Intelligent Customer Segmentation in Digital Commerce Using K-Means Clustering

Damla Demir[1]; Gökçe Karahan Adalı[2]

[1]Graduated Student, Department of Management Information Systems, Haliç University Turkey
[2]Assistant Professor, Department of Software Engineering,İstanbul Sabahattin Zaim University, Turkey

Publication Date: 2025/12/16

Abstract: The rapid rise in e-commerce has forced companies to have good knowledge of customer behavior and tailor the marketing strategies accordingly. This study discusses the appropriateness of the K-Means algorithm for customer segmentation from behavioral and demographic data obtained by a systematic Likert-scale survey. Clusters with high interpretability were obtained and validated through silhouette analysis with values up to 0.75, indicating high internal consistency. Key findings show that female interviewees prefer shopping by mobile to a far greater extent than male interviewees, while male interviewees are more responsive to promotional emails and SMS. Younger and middle-aged users are similarly more susceptible to social media advertising, with older segments having more neutral or selective orientations. These results illustrate the complexity of customers' behavior and that demographic and behavioral data should be combined in segmentation studies. By its demonstration of the value of clustering techniques in providing insightful customer profiles, this study contributes to practical and methodological applications to data-driven decision-making in e-commerce. Future research is encouraged to expand the dataset size and incorporate more advanced methods such as predictive modeling and sentiment analysis to further improve segmentation precision.

Keywords: *K-Means Clustering, Customer Segmentation, E-commerce Analytics.*

How to Cite: Damla Demir; Gökçe Karahan Adalı (2025) Intelligent Customer Segmentation in Digital Commerce Using K-Means Clustering. *International Journal of Innovative Science and Research Technology,* 10(12), 757-768. https://doi.org/10.38124/ijisrt/25dec513

I. INTRODUCTION

In today's technological world, institutions and organizations are undergoing digital transformation. The increasing number of e-commerce websites clearly demonstrates how critical this transformation is for businesses. The most important way to keep up with this change is by analyzing customers in the best possible way. Just as we recognize our customers' faces, greet them, and communicate with them in real life, we must be able to do the same in digital environments. In other words, we need to thoroughly understand our customers' behaviors and characteristics in the digital space as well. Establishing an e-commerce website allows businesses to get to know their customers better, conduct various analyses based on their needs, and accurately evaluate these analyses using real data. This is achieved through segmentation with accurate data. "Market segmentation is one of the most important key points to achieving goals for all businesses. Market segmentation is a process to divide customers into homogeneous groups which have similar characteristics such as buying habits, life style, food preferences etc." [10]. Segmentation helps businesses tailor their marketing approaches, improve customer satisfaction, and make decisions based on insights into customer behaviors and demographic information [11]. Clustering is frequently applied for customer segmentation across various sectors like telecom, e-commerce, sports, advertising, and sales [3].

Planning, organizing and leading these segments' data is a very valuable area which demonstrates knowledge management level of businesses. In here, businesses must arrange their own budget. For example, mass production amount, profit amount or sales amount must be analyzed very carefully by the strategic planning department of the enterprise. Market segmentation's core purpose is to predict customer needs, driving profitability by optimizing product availability (quantity, timing, and cost) for the target audience" [8]. Segmentation helps businesses tailor their marketing approaches, improve customer satisfaction, and make decisions based on insights into customer behaviors and demographic information [11].

Customer segmentation plays an important role for businesses to better understand their target audience and to optimize marketing efforts. Over time, segmentation strategies have turned into AI-driven techniques from traditional approach.

➢ *Traditional Segmentation:*

Demographic Segmentation Based on age, gender, income and occupation.

Geographic Segmentation Customers are segmented based on their location such as country, city, regional.

Psychographic Segmentation – Groups customers by lifestyles, values and interests.

Behavioral Segmentation – Focuses on purchasing behavior, using of products periodically, brand loyalty.

This segmentation results come from generally based on surveys, market research and analyses and historical sales data.

➢ *AI Based Segmentation:*

Clustering Algorithms (K-Means, DBSCAN, Hierarchical Clustering) – Techniques, algorithms make grouping to the customers based on similarities in data.

Predictive Analytics – It uses past data to forecast and guessing the future conditions, actions.

Real-Time Data Processing – AI adapts to changes in customer preferences dynamically.

Personalized Recommendations – AI can develop different messages or create different recommendations to each individual customers based on the customer preferences and buying habits. AI can join an interactive communication with own customers. This segmentation is more adaptive, adjustable and accurate, helping businesses reach highly improved positions when compared to their competitors.

The aim of this research is to serve as a valuable resource for data scientists and business analysts, while also providing essential guidance for businesses seeking to adapt to digital transformation. This article focuses on how businesses, particularly those operating their own e-commerce systems, can effectively segment their customers and leverage the k-means clustering technique when designing targeted advertising strategies. The methodological structure of this study loosely aligns with the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework, which is widely recognized as a standard for organizing data mining projects [21].

➢ *The Role of AI and Data Mining in E-Commerce*

Artificial intelligence has permeated nearly all sectors in the modern era. In digital spaces, individuals generate a wealth of personal data, encompassing platform registration dates, financial details like bank account information and balances, and browsing histories. AI systems actively monitor and analyze a spectrum of user interactions, including website activity, purchasing patterns, and inclinations toward discount tracking. The e-commerce sector has been profoundly influenced by AI, which has led to better customer interactions, more efficient processes, and increased business development [1].

E-commerce platforms (web applications, websites) utilize data mining techniques to understand customer behavior, increase sales, prevent fraud, and monitor and optimize overall business processes. Customers' shopping habits, preferences, and demographic information are collected and analyzed through data mining. One of the most used clustering algorithms for customer segmentation is the K-Means technique. Globally recognized companies such as Amazon and Netflix leverage data mining techniques to offer product recommendations based on customer preferences.

AI's ability to customize e-commerce experiences stands as a top marketing strategy. By analyzing user-specific data, AI and machine learning uncover key trends. For example, AI-based tools can monitor customer behavior across various digital platforms, such as emails, apps, and websites, to assess their online engagement [14].

➢ *Clustering Techniques for Customer Segmentation*

A well-executed customer segmentation strategy helps e-commerce businesses enhance brand loyalty, improve customer retention, and differentiate themselves in the global market. By understanding customer needs and preferences, companies can offer tailored products or services, ensuring a more personalized experience. In this segmentation process, clustering algorithms are widely used to identify similar groups. Clustering, as an unsupervised learning method, groups customers based on specific characteristics [4]. As a result, businesses can refine their strategies, optimize CRM service processes, and deliver the best possible performance to their customers. Additionally, effective segmentation aids in optimizing pricing strategies, financial planning, and overall economic investments. By applying clustering techniques, customers can be grouped based on their online shopping behaviors, enabling a deeper understanding of their purchasing patterns. The insights gained from this process shed light on crucial customer expectations in the digital marketplace. Leveraging this information allows businesses to refine their customer relationship management strategies, ultimately enhancing satisfaction and fostering long-term brand loyalty [23].

One significant benefit of clustering lies in its capacity to minimize data storage needs and computational effort by summarizing multiple features under unified cluster labels. It also aids in addressing missing values by inferring them from similar instances within the same cluster. Moreover, clustering contributes to data privacy by linking data to group-level patterns rather than to individual-level records [7].

