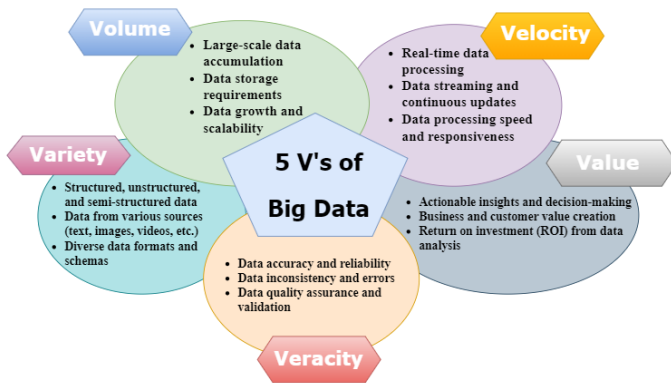


Fig. 1. 5 V's of Big Data

II. LITERATURE REVIEW

In the era of e-commerce, businesses strive to maximize customer value by effectively segmenting their customer base. This literature review aims to explore various research articles and review papers that discuss the use of RFM analysis and big data in customer segmentation for e-commerce. The insights gained from these studies can help businesses develop targeted marketing strategies, improve customer satisfaction, and increase profitability.

Customer Segmentation on Online Retail using RFM Analysis: Big Data Case of Bukku.id. This study by Mohamad Abdul Kadir and Adrian Achyar presents a case study of Bukku.id, an online retail platform, utilizing RFM analysis for customer segmentation. The authors highlight the importance of leveraging big data to analyze customer behavior and identify high-value segments, leading to personalized marketing campaigns and improved customer satisfaction. [1]

Customer Segmentation using RFM Model and K-Means Clustering Shirole, Salokhe, and Jadhav explore the application of the RFM model combined with K-means clustering for customer segmentation. Their research highlights the effectiveness of this approach in identifying distinct customer groups based on recency, frequency, and monetary value. The

findings contribute to the development of targeted marketing strategies and customer retention programs. [2]

Customer Segmentation Based On Recency Frequency Monetary Model: A Case Study in E-Retailing KABASAKAL presents a case study focused on customer segmentation in the e-retailing industry using the Recency Frequency Monetary (RFM) model. By analyzing transactional data, the study identifies different customer segments and discusses their characteristics and preferences. The research emphasizes the value of RFM analysis in understanding customer behavior and designing personalized marketing strategies. [4]

An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm Wu, Shi, Lin, Tsai, Li, Yang, and Xu present an empirical study on customer segmentation based on purchase behaviors using the RFM model and K-means algorithm. The research explores the relationship between customer purchase behaviors and their value to the business. The study contributes to the understanding of customer segmentation techniques and their application in e-commerce. [5]

A review of data mining methods in RFM-based customer segmentation Ernawati, Baharin, and Kasmin provide a comprehensive review of data mining methods in RFM-based customer segmentation. The authors explore various algorithms and techniques used to analyze RFM data, such as clustering, decision trees, and association rules. The review helps to establish a solid foundation of knowledge for researchers and practitioners interested in implementing RFM analysis in customer segmentation. [6]

A hybrid big data analytical approach for analyzing customer patterns through an integrated supply chain network Wang, Tsai, and Ciou propose a hybrid big data analytical approach for analyzing customer patterns in an integrated supply chain network. By combining RFM analysis with other data mining techniques, the authors demonstrate how businesses can gain valuable insights into customer behavior and preferences. The study emphasizes the significance of integrating multiple data sources to enhance the accuracy and effectiveness of customer segmentation. [7]

Customer online shopping experience data analytics: Integrated customer segmentation and customized services prediction mode. Wong and Wei discuss the integration of customer segmentation and customized services prediction through data analytics in the online shopping experience. The study emphasizes the importance of utilizing big data analytics to enhance customer segmentation accuracy and predict customer preferences. The findings contribute to improving the overall online shopping experience and customer satisfaction. [8]

Multi-Behavior RFM Model Based on Improved SOM Neural Network Algorithm for Customer Segmentation Liao, Jantan, Ruan, and Zhou propose a multi-behavior RFM model based on an improved Self-Organizing Map (SOM) neural network algorithm. The research demonstrates the effectiveness of this model in capturing complex customer behavior patterns and identifying customer segments with

diverse preferences. The study highlights the advantages of using advanced machine learning techniques in RFM-based customer segmentation. [9]

