Customer Behavior: Prediction and Analysis Using Unsupervised Machine Learning Techniques

Abstract- This research investigates consumer behavior using unsupervised machine learning methods applied to a retail purchase dataset. The study aims to identify actionable trends for improved decision-making, categorize customers into distinct segments, and understand underlying purchase patterns. Employing K-means and hierarchical clustering techniques, along with dimensionality reduction using the Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE), the research develops a framework for analyzing complex, high-dimensional customer data. The findings identify different consumer groups with distinctive buying patterns which offer insightful information for focused advertising, customized client interactions, and efficient use of resources. This data-driven strategy demonstrates how unsupervised machine learning may improve company plans and lead to long-term success in the retail industry.

Index Terms- Customer Behavior, Unsupervised Machine Learning, Clustering, K-means Clustering, Hierarchical Clustering, Dimensionality Reduction, Principal Component Analysis (PCA), t-SNE, Retail Data Analysis, Customer Segmentation, Market Segmentation

I. INTRODUCTION

Understanding Customer Behavior is what forms a business and accelerates its growth. The customers' experience, interaction, purchases, and preferences is what provide insights of the customer needs, wants or the value they assign to products and services delivered by the company. This information plays a very important role in business expansion, increasing customer satisfaction, and managing resources effectively. Studying customer behavior is crucial for stabilizing retail operations such as streamlining supply chains, reducing wastage, and improving stocking strategies. However, significant challenges lie within the unstructured nature of the customer data, making it difficult to derive valuable details through conventional methods.

The traditional methods to analyze unstructured customer data fall short of capturing the dynamic relationships among the variables which becomes more complex when the data is generated through real-time customer interactions, this is where machine learning plays an important role.

Machine Learning techniques especially unsupervised learning, emerge as an invaluable tool in handling data with such complexities. Key insights provided by customer segmentation can help businesses with enhanced customer support and implement strategies to reach the target. These techniques help businesses to analyze vast, unstructured datasets and extract meaningful patterns, hidden relationships, and underlying trends which would help the management with more informed decision-making. This research aims to uncover the complex structures within the retail dataset, offering a deeper understanding of customer segments, marketing effectiveness, demographics, and purchasing behaviors. By integrating clustering and dimensionality reduction techniques, we seek to develop a framework that would help decode the customer data. This research would help transform businesses by looking into the unmet customer needs that can guide the development of new products or services, recognize customers' willingness to pay across different segments, and help ensure both customer satisfaction along with long-term profitability for businesses.

II. OBJECTIVE

This term paper investigates consumer behavior using unsupervised machine learning methods applied to a retail purchase dataset. The primary goals are to identify trends for improved decision-making, categorize clients, and understand purchase patterns. Clustering has been employed to put customers into segments and visualize high-dimensional data with increased precision. We chose these methods because they can analyze data without needing pre-set categories. They reveal hidden patterns and the diversity in how customers shop which helps companies create better marketing, more customized experiences, and stronger customer relationships. By bridging analytical methods with practical insights, the paper stresses the value of data-driven approaches in understanding and predicting customer behavior.

III. LITERATURE REVIEW

The ever-increasing amount, variety, speed, and complexity of data produced in today's digital ecosystem is known as big data (O. O. Adebola, 2019) . The widespread availability of digital data and the internet has enabled businesses to use big data to better understand the interests and behavior of their customers (Kalaiah, 2023) . The origins of artificial intelligence (AI) as a formal field of inquiry can indeed be traced back to Alan Turing's seminal 1950 paper titled *"Computing Machinery and Intelligence,"* published in *Mind*. (Kuiper Erik, 2019) Turing's work laid the philosophical and theoretical foundation for AI by exploring the possibilities of machine learning and intelligence. However, the possibility of new technology altering the nature of work seems genuine, even though the existential threat posed by robotics and artificial intelligence is still firmly restricted to science fiction. The Bank of England estimates that 15 million jobs will be lost by 2035, whereas (Balaram, 2018) PwC predicts that 7 million jobs will be lost by 2040. Regardless of the prognosis you choose to accept, this image is concerning.