➢ *K-Means Clustering Implementation*

The K-Means clustering method is one of the best techniques for customer segmentation. It is a method that does not rely on any labeled data, allowing the machine to analyze the provided data without making any forward-looking predictions. Instead of this, it assesses the current state of the business and explains how it should act within the short term. The k-means clustering algorithm operates by categorizing data points into clusters by using a mathematical distance measure, usually euclidean, from the cluster center. The objective is to minimize the sum of distances between data points and their assigned clusters. Data points that are nearest to a centroid are grouped together within the same category. A higher k value, or the number of clusters, signifies smaller

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec513

clusters with greater detail, while a lower k value results in larger clusters with less detail. “This form of cluster analysis is broadly utilized in various data science applications, such as dividing markets into segments, organizing text documents, segmenting images, and reducing image size. K-means is a preferred technique in these areas due to its simplicity, speed, and reliable performance.” [8]. According to Kuhn and Johnson [12], unsupervised learning methods such as K-means clustering are often used as a precursor to supervised modeling tasks, helping to reveal latent patterns within the data. The K-means algorithm encounters difficulties in selecting the most suitable number of clusters.” [16]. While K-Means is a fast and efficient method for customer segmentation, its inability to directly adapt to changing customer behavior remains one of its major limitations.

II. LITERATURE REVIEW

In recent years, numerous studies have employed clustering techniques, particularly the K-Means algorithm, to perform customer segmentation based on survey data. These studies typically follow a structured methodology involving survey design, Likert-scale responses, data preprocessing (e.g., normalization and feature selection), application of K-Means, and evaluation using methods like silhouette analysis and PCA.

For instance, [6] demonstrated the effectiveness of clustering techniques on survey-based customer data by grouping telecom users according to service usage patterns. Their methodology, like this study, involved converting categorical responses into numerical form and applying K-Means to identify homogeneous customer groups. [20] conducted a similar segmentation study using survey data in the retail sector. They emphasized the importance of feature selection in determining cluster quality and used PCA to reduce dimensionality before clustering, paralleling the approach taken in this research.

Some researchers applied K-Means on Likert-scale e-commerce survey data to discover customer segments with varying preferences for digital payment options [13]. Their data preprocessing steps included encoding categorical variables and normalization, followed by silhouette analysis to assess clustering validity.

Some researchers focused on segmenting users of an online platform using a methodology that closely resembles the current study: a self-administered questionnaire, application of K-Means, visualization through PCA, and validation using silhouette scores [24]. The study demonstrated that survey-based segmentation is both scalable and interpretable, particularly when clusters are used to design differentiated marketing strategies. Similarly, [15] employed an optimized K-Means algorithm on transactional retail data and identified three meaningful customer clusters that supported targeted marketing initiatives. [5] combined PCA and K-Means in their analysis of online retail data, which improved the clarity and accuracy of the resulting segments by reducing dimensionality prior to clustering. In a more advanced framework, [21] integrated Q-learning–based differential evolution with K-Means and PCA, achieving over 95% classification accuracy and demonstrating the potential of reinforcement learning in segmentation workflows. Additionally, [18] introduced K-Sil, a silhouette-guided variation of K-Means, which gave greater weight to well-clustered instances and led to improved internal cohesion and cluster separation. [17] also analyzed purchasing behavior on the Amazon platform using K-Means and identified five distinct customer segments, illustrating how behavioral data can be used to enhance personalization strategies.

These recent contributions reinforce the methodological validity of combining clustering techniques with survey or transactional data, particularly in contexts where customer-centric decision-making is prioritized [9].

By investigating several clustering methods to segment customers in the UK retail sector, finding that K-means clustering effectively identified homogeneous purchasing groups when applied to large-scale online retail data. In their comparative analysis, the authors report that K-means clustering outperformed alternative methods—such as DBSCAN and GMM—in yielding interpretable and actionable customer segments from a dataset of over 540,000 records.

Tools like Orange and R have also been utilized in several empirical studies. For example, [2] used Orange’s drag-and-drop interface to perform K-Means clustering on user satisfaction surveys in an academic context. Their research underscored the accessibility of visual data mining tools in academic and business environments alike.The combination of statistical tools (such as R) with interactive environments (like Orange) increases the accessibility and replicability of such segmentation efforts, especially for SMEs.

III. METHOD

This study employs the K-means clustering algorithm for customer segmentation. The analysis is designed to uncover homogeneous customer groups (segments) within the dataset. These segments are intended to inform marketing decisions, improve personalization (differentiation strategies), and enhance customer relationship management strategies.

The data analysis was conducted using both Orange and R software environments to ensure methodological robustness and reproducibility. Orange was utilized for its user-friendly, visual programming interface, allowing for rapid prototyping of the data mining workflow, including preprocessing, clustering, and visualization through PCA. In parallel, R was employed for statistical validation, calculation of silhouette scores, and supplementary visualizations. The combined use of Orange and R enabled a comprehensive and cross-validated approach to customer segmentation based on the survey data.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec513

➢ *The Following Research Questions were Addressed Throughout the Study:*

- RQ1: Do women enjoy shopping via mobile applications more than men? (This insight can guide businesses in designing or adjusting their mobile app interface accordingly.)
- RQ2: Does the average monthly income influence whether individuals shop regularly on e-commerce platforms?
- RQ3: Are discount messages from e-commerce websites followed more by men or women?
- RQ4: Is there a relationship between the ages of participants and the effectiveness of social media ads? (This can help determine which type of customers should receive such ads, especially considering advertising costs.)

➢ *Research Design*

This study aims to perform customer segmentation using clustering techniques, with a particular focus on the K-means algorithm. The research design is based on an applied quantitative research approach. Data was collected from a self-administered survey. The research employed quantitative and descriptive design, aiming to understand consumer purchasing behavior in the context of e-commerce. A structured questionnaire consisting of 22 questions was developed for this purpose. The first six questions focused on demographic variables, including gender, age, marital status, household size, education level, and monthly income. The remaining questions were designed to capture behavioral and psychographic characteristics such as purchasing criteria, shopping habits, and customer priorities when using e-commerce platforms. These items were specifically formulated to facilitate customer segmentation and to explore the underlying patterns in consumer behavior. To minimize response bias and increase comparability, the questions primarily utilize a Likert scale format.

➢ *Data Collection Method*

The data collection method in this study primarily relies on primary data, which was obtained through a self-administered survey. The survey was designed and distributed by the researcher to collect relevant information directly from participants, specifically focusing on customer behaviors, preferences, and demographic characteristics. The survey included closed-ended questions. To facilitate the conversion of qualitative responses into numerical data, the survey incorporated a Likert scale.

This scale allowed participants to express their opinions and attitudes on various topics using a set of predefined response options, typically ranging from strongly agree to strongly disagreement. By assigning numerical values to each response, the qualitative data could be easily quantified, enabling more straightforward analysis. The use of the Likert scale ensures that subjective opinions can be represented in a consistent and measurable way, making it suitable for statistical analysis and clustering techniques.

The data was collected through an online questionnaire created using Google Forms. The survey was conducted online and made available to participants through digital platforms. The dataset was obtained through a survey conducted with 78 participants, consisting of 36 males and 36 females.

The data was collected from participants living in Turkey. In terms of educational background, 18 participants held a high school diploma, 6 had completed a master's degree, 2 had completed primary school, and 2 had completed secondary school. The age distribution of the participants was as follows: 28 participants were between 18–24 years old, 13 were between 25–34, 21 were between 35–44, 10 were between 45–54, and 6 participants were aged 55 or above. In terms of monthly income, 40 participants reported earning between 30,000 and 50,000 TRY (Turkish Lira), 24 participants between 50,001 and 70,000 TRY, 11 participants between 70,001 and 90,000 TRY, and 3 participants reported earning more than 90,000 TRY. To ensure the completeness of the dataset and the consistency of responses, all questions were made mandatory, preventing participants from submitting the form without answering each item. In addition, the use of Likert-type scales helped reduce the likelihood of extreme or inconsistent answers (outliers), thereby enhancing the reliability of the collected data. Ensuring data accuracy and consistency is a crucial step in making datasets more reliable and meaningful (a process commonly known as data cleaning or data preparation). This process involves detecting flawed or contradictory entries and resolving them effectively [16].