The impact of big data market segmentation using data mining and clustering techniques Balas and Jain investigate the impact of big data market segmentation using data mining and clustering techniques. The study emphasizes the role of RFM analysis in effectively segmenting customers and designing targeted marketing campaigns. The research provides insights into the application of big data analytics in customer segmentation and its impact on business performance. [10]

A Review On Customer Data Analytics In E-Commerce Shukla, Yagyasen, Gupta, Singh, and Shivanshu conduct a review on customer data analytics in e-commerce. The authors discuss the significance of customer segmentation for businesses and the utilization of RFM analysis and big data techniques to gain actionable insights. The review provides an overview of the current trends and challenges in customer data analytics in the e-commerce industry. [11]

The reviewed literature demonstrates the importance of leveraging RFM analysis and big data in customer segmentation for e-commerce. The studies highlight the effectiveness of RFM-based models, clustering algorithms, and advanced data mining techniques in identifying valuable customer segments, designing personalized marketing strategies, and maximizing customer value. By understanding customer behavior and preferences, businesses can enhance customer satisfaction, increase sales, and gain a competitive edge in the e-commerce market.

III. DATASET DESCRIPTION

This dataset contains all purchases made for an online retail company based in the UK during an eight month period. It is a comprehensive collection of data related to online retail transactions. It provides valuable information about customer behavior, product details, and sales transactions. The dataset includes attributes such as customer ID, invoice number, product description, quantity, unit price, and transaction date. This dataset enables researchers and practitioners to analyze various aspects of online retail, including customer segmentation, purchase patterns, and sales performance. By utilizing this dataset, researchers can gain insights into customer preferences, identify trends, and develop effective strategies for maximizing customer value in the e-commerce domain. [2]

IV. PROBLEM STATEMENT

In the field of online retail, effectively dividing customers into groups based on their purchasing behavior and identifying the most profitable segments is crucial for businesses to optimize marketing strategies and improve customer satisfaction. This research project focuses on using RFM analysis and big data analytics to tackle the following challenges: First **Segmentation**: Developing an efficient method to categorize customers into distinct groups with similar purchasing patterns, Second **Profitability**: Identifying the customer segments

that generate the highest profits, enabling targeted marketing and strategies to increase revenue and customer lifetime value, and Third Method **Comparison**: Comparing different RFM analysis techniques to determine the most accurate, efficient, and actionable approach for customer segmentation in online retail. By addressing these challenges, this research aims to help businesses gain a deeper understanding of customer behavior, tailor marketing strategies to specific segments, and maximize customer value and profitability in the online retail industry.

V. CUSTOMER SEGMENTATION

Customer segmentation is a vital strategy in the field of online retail, especially when leveraging RFM analysis and big data analytics. By segmenting customers based on their purchasing behavior and characteristics, businesses can gain valuable insights into their customer base, enabling personalized marketing efforts, enhanced customer satisfaction, and increased profitability. The combination of RFM analysis and big data analytics allows for the identification of distinct customer segments, the determination of their profitability, and the selection of the most effective segmentation method. This empowers businesses to tailor their marketing strategies, optimize resource allocation, and maximize customer value, ultimately driving long-term success in the competitive landscape of online retail [1]

VI. RFM ANALYSIS

RFM analysis, a powerful technique in the realm of online retail and customer segmentation, holds significant potential for maximizing customer value. By analyzing customers' Recency, Frequency, and Monetary aspects, businesses can gain insights into their purchasing behavior and identify patterns that allow for effective segmentation. Leveraging big data analytics, RFM analysis enables businesses to categorize customers into distinct groups with similar characteristics, providing a deeper understanding of customer preferences and needs. This method facilitates targeted marketing efforts, personalized communication, and the delivery of tailored experiences, ultimately enhancing customer satisfaction and driving business profitability. With its ability to uncover valuable customer insights and guide strategic decision-making, RFM analysis emerges as a valuable tool in optimizing marketing strategies and elevating customer-centric approaches in the dynamic landscape of e-commerce. [4]

Dimensions	Description
Recency	The duration since the date of last purchase
Frequency	The total count of purchases
Monetary	The average amount purchased