(Shelley M. Cazares, 2020) You do not have an answer key when learning without supervisio n.Rather, you enter the data into the system, and it will determine how best to arrange it—how to rack and stack it so that human users like you or me can better interpret it. (Albert Orriols-Puig, 2009) To extract important information that the traditional analysis methodology was unable to identify, several writers have suggested using supervised machine learning techniques, which are informed with minimal prior knowledge about the problem. Following these efforts, one intriguing solution to these issues seems to be the employment of unsupervised learning techniques, which are ignorant of the problem structure and allow the machine to extract intriguing, practical, and undiscovered market knowledge. According to Larose (2014), unsupervised machine learning (UML) does not specify a goal variable and instead depends only on input data. Conversely, Supervised Machine Learning (SML) algorithms are given a specific goal (such as a target variable) to group data (Larose, 2014; Prasad, 2016; Walter & Bekker, 2017). This article focuses on UML for data-driven customer segmentation.

Understanding consumer behavior is essential to modern corporate strategy because it enables businesses to forecast demand, enhance customer satisfaction, and optimize their product offerings. (Pejman Ebrahimi, 2022) The unified theory of acceptance and use of technology (UTAUT), diffusion of adoption (DOI), theory of planned behavior (TPB), theory of reasoned action (TRA), and technology acceptance model (TAM) are some of the most popular theories for determining consumer purchasing behavior. (Mushtaq Ahmad Shah, 2024) Data analysis and real-world case studies provide insights into the predictive modeling of consumer behavior across a range of businesses. Social networks give marketers a lot of opportunities to interact with consumers and enable interactive user engagement. (Pejman Ebrahimi, 2022) In many developed and developing countries, people utilize social networking sites to buy items. Furthermore, customers' inclination to purchase goods in markets has been significantly impacted by COVID-19.

Consumer behavior encompasses a variety of behaviors, including brand involvement, preferences, buying patterns, and reactions to marketing initiatives. Businesses can obtain important insights by analyzing these behaviors, which include grouping customers according to shared traits like purchasing patterns or preferences, forecasting future customer requirements or market changes, identifying at-risk customers to put retention strategies in place, and creating customized marketing campaigns for particular audience segments. (OKONKWO, 2022) Businesses that use data-driven tactics and continuously improve their predictive models may be better equipped to foresee and address customer attrition, fostering enduring relationships with clients and long-term business success. Traditional methods of analysis are no longer adequate due to the increasing volume and complexity of client data from digital platforms. (Dung, 2021) Despite the potential of machine learning to address issues related to predicting customer behavior, there is a dearth of studies in this area. In addition, preprocessing data is necessary before using a machine learning model. Assumptions on feature independence, data distribution, and the relationship between predictor and result are important to machine learning models. (OKONKWO, 2022) Deviations from these presumptions could result in poor model performance or erroneous projections. Also, not every dataset can benefit from the same approach. Preprocessing techniques and machine learning models must therefore be studied in order to enhance the outcomes for a particular dataset. Furthermore, marketing departments frequently don't have a basic understanding of data-driven segmentation techniques. To create precise behavioral profiles or user segments, organizations frequently struggle to extract knowledge from data and choose the best machine learning algorithms. Without depending on predetermined labels or classifications, unsupervised machine learning (ML) is a potent solution that helps companies find patterns and relationships in big datasets. (Ali Rachini, 2024) Former segmentation techniques frequently fail, leading to less ideal marketing tactics and decreased customer satisfaction. By strengthening application security for sensitive data, the research broadens its influence, leading to a framework that uses data exploration, clustering, and dimensionality reduction to advance the classification of customer behavior through unsupervised learning. (Ali Rachini, 2024) Evaluations show that the following algorithms have exceptional accuracy levels: ANN (99.4%), CNN (99.3%), SVM (99%), Random Forest (99%), and Logistic Regression (95%). (Mushtaq Ahmad Shah, 2024) Methods like supervised random forests and unsupervised K-means clustering, with an 84% forecast accuracy, demonstrate how well these strategies work to divide the client base according to their spending patterns and income.

(Chen, 2009) The introduction of clustering improved the top decile lift when comparing the hybrid models to the benchmark scenario without clustering. One technique for grouping data, hierarchically, is agglomerative clustering. (Gopal & Jacob, 2022) Cases must be combined until the desired number of clusters is reached. The Elbow Method can be incorporated to calculate the number of clusters that must be generated. It is further demonstrated that the hybridization technique uses cluster labels as input for the decision trees. The finest hybrid models yield intriguing characteristics and guidelines that might assist marketing professionals in identifying churners from the data. Most organizations have recognized the importance of customer relationship management and the use of computational capabilities to achieve a competitive market advantage (Anant Katyayan, 2022).