➢ *Data Preprocessing Techniques*

A preliminary PCA was conducted to assess the data structure, which revealed that the first principal component accounted for nearly all the variance. However, the clustering analysis was carried out using the original variables.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec513

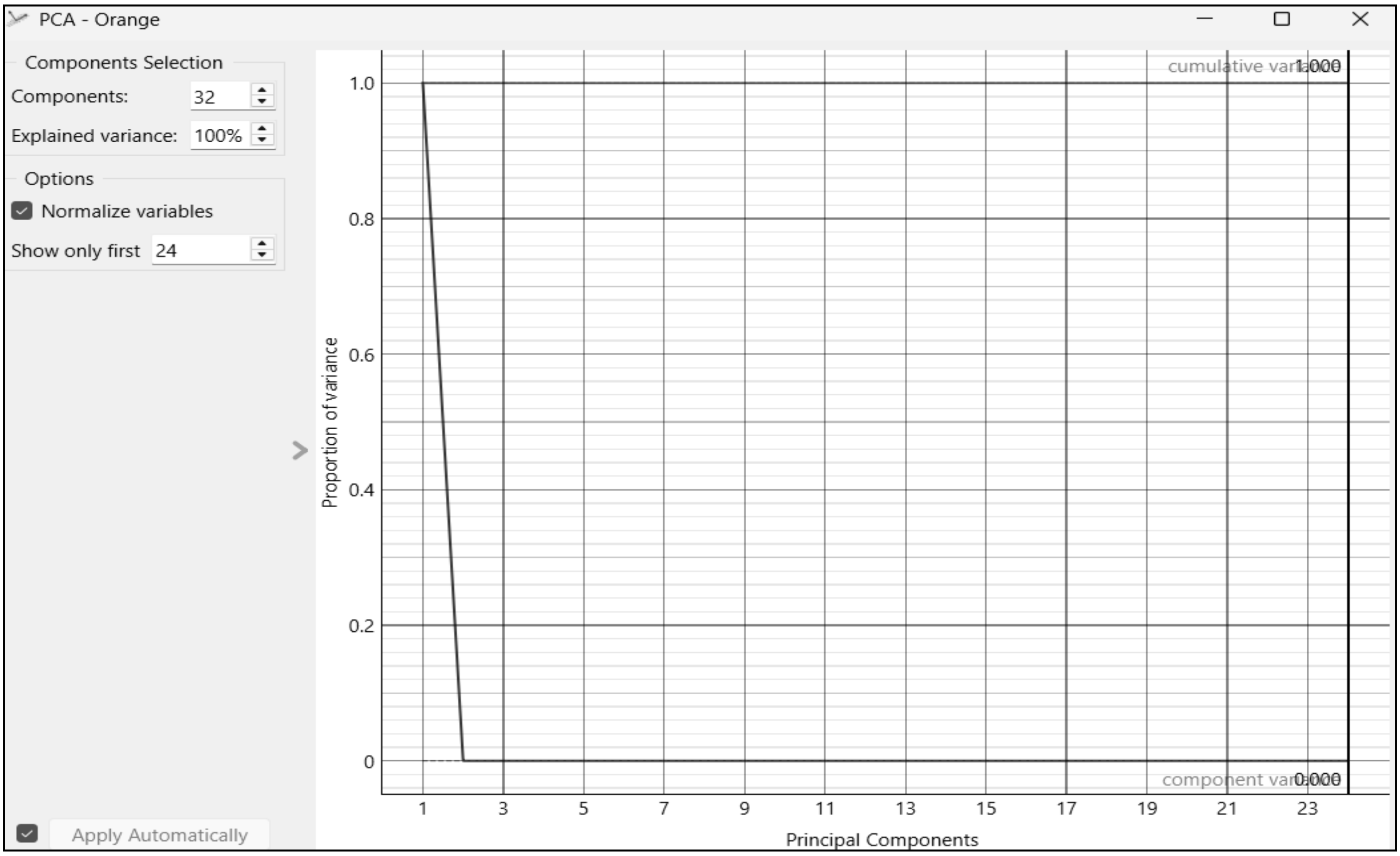

Fig 1 PCA Report (Created by the Author Using Orange, 2025)

According to the Scree Plot obtained from the PCA analysis, the first principal component explains almost all (100%) of the entire variance. This shows that the basic structure of the dataset can be largely represented in a single dimension. The contribution of the other components to the variance is negligible.

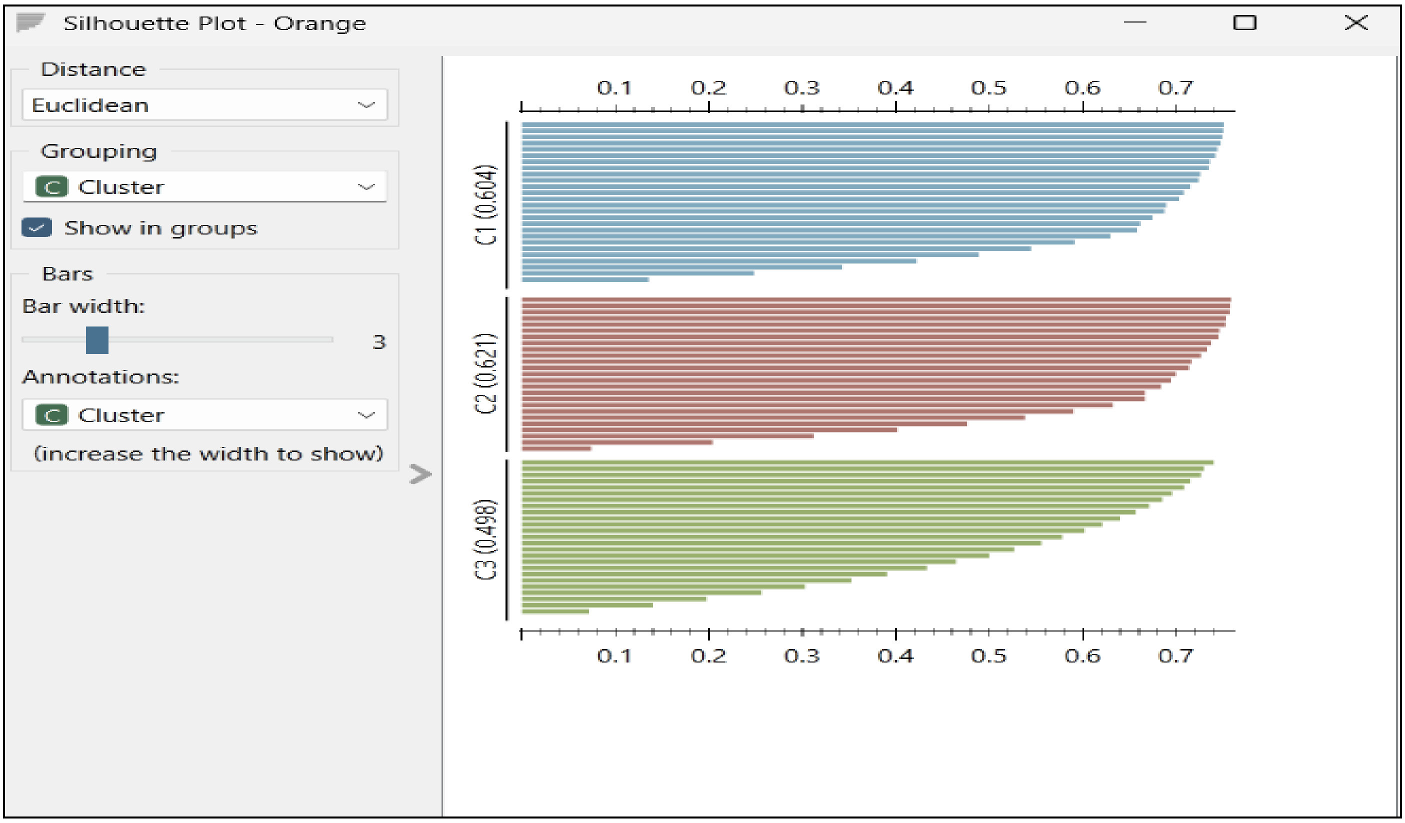

Fig 2 Silhouette Plot (Created by the Author Using Orange, 2025)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec513