TABLE I
DIMENSIONS OF THE RFM MODEL

VII. CLUSTER ANALYSIS

Cluster analysis plays a crucial role in the context of maximizing customer value in e-commerce by leveraging RFM analysis and big data for effective segmentation. By employing cluster analysis techniques, businesses can identify distinct groups or clusters of customers based on their shared characteristics, preferences, and purchasing behaviors. This method goes beyond traditional segmentation approaches and allows for a more granular understanding of customer segments, enabling businesses to tailor their marketing strategies, product offerings, and customer experiences to specific cluster needs. With the aid of big data analytics, cluster analysis uncovers hidden patterns and relationships among customers, facilitating targeted marketing campaigns, personalized recommendations, and improved customer satisfaction. By effectively applying cluster analysis in the e-commerce domain, businesses can unlock valuable insights and optimize their customer value management strategies, leading to enhanced customer loyalty, increased revenue, and sustained growth in the highly competitive online marketplace. [1] [3]

VIII. PROPOSED FRAMEWORK

The proposed framework for maximizing customer value in e-commerce involves implementing data preprocessing and feature engineering techniques to prepare the data for analysis, followed by the application of cluster techniques such as k-means clustering to segment customers effectively. This framework aims to leverage RFM analysis and big data to identify distinct customer groups and tailor marketing strategies, ultimately enhancing customer value and driving business growth.

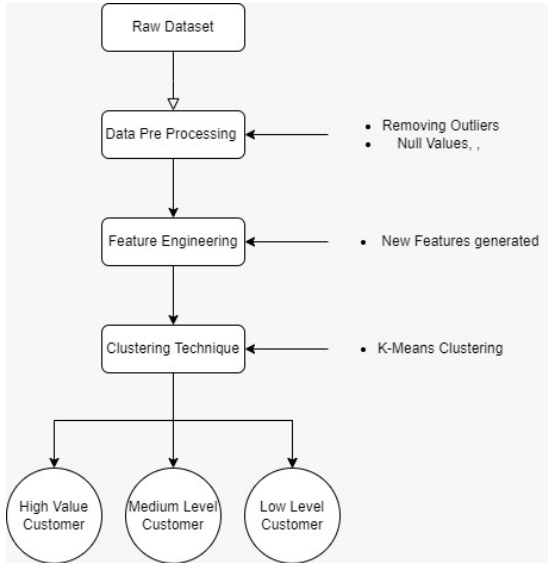


Fig. 2. Proposed Framework of our Research

IX. METHODOLOGY

A. Data Collection and Preprocessing:

The stages of a business involve focusing on understanding the purpose and meeting the needs of the business, based on its valuation. Once this understanding is achieved, an initial data mining plan is designed to effectively reach the desired goals. This study specifically examines an online retail e-commerce website operated by a UK-based retailer. The dataset used in this study encompasses sales transactions that occurred between the years 2010 and 2011. Upon collecting the data, it is crucial to become familiar with its content and gain a deeper understanding of its nature. As previously mentioned, this paper utilizes data from an online retail e-commerce website operated by a UK retailer, specifically analyzing sales transactions recorded from December 2010 to December 2011. The dataset comprises eight attributes, which are detailed in Table 1. [2]

No	Attribute	Description
1	Invoice No	Auto generated invoice with "C" prefix
2	Stock Code	Identifier for an item in stock.
3	Description	Customer's purchased item description
4	Quantity	Items purchased in a single order.
5	Invoice Date	Purchase date and time.
6	Unit Price	Product price per unit
7	CustomerID	Identifier assigned to a customer
8	Country	Shipping destination country

TABLE II
ATTRIBUTES DURING THE DATA UNDERSTANDING PHASE

Recency, frequency, and monetary values are tools used to identify an organization's top customers. The RFM model, consisting of three factors, helps analyze customer behavior: recency indicates how recently a customer made a purchase, frequency measures how often a customer buys, and monetary represents the amount spent by the customer. The model categorizes customers based on these factors, allowing for the ranking of customers in each category. [2] [8]

Recency	Frequency	Monetary	prediction
151.0	182.0	4310.0	1
97.0	31.0	1797.24	3
40.0	73.0	1757.55	3
332.0	17.0	334.4	0
94.0	95.0	1545.41	2
226.0	4.0	89.0	0
254.0	58.0	1079.4	2
236.0	13.0	459.4	0
267.0	59.0	2811.43	3
172.0	19.0	1168.06	2