Data mining techniques for customer segmentation could assist businesses in customer-focused marketing and developing unique strategies for a wide range of clients. (Shen, 2021) Numerous studies have been conducted on the RFM (Recency, Frequency, Monetary Value) model and clustering algorithms for consumer segmentation; (J. Divya Udayan, 2024) which gives companies a competitive advantage rewarding a better understanding of consumer behavior and guides the best possible growth paths. However, little research has been done on the relationships between customer groups and the goods they have bought. To interpret the results, feature dimensions must be decreased because humans cannot perceive space in more than three dimensions. (Shen, 2021) Pairwise similarities between the original input observations and the corresponding low-dimensional places in the pairwise similarities of the embedding are characterized by two distributions, with T-SNE minimizing the divergence between them. The use of ensemble learning (EL), (Hicham & Karim, 2022) an effective teaching method; has grown in popularity in recent years. EL uses a meta-classifier to combine findings from several techniques of classification. Additionally, it is better than other methods as it integrates the most precise elements of many machine learning techniques to create forecasts that are more accurate than those produced by any of the ensemble's algorithms. (Mahboob, 2015) Hartigan (1985) explains that there isn't a single optimal criterion for every issue.

IV. DATA AND METHODOLOGY

This study used a dataset gathered from retail sales and customer transaction records, consisting of thousand data points. It included variables like purchase frequency, transaction amounts, product categories, and customer demographics. The data was cleaned to ensure it was reliable and consistent, with a focus on addressing missing values and outliers. The dataset provides a structured view of customer behavior, facilitating analysis for clustering and dimensionality reduction techniques.

To help find patterns, structures, and correlations in datasets, **clustering techniques** are employed in data analysis to group related data points into clusters. A cluster is a group of observations that differ from those in other clusters yet are comparable to each other within the same cluster. The dataset was divided into customer groups according to purchasing trends for **K-Means Clustering. The Elbow Method** was used to determine the ideal number of clusters and the analysis of cluster centroids revealed the important behavioral tendencies. K-means clustering is a straightforward and popular method for dividing a data collection into K different clusters. K-means clustering is carried out by determining the desired clusters. After K has been determined, each observation must be assigned to precisely one of the K clusters:

Step 1: Give each observation a random number between 1 and K. These act as the observations' first cluster allocations.

Step 2: Continue iterating until cluster allocations are stable.

a) Determine the cluster centroid for each of the K clusters.

b) Assign each observation to the cluster with the closest centroid.

Hierarchical Clustering began with feature selection so that only the most relevant attributes were included in the analysis. After that, the features were scaled and standardized to get rid of any biases that may have arisen from differences in scale. A linkage matrix was then computed to determine the hierarchical relationships between data points. This process was visualized using a **dendrogram** which in turn gave insights into the clustering structure. Lastly, cluster labels obtained from the dendrogram were added to the dataset, to get a clear interpretation of the formed groups and enable further analysis. was performed to reduce the number of dimensions or features in a data set. The goal of dimensionality reduction is to decrease the data set's complexity by reducing the number of features while also keeping the most important properties of the original data.

Two methods were used in order to supplement clustering. **Principal Component Analysis (PCA)**: PCA was utilized to minimize the dataset's dimensionality while maintaining the highest possible variance. This method projected the dataset into a lower-dimensional space, which aided in grouping and enhanced interpretability. The first few principal components are often employed while the remaining principal components are omitted in principal component analysis (PCA), a technique for calculating the principal components of the data. It is frequently utilized.

For **dimensionality reduction**, each data point is projected onto the first few principal components in order to produce lower-dimensional data while retaining the greatest amount of variety in the data**. t-SNE:** This technique was used to display the high-dimensional customer data in two dimensions, giving cluster distributions and overlaps a more lucid appearance. **T-Distributed Stochastic Neighbor Embedding (T-SNE)** works especially well for visualizing high-dimensional datasets. "The SNE transaction formation is a bijection between the mapping space (2-dimensional or 3-dimensional) and the original multidimensional feature space. A distribution that describes pairwise similarities of the original input observations and another that explains pairwise similarities of the corresponding low-dimensional points in the embedding are the two distributions that T-SNE minimize the divergence between.

V. DATA DESCRIPTION AND FEATURE ENGINEERING

The study has incorporated a retail transaction dataset consisting of 1000 unique transactions sourced from the Data science and ML repository. Table 1 Give a summary of variables in the dataset.

Table 1: Variables in the dataset