Silhouette value indicates how well each data point fits into its assigned cluster, in other words, how strongly it belongs to that cluster. According to Figure 2, each cluster has positive silhouette scores that are close to 1, which means the points within each cluster are tightly grouped and well-separated from other clusters. The silhouette value ranges from 0.1 to 0.7, with most data points leaning toward 0.7. This suggests that the clusters are clearly separated, most of the data points have been assigned to the correct cluster, and there is minimal overlap between clusters. Ultimately, this indicates that the analyses we perform later will likely have a high level of accuracy.

IV. FINDINGS

In this section, various analyses were carried out to examine the relationships between demographic variables and participants' online shopping behaviors. Specifically, clustering techniques were employed to identify potential patterns and segment participants based on their purchasing habits and income levels. These analyses were conducted using both Orange Data Mining and R programming environments to ensure methodological rigor and reproducibility.

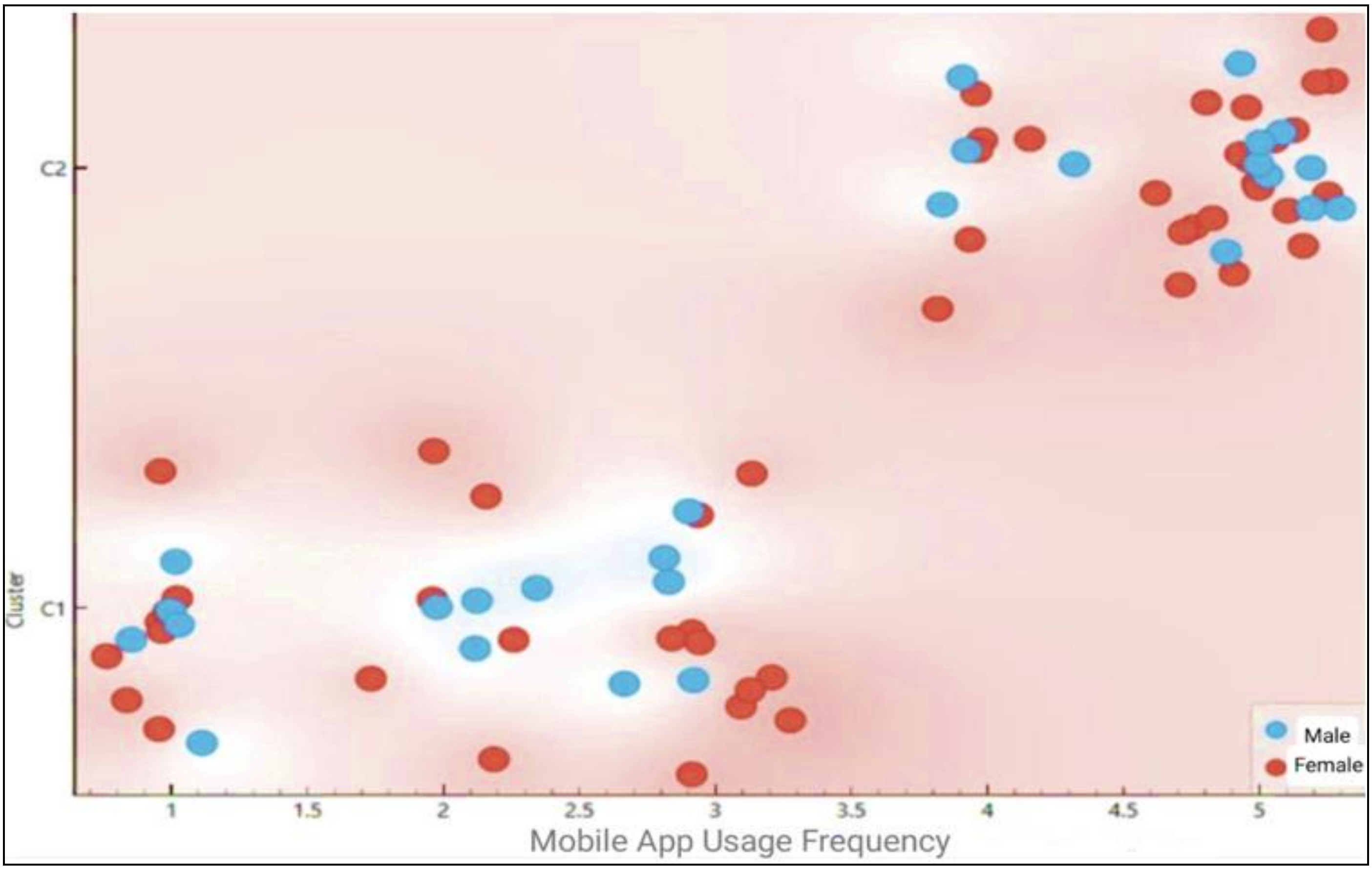

Fig 3 Gender Clusters Regarding Via Mobile App. Purchasing (Created by the Author Using Orange, 2025)

Figure 3 shows the distribution according to cluster groups. The graph is divided into 2 clusters (C1 and C2). Cluster C1 is denser with lower scores (between 1-3), meaning that individuals in this group prefer shopping from mobile applications less. Cluster C2 is denser with higher scores (4-5). This group has a more positive view of mobile shopping.

Women (red dots) are generally present in both clusters but are especially concentrated at scores of 4 and above. This may indicate that women are more prone to mobile shopping. Men (blue dots) are more visible in cluster C1 and are clustered at lower scores (between 1-3). This may suggest that men are more distant from mobile shopping. Cluster C1: Consists of individuals who prefer shopping from mobile applications less. Although both women and men are included in this cluster, men are predominant. Cluster C2: Consists of individuals who prefer mobile shopping more. Female participants are more concentrated in this cluster.

Most of the red dots are concentrated at levels 4 and 5. This shows that women highly prefer shopping from mobile applications. There is a stronger tendency among women in this regard; they are almost homogeneously clustered in high scores. Men are spread out at lower levels: they have a wide distribution between 1 and 5. This shows that there is more diversity in mobile shopping preference among men, with some men embracing this shopping style and others staying away from it.