Fig. 3. Feature Engineering and Prediction

B. Feature Engineering and RFM Analysis:

After preprocessing, the next step involves feature engineering, where additional customer-centric features will be derived from the collected data. These features may include customer lifetime value, purchase frequency, average order value, and recency of purchases. Subsequently, RFM analysis will be conducted, assigning RFM scores to each customer based on their recency, frequency, and monetary value. This analysis will enable the segmentation of customers into distinct groups, identifying high-value customer segments [5]

C. Cluster Analysis and Evaluation:

In this phase, cluster analysis techniques, such as k-means clustering, hierarchical clustering, or other appropriate algorithms, will be applied to the RFM scores to create customer segments. The clusters will be evaluated based on their homogeneity and heterogeneity, as well as their interpretability and relevance to business objectives. The identified clusters will then be analyzed to gain insights into customer behaviors, preferences, and profitability. Finally, the effectiveness of the segmentation approach and its impact on customer value maximization will be assessed using appropriate evaluation metrics and statistical tests. [10]

By following this process, we can apply RFM, big data and cluster analysis techniques to effectively segment customers in e-commerce, leading to customized marketing strategies, customer benefits that make it's better and improved business performance. [3]

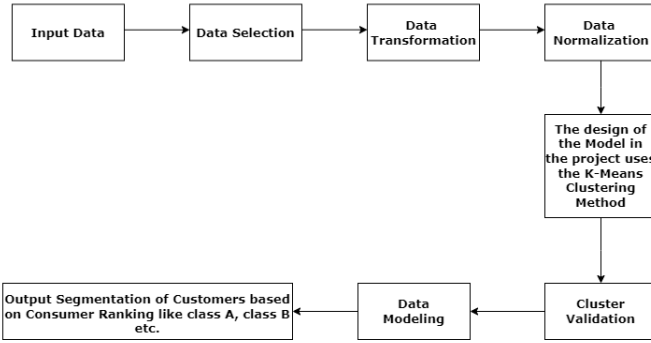


Fig. 4. Research Flow Diagram

X. RESULTS AND FINDINGS

We analyzed data from a UK online retail store for 13 months (December 2010 to December 2011) and identified four attributes. Even after preparing the data, there were some imbalances in the RMF model, so we made further adjustments. To understand customer behaviors better, we created histogram plots to visualize the distribution of Recency, Frequency, and Monetary values in the RFM analysis model, giving us insights into how customers are behaving across these dimensions.