As part of this process, a clustering analysis was performed using Orange to investigate whether average monthly income influences regular shopping behavior on e-commerce websites.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec513

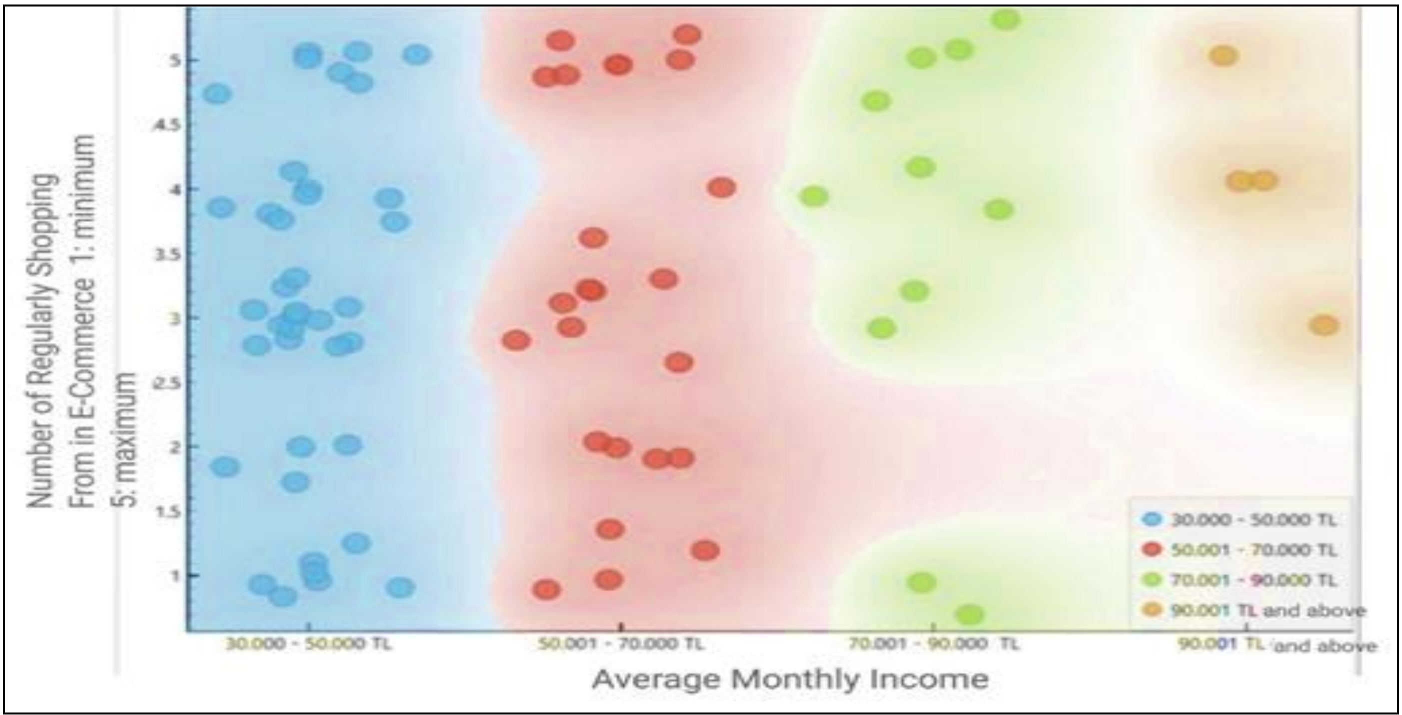

Fig 4 Income Based Shopping Cluster Analysis (Created by the Author Using Orange, 2025)

Participants with a monthly Income of 30.000–50.000 TL were mostly concentrated in the 1–3 range, indicating that they either do not shop regularly or are undecided. Regular shoppers in this segment were relatively few.

Participants with a monthly income of 30.000–50.000 TL predominantly selected values between 1 and 3, indicating that they either do not shop regularly or are undecided about online shopping. Regular buyers were relatively limited within this segment. In contrast, those earning 50.001–70.000 TL showed a response concentration between 3 and 5, suggesting more consistent online shopping habits. This income bracket likely includes individuals who are both financially stable and more familiar with digital platforms.

The 70.001–90.000 TL income category showed clustering around scores 3 and 4, pointing to a generally positive tendency toward online shopping, though full agreement (level 5) was less frequently observed. Finally, the 90.001 TL and above segment had the lowest representation in the sample. Their responses ranged between 2 and 4, reflecting a more moderate or selective engagement with e-commerce. This may imply a more discerning approach to online purchases among higher-income individuals.

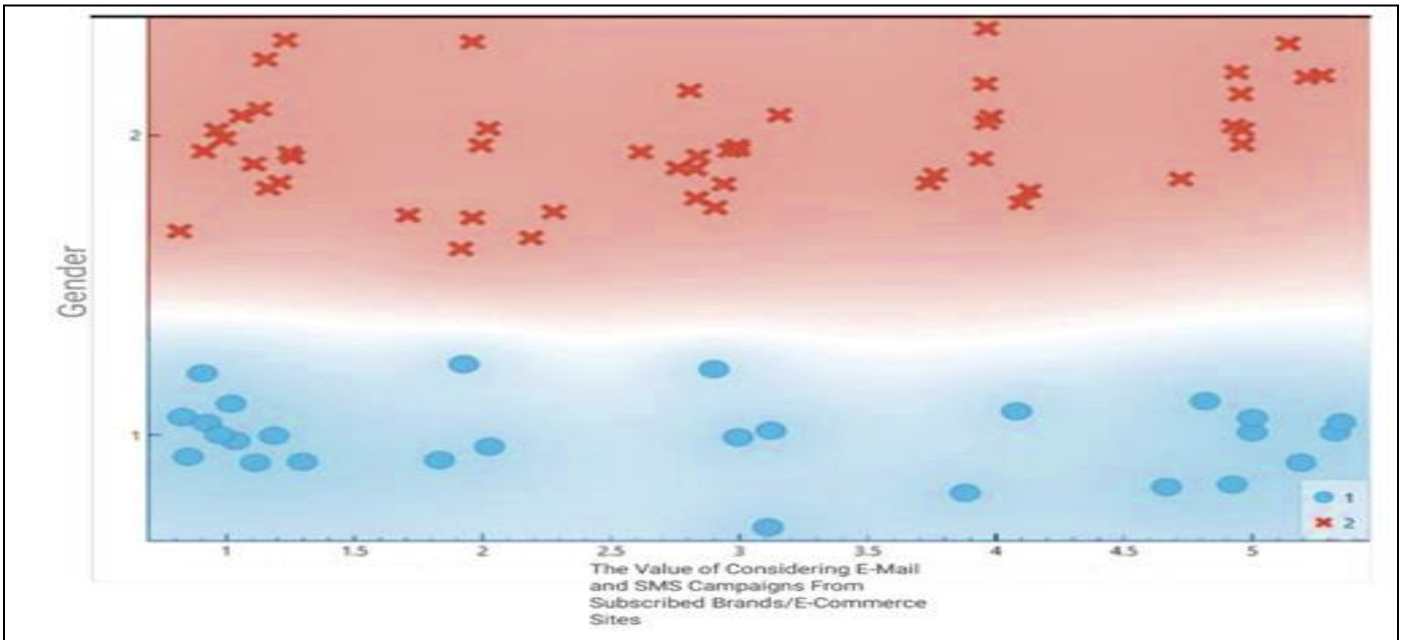

Fig 5 Cluster Analysis of Following Messages (Created by the Author Using Orange, 2025)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec513

Figure 5 presents the distribution of participants' responses regarding their attention to email and SMS campaigns sent by subscribed e-commerce brands, segmented by gender. The x-axis represents agreement levels on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), while the y-axis shows gender (coded as 1 = female, 2 = male).
To better understand patterns within the data, a clustering analysis was performed using Orange. The plot visualizes the resulting clusters with color-coded regions and point distributions, enabling a clearer interpretation of group-specific behaviours. The analysis reveals that female participants (blue) are mostly concentrated between agreement levels 1 and 3, indicating relatively low interest in marketing communications. Very few female respondents selected levels 4 or 5, suggesting limited responsiveness to promotional messages. In contrast, male participants (red) are more widely distributed across the higher end of the scale, with a notable number selecting levels 4 and 5. This suggests that men in this sample are more inclined to engage with brand-generated marketing content.