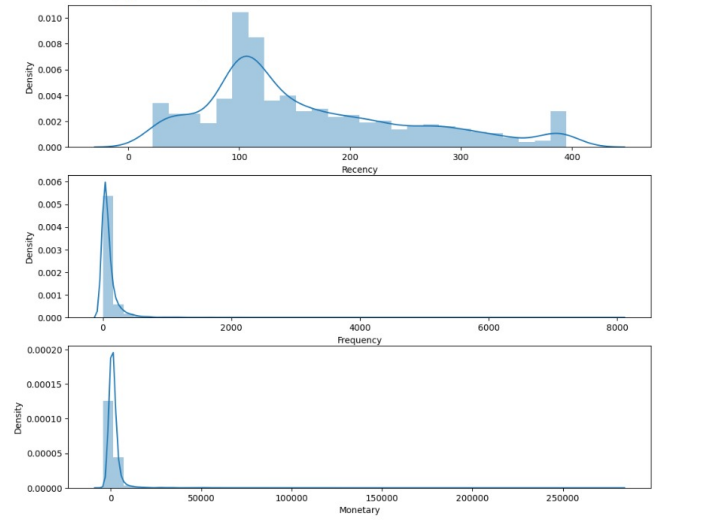


Fig. 5. Skewness in RFM Model

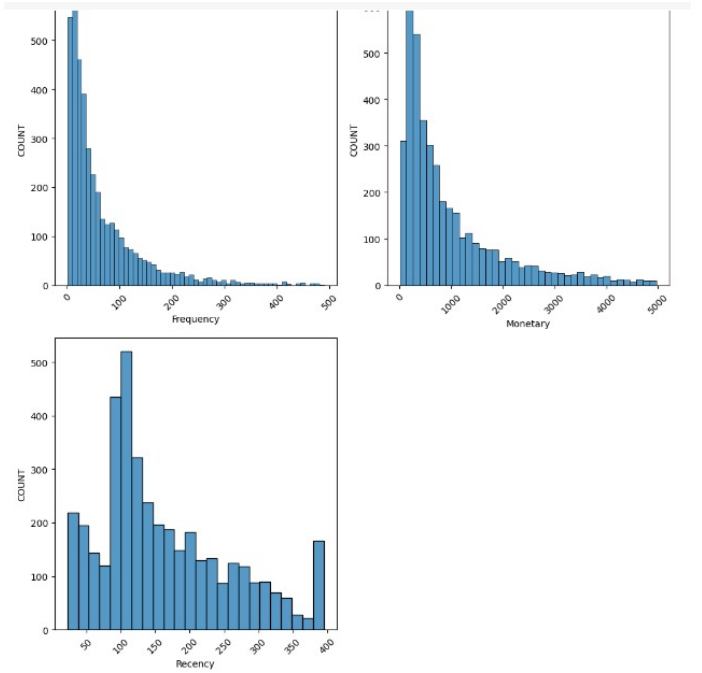


Fig. 6. Visualize data after removal outliers

We used histogram plots to visualize the data after removing outliers, providing a clearer representation of the distribution and patterns within the dataset.

The elbow method was utilized to analyze the relationship between the number of clusters and the within-cluster sum of squares. By plotting the number of clusters against the corresponding within-cluster sum of squares, we could identify the optimal number of clusters to use in the RFM analysis. The point where the plot forms an "elbow" shape indicates the number of clusters that best balances the reduction in within-cluster sum of squares and the complexity of the clustering solution. [5]

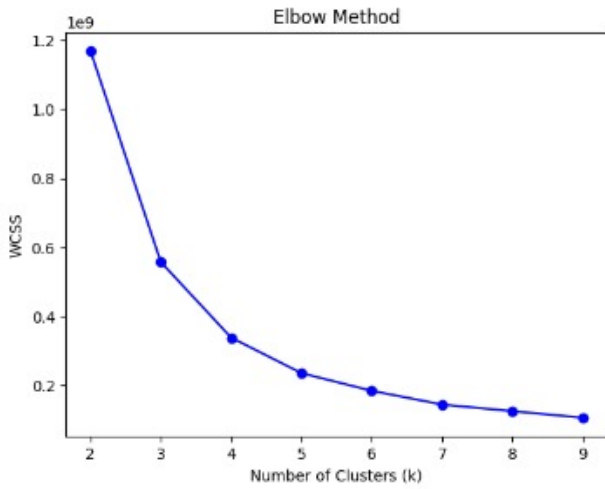


Fig. 7. visualize the data through elbow method wrt WCSS and No. of clusters

We utilized scatterplots to visualize the RFM clustering results, focusing on the relationship between the monetary value and frequency of customer purchases. This allowed us to observe any patterns or clusters formed based on these two dimensions and gain insights into the customer segmentation based on their spending habits and purchase frequency. [1]

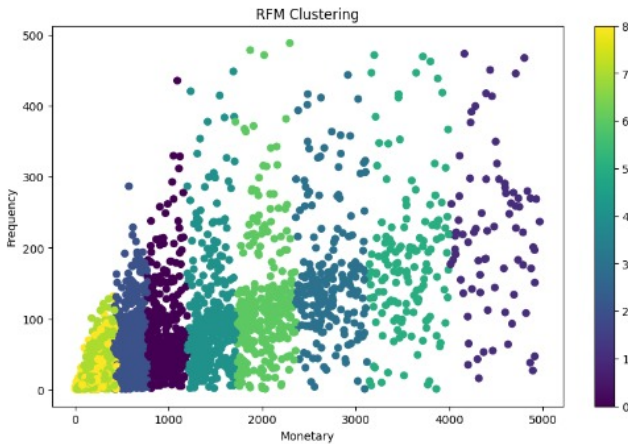


Fig. 8. visualize the data through scatterplot with respect to monetary and frequency.

Scatterplots were employed to visually represent the RFM clustering outcomes, specifically highlighting the association between the monetary value and recency of customer transactions. By examining the scatterplot, we could identify potential groupings or clusters based on customers' spending amounts and the time elapsed since their last purchase, aiding in the segmentation of customers according to their recent purchasing behavior and monetary contributions. [4]



Fig. 9. visualize the data through scatterplot with respect to monetary and recency.