These findings indicate a gender-related difference in responsiveness to digital marketing, with male participants appearing more receptive. Clustering helped to highlight this behavioural segmentation, offering potential guidance for targeted communication strategies in e-commerce. To explore the relationship between age and the effectiveness of social media advertising, a clustering analysis was performed on participant responses to the relevant survey item. Understanding how different age groups perceive digital advertisements is critical for designing age-appropriate marketing strategies in e-commerce. The findings were visualized to reveal patterns of agreement across age segments.

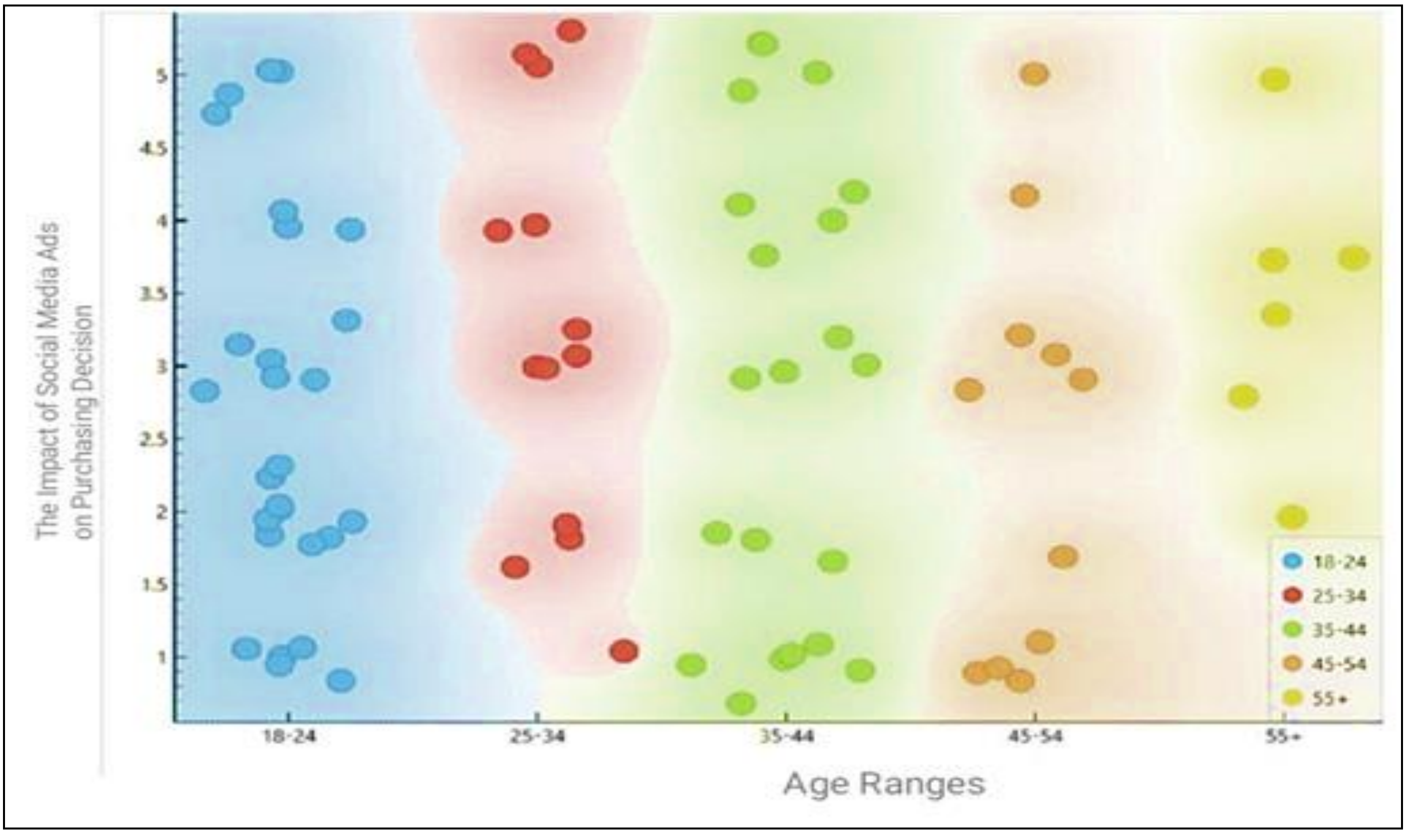

Fig 6 Relationship Between Age and Power of Ad. Effective (Created by the Author Using Orange, 2025)

Figure 6 indicates the responses of respondents regarding the influence of social media ads, segmented according to age groups and measured on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The 18–24 years old age group (blue), representing most of the sample, demonstrated a widespread primarily between levels 2 and 4. The presence of extreme values (1 and 5) suggests high individual diversity in ad responsiveness within this group. Participants aged 25–34 years (red) demonstrated closer grouping between 2 and 4, representing a moderate and more consistent perception of social media ads. The 35–44 age group (green) showed answers bunched around 3 and 4, suggesting a high level of agreement with the effectiveness of such advertisements. Smaller in size but still showing agreement levels between 3 and 4, the 45–54 group (orange) suggested that this age group is still active and open to social media marketing instruments. Last, the 55+ age group (yellow), the lowest represented age group in the sample, was dispersed between 2 and 4, suggesting a more conservative or neutral stance towards social media advertising.

These patterns indicate that age plays a role in shaping perceptions of social media ads, with younger and middle-aged adults showing greater variability and susceptibility, whereas older adults demonstrate more consistent yet reserved responses.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec513

V. DISCUSSION AND CONCLUSION

This section discusses the clustering analysis results of the data obtained through the survey form used in the study and aims to transform these results into meaningful findings. The data was analyzed using the Orange data mining software and a segmentation process was performed on various questions reflecting the participants' attitudes towards e-commerce shopping.

An appropriate number of clusters were determined for each question group and these clusters revealed different tendencies in terms of consumer behavior. In line with the analysis steps technically explained in the methodology section, this section will focus on the behavioral characteristics of the clusters that emerged. The findings are not only numerical but also interpretative in nature and are aimed to provide clues about the customer profile of each cluster. In this way, e-commerce sites will be able to make more strategic decisions regarding the shopping habits, trust levels, digital literacy and expectations of different user groups.

As a result of cluster analyses, it was observed that some participant groups were quite prone to technology and online shopping, sensitive to discounts and campaigns; while some groups were more distant towards the digital environment, lacking confidence or exhibiting indecisive tendencies. This situation shows that users do not exhibit a homogeneous structure in terms of e-commerce businesses and that reaching every user with the same marketing approach may not be sufficient. Therefore, the findings obtained reveal how important consumer segmentation is in marketing communication.

In this section, the importance of the determined clusters for e-commerce businesses will also be emphasized; in particular, how concepts such as customer loyalty, perception of trust, and user experience are shaped in terms of intra-cluster differences will be explained. As a result of all these analyses, it is emphasized that businesses need to develop their digital strategies in a more targeted manner by recognizing user profiles more accurately.

In this section of the study, a detailed evaluation of the clustering analyses applied to understand the multivariate structure of the survey results obtained during the data collection process was made. In line with the analysis steps specified in the previous section, the K-means algorithm was applied to the observations related to each question using the Orange data mining tool and meaningful clusters were obtained depending on the internal structure of the data. In this section, unlike the general analysis flow specified in the methodology, the focus was on the logical basis of the applied technical process and the qualitative representation forms of the clusters.

The basic working principle of the K-means algorithm is to group observation points based on Euclidean distance, to maximize the internal similarity of each group (cluster) and to increase the difference between clusters. In this study, the number of clusters was determined based on the natural categories of some predefined variables (e.g. salary levels, shopping frequency, etc.). However, due to the nature of the algorithm, observations in similar value ranges can be collected in the same cluster, and this may cause partial differences between theoretical classifications and practical cluster distributions.

As a result of clustering, each individual was assigned to a cluster, and the Silhouette coefficient was used for the internal consistency of the clusters. The fact that this coefficient was 0.75 for both clusters shows that the clusters are quite distinct and meaningful. If the Silhouette value is greater than 0.5, clustering is reliable, and in this respect, a solid segmentation has been made. Women represent cluster C1, and men represent cluster C2. The fact that cluster C1 collects higher scores indicates that women are more prone to mobile shopping. This points to the applicability of gender-differentiated design strategies in the user interface (UI/UX) development process. The data mining process can be used as a powerful predictive tool in terms of customer segmentation in marketing strategies. It can be suggested that more visual, intuitive and personalized interfaces be developed for female users, especially in e-commerce mobile applications.

Each analysis mentioned above not only classifies the individual responses of the participants but also reveals how diverse and dynamic customer behavior is in the e-commerce sector. Therefore, using such segmentation data while creating digital marketing strategies will make significant contributions to improving user experience, target audience management and customer loyalty.

A. Summary of Key Findings

This study addressed the issue of customer segmentation in the e-commerce sector and conducted research to understand customer behavior using data mining techniques. The aim of the study was to reveal how customer segmentation can make companies' marketing strategies more effective and the impact of personalized services on customer satisfaction. In the research process, different customer groups were created using cluster analysis and machine learning techniques and an in-depth analysis was conducted on the shopping habits, demographic characteristics and interaction levels of these groups. As a result of the analysis, it was seen that certain customer segments have significantly different shopping behaviors from each other. For example, it was determined that young users shop more frequently and at lower prices, while middle-aged customers prefer to shop at higher prices. As a result, it has become clear that e-commerce companies need to develop targeted marketing strategies by segmenting their customers. Segmentation plays a major role not only in increasing sales but also in creating customer loyalty. In addition, analysis of customer behavior will enable companies to make decisions on better stock management, campaign planning and personalized offers.

The study's results reveal distinct customer segments, each characterized by unique purchasing tendencies and engagement levels. These findings provide valuable insights for businesses to:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec513

- Develop targeted marketing campaigns tailored to different customer groups.
- Optimize product offerings to increase customer satisfaction.
- Improve customer retention strategies by more effectively understanding shopping behaviors.

Considering the findings obtained in this study, some important recommendations are presented for e-commerce companies:

- Customer Data Collection and Analysis Should Be Prioritized: E-commerce platforms should continuously collect data to analyze the behaviors and interactions of their customers more effectively. This data will form the basis for more accurate and effective customer segmentation. Proper collection and analysis of data is critical to the success of marketing strategies.
- Dynamic Segmentation Applications Should Be Implemented: Traditional segmentation models classify customers based solely on their demographic information. However, there may be cases where this model is inadequate. More dynamic and constantly changing segmentation strategies should be developed. Online shopping habits are changing rapidly, so segmentation strategies should be reviewed periodically.
- Machine Learning Models Should Be Used: In addition to machine learning methods such as cluster analysis used in this study, it is also recommended to apply more advanced techniques such as predictive modeling. Predicting future customer behavior can provide great advantages for inventory management and sales forecasts.
- Long-Term Strategies Should Be Developed for Customer Loyalty: Achieving customer loyalty requires more than just making sales. In this context, long-term strategies that will add value to customers should be implemented. Customers can be offered rewards, advantages and special offers not only for shopping but also for showing loyalty to the brand.
- Data Security and Privacy Should Be Considered: The security of customer data is one of the most important issues for e-commerce companies. Proper protection of data and providing customers with secure shopping environments will increase customer trust. In this context, compliance with data security laws should be ensured and customer privacy should always be kept at the forefront.

For e-commerce companies, customer segmentation is more than just a marketing tool. This process is a critical factor in making efficient decisions in customer relationship management, sales strategies, inventory management, and other business processes. Effective use of data mining techniques and machine learning methods will help companies make more accurate and effective segmentations. Adopting customer-focused strategies will enable companies to gain a stronger position in the market.

The main findings reveal that customer segmentation is not limited to demographic data only, but also should be analyzed in depth with behavioral and psychographic data. It is also emphasized that the effectiveness of personalized marketing strategies and the importance given to data security are critical to the success of e-commerce companies. The study shows that data mining techniques have the potential to increase customer satisfaction and loyalty in the e-commerce sector.

Customer Segmentation Variety: The findings revealed that customer segmentation is divided into quite different groups. Customers are divided into distinctly different segments based on characteristics such as demographic factors (age, gender, income level, etc.) and shopping habits. Young users shop frequently with low budgets, while middle-aged users prefer to shop with higher budgets. Impact of Personalized Marketing: It has been observed that personalized marketing strategies implemented after customer segmentation significantly increase sales. The Effect of Demographic Characteristics on Purchasing Behavior: Demographic factors (age, gender, income) have been found to have a significant effect on shopping behavior. For example, customers in the high-income group generally prefer more expensive products, while customers in the low-income group make price-focused choices. This shows that marketing strategies should be tailored to segments.

The Effectiveness of Data Mining Methods: It has been observed that the data mining and machine learning techniques used (cluster analysis, predictive modeling, etc.) increase the accuracy of customer segmentation and enable more accurate marketing decisions. These techniques have enabled companies to predict customer trends in advance and develop strategies accordingly.The Increasing Role of Mobile Shopping: The rate of shopping done via mobile devices has increased significantly. Customers' demand for mobile-friendly websites and applications is of great importance in terms of user experience. Mobile shopping is especially preferred by young users.

B. Limitations of the Study

Although this study has presented important findings by focusing on the analysis of e-commerce data with customer segmentation and data mining techniques, some limitations have emerged during the research process. First, the data set used consists of both publicly available sources and survey results obtained from a limited number of participants. In particular, the responses of a total of 78 participants were evaluated in the survey section. Although this number is considered sufficient for basic analyses, not working with a larger sample size limits the generalizability of the findings. The fact that most of the participants belong to a certain age range and region makes it difficult for the results obtained to fully represent the entire e-commerce customer base. This situation limits the ability to evaluate the impact of demographic differences on customer behavior from a broader perspective.

On the other hand, a significant portion of the data used in the study is based on the participants' own statements. This may also bring about subjective elements such as social desirability bias, misunderstanding or exaggeration. Since the responses obtained through the survey are based on the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec513

individuals' own perceptions, they may not always fully reflect real behaviors. In addition, there are uncertainties regarding the currency of some open datasets used, and it may not be clear whether these datasets reflect current consumer trends on e-commerce platforms. This affects the validity of the customer segments analyzed over time.

Another limitation of the study is related to the scope and technical capacity of the analysis tools used. In this study, the Orange data mining tool, which is particularly notable for its user-friendly interface, was preferred. Although Orange provides very efficient results in terms of basic cluster analysis and visualization, it offers limited opportunities in terms of applying advanced machine learning techniques or customizing models. More advanced algorithms that can be done in environments such as Python or R, which are coding-based platforms, were not included in the scope of this study. This limited the depth and diversity of the segmentation models obtained. In addition, the use of only K-means and similar basic algorithms in the study created potential limitations in terms of discovering more complex patterns.

A longterm observation technique that would accurately reflect the evolution of consumer behaviorover time is not included in this study, which is predicated on the examination of data from a certain time.

Customer preferences in the e-commerce sector are affected by many factors such as seasonal changes, campaigns or external economic factors. Therefore, follow-up studies conducted at different time intervals can reveal changes in customer behavior more accurately. Changing shopping habits and technological development over time require such studies to be constantly updated. Finally, the linguistic and cultural context of the study is also a limitation. The survey form and the general structure of the study were prepared in Turkish and targeted users in Turkey. In this context, the results obtained may not exactly coincide with user profiles in different cultural structures or countries. User attitudes, especially on data security, personal privacy and online shopping, may vary culturally. Therefore, the findings obtained from the study should be evaluated only in the context of Turkey. In conclusion, although this study provides important findings that will contribute to literature, it should be evaluated within the framework of the limitations mentioned above. These limitations will guide future research and form the basis for studies to be conducted with larger samples, in different geographies and with advanced techniques.