XI. CONCLUSION

In conclusion, this research project focused on maximizing customer value in e-commerce by using RFM analysis and big data for segmentation. K-means clustering is identified as the most robust method for clustering. The literature review strongly supports the use of K-means clustering for customer segmentation, which is an effective process for real-time recommendations. Following customer segmentation, clusters are created, and different recommendations are applied to each cluster through a recommendation system. Each cluster has unique recommendations tailored to its characteristics. K-means clustering performs exceptionally well on large datasets. Feature engineering is applied to days, weeks, and time to extract better insights. It is observed that a minimum of three clusters yields optimal results, and the methodology can accommodate up to five clusters. The combined use of RFM analysis, clustering, and feature engineering produces the best outcomes. For future recommendations, further improvement in results and the creation of additional clusters can be explored through enhanced feature engineering techniques. Customer segmentation provides valuable insights into customer behavior and can be used to develop effective marketing strategies and promotions for loyal customers. Additionally, further comparisons with other tools and methods can be explored to enhance the developed system. Overall, this research contributes to the field of e-commerce and provides practical insights for businesses aiming to optimize their customer segmentation strategies and maximize customer value.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to our course instructor, Dr. Amanullah Yasin, for his valuable guidance and support throughout the research project. I would also like to extend a special thanks to my project mate, Faisal Mumtaz, for his collaboration and contribution to the success of this endeavor. Their involvement and assistance have been truly invaluable

REFERENCES

- [1] M. A. Kadir and A. Achya, "Customer Segmentation on Online Retail using RFM Analysis: Big Data Case of Bukku.id," European Alliance for Innovation (EAI), 2019-09-09.
- [2] R. Shirole, L. Salokhe, and S. Jadhav, "Customer Segmentation using RFM Model and K-Means Clustering," International Journal of Scientific Research in Science and Technology, vol. 8, no. 3, pp. 591-597, May-June 2021.
- [3] R. W. B. Sembiring, F. A. Mohammed, and K. Chairuang, "Customer segmentation based on RFM model using K-Means, K-Medoids, and DBSCAN methods," Lontar Komputer, vol. 11, no. 1, e-ISSN 2541-5832, April 2020.
- [4] I. Kabasakal, "Customer segmentation based on Recency Frequency Monetary model: A case study in E-Retailing," J. Inf. Technol., vol. 13, no. 1, pp. 47-56, DergiPark Akademi, 2020.
- [5] J. Wu, L. Shi, W.-P. Lin, S.-B. Tsai, Y. Li, L. Yang, and G. Xu, "An empirical study on customer segmentation by purchase behaviors using a RFM model and K-means algorithm," Math. Probl. Eng., vol. 2020, Article ID 8884227, 7 pages, Hindawi, 19 Nov 2020.
- [6] E. Ernawati, S. S. K. Baharin, and F. Kasmin, "A review of data mining methods in RFM-based customer segmentation," J. Phys.: Conf. Ser., vol. 2021, Annual Conf. ANCOSSET - 2020, IOP Publishing.
- [7] S.-C. Wang, Y.-T. Tsai, and Y.-S. Ciou, "A hybrid big data analytical approach for analyzing customer patterns through an integrated supply chain network," J. Ind. Inf. Integr., vol. 20, Dec. 2020.
- [8] E. Wong and Y. We, "Customer online shopping experience data analytics: Integrated customer segmentation and customized services prediction mode," Int. J. Retail Distrib. Manag., vol. 46, no. 4, pp. 340-354, April 2018.
- [9] J. Liao, A. Jantan, Y. Ruan, and C. Zhou, "Multi Behavior RFM Model Based on Improved SOM Neural Network Algorithm for Customer Segmentation," IEEE Access, vol. 10, pp. 122501-122512, Nov. 18, 2022.
- [10] V.E. Balas and L.C. Jain, "The Impact of Big Data Market Segmentation Using Data Mining and Clustering Techniques," Journal of Intelligent and Fuzzy Systems, vol. 38, no. 5, pp. 6159-6173, May 2020.
- [11] V. Shukla, D. Yagyasen, R. Gupta, S. Singh, Shivanshu, "A Review on Customer Data Analytics in E-commerce," International Research Journal of Modernization in Engineering Technology and Science, vol. 5, May 2023.

Our research work is an original contribution in the field of maximizing customer value in e-commerce through RFM analysis and big data for effective segmentation. While we have drawn guidance from other research papers, we have ensured to provide proper credit through citations, acknowledging the valuable insights and contributions from previous studies.

Abdul Baqi Malik and Faisal Mumtaz