C. Recommendations for Future Research

The findings of this study demonstrate how data mining techniques can enhance the effectiveness and interpretability of customer segmentation in the e-commerce sector. However, considering the limitations of the study, it is possible to make several suggestions for future studies. First, customer profiles can be revealed more clearly by increasing the sample size and expanding it to include more diverse demographic groups.

In this direction, studies conducted with more balanced samples representing different age groups, income levels, education levels and geographical regions will increase the accuracy and generalizability of segmentation analyses. In addition, it is recommended that studies be conducted using more advanced machine learning methods instead of the basic clustering algorithms used in this study. With algorithms such as artificial neural networks, decision trees, and support vector machines, it will be possible to discover more complex patterns underlying customer behavior. The use of such models in application areas such as customer lifecycle value estimation, abandonment risk analysis, and personalized marketing strategy development may have strategic value for businesses.

Future studies can benefit significantly from the integration of diverse data sources. For example, combining quantitative data (e.g., purchase history, click behavior) with qualitative data (e.g., customer comments, social media interactions, sentiment analysis results) may offer a more multidimensional understanding of customer profiles. In this context, the integration of natural language processing (NLP) techniques and text mining methods can enable more in-depth and comprehensive analyses. At the same time, it will be possible to obtain more dynamic and up-to-date information about digital consumer behavior with the data obtained from social media platforms. In addition, it is recommended that research should not be limited to the consumer perspective only but also be supported by analyses focusing on companies' data use and segmentation strategies. Thus, not only customer behaviors, but also how companies use this data, which tools they prefer, and how segmentation studies are reflected in marketing strategies can be included in the research agenda.

Such holistic approaches will guide not only academic studies but also industry professionals.

Finally, it would be greatly beneficial to conduct comparative international studies examining the impact of cultural factors on customer segmentation. The question of whether the findings obtained in Turkey are similar or different in different countries will reveal the extent to which consumer behavior is affected by the cultural context. Such comparative studies can play an important role in the development of segmentation strategies and global marketing plans of multinational companies.

This research highlights the potential of data mining and machine learning in customer segmentation within the e-commerce sector. By extracting meaningful patterns from customer data, businesses can implement more effective marketing and personalization strategies. The study contributes to the growing field of customer analytics and offers practical implications for data-driven decision-making in competitive digital marketplaces.

As a result, customer segmentation studies are becoming increasingly sophisticated and multidimensional, driven by technological advancements and progress in data science. This trend highlights a growing need for comprehensive, interdisciplinary, and data-driven research. Guiding future studies in this direction will not only enrich academic literature but also enhance data-informed decision-making processes within the e-commerce sector.

REFERENCES

[1]. Aggarwal, D. (2023). Exploring the Role of AI for Online Grocery Shopping through Enhancing Personalized Recommendations and Customer Segmentation. Technoarete Transactions On Advances In Computer Applications (TTACA), 2(3).

[2]. Alzahrani, R., Habib, M., & Khan, M. A. (2022). Student Satisfaction Clustering using K-Means with Orange Data Mining Tool. International Journal of Advanced Computer Science and Applications, 13(2), 34-40.

[3]. Analytics Vidhya (2025). *What is k-means clustering?* https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/ *Accessed on 02.03.2025*)

[4]. Cooil, B., Aksoy, L., & Keiningham, T. L. (2008). Approaches to Customer Segmentation. Journal of Relationship Marketing, 6(3–4), 9–39. https://doi.org/10.1300/J366v06n03_02

[5]. Daruvuri, S., Raj, A., & Tripathi, M. (2025). Improving customer segmentation through PCA-enhanced K-Means clustering in online retail datasets. *Journal of Big Data*, 12(1), 48. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-025-01111-y

[6]. García-Murillo, M., & Annabi, H. (2002). Customer Knowledge Management. Journal of the Operational Research Society, 53(8), 875–884.

[7]. Google Developers. (2024). *What is clustering?* Accessed on 19.03.2025, https://developers.google.com/machine-learning/clustering/overview

[8]. IBM, (2025, March 02) *K-means clustering*. https://www.ibm.com/think/topics/k-means-clustering#:~:text=K-means%20is%20an%20iterative,the%20characteristics%20of%20the%20data

[9]. John, I., Shobayo, O., & Ogunleye, O. (2024). *Clustering customer segments in UK retail using RFM and machine learning algorithms*. ArXiv preprint. https://arxiv.org/abs/2402.04103

[10]. Kashwan, K. R. & C M, Velu. (2013). Customer Segmentation Using Clustering and Data Mining Techniques. International Journal of Computer Theory and Engineering. 5. 856-861. 10.7763/IJCTE.2013.V5.811.

[11]. Kirubakaran, Vignesh Saravanan & G, Sakthi. (2025). Customer Segmentation using Clustering Techniques: Data-Driven Approach to Enhance Marketing Strategy. 10.1109/ICSCNA63714.2024.10864053.

[12]. Kuhn, Max & Johnson, Kjell. (2013). Applied Predictive Modeling. 10.1007/978-1-4614-6849-3.

[13]. Lau, M. M., Cheung, R., Lam, A. Y., & Chu, Y. T. (2019). Segmenting online consumers by motivations for using e-commerce platforms. Journal of Retailing and Consumer Services, 47, 70-78.

[14]. Nimbalkar, A. A., & Berad, A. T. (2021). The increasing importance of AI applications in e-commerce. Vidyabharati International Interdisciplinary Research Journal, 13(1), 388-391.

[15]. Nugroho, M. A., Darmawan, M. A., & Yudhistira, A. (2024). Customer segmentation using optimized K-Means clustering in retail data. *Journal of Intelligent Data Science and Systems*, 4(2), 112–120. https://www.idss.iocspublisher.org/index.php/jidss/article/view/236

[16]. Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. IEEE Data Eng. Bull., 23(4), 3-13.

[17]. Rimakka, H., Öztürk, A., & Yilmaz, S. (2023). User segmentation based on purchasing habits and preferences on the Amazon platform using K-Means clustering. *International Journal of Data Science Research*, 6(3), 201–210. https://www.researchgate.net/publication/377972483

[18]. Semoglou, V., Papakostas, G. A., & Likas, A. (2025). K-Sil: A silhouette-based improvement of K-Means clustering for better customer segmentation. *Expert Systems with Applications*, 233, 120900. https://doi.org/10.1016/j.eswa.2025.120900

[19]. Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration of K-Means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering*, *336*(1), 012017. https://doi.org/10.1088/1757-899X/336/1/012017

[20]. Tuma, M. N., Decker, R., & Scholz, S. W. (2011). A survey-based customer segmentation using clustering of latent class probabilities. European Journal of Operational Research, 213(3), 564-572.

[21]. Wang, T. (2025). Reinforcement learning-based hybrid clustering for customer behavior analysis. *Applied Soft Computing*, 145, 110401. https://doi.org/10.1016/j.asoc.2025.110401

[22]. Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1, pp. 29-39).

[23]. Wu, R. S., & Chou, P. H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. Electronic Commerce Research and Applications, 10(3), 331-341. https://doi.org/10.1016/j.elerap.2010.11.002

[24]. Zhou, Y., Cheng, L., & Xu, H. (2020). Personalized Recommendation System Based on Customer Segmentation Using K-means Clustering and PCA. Procedia Computer Science, 174, 838-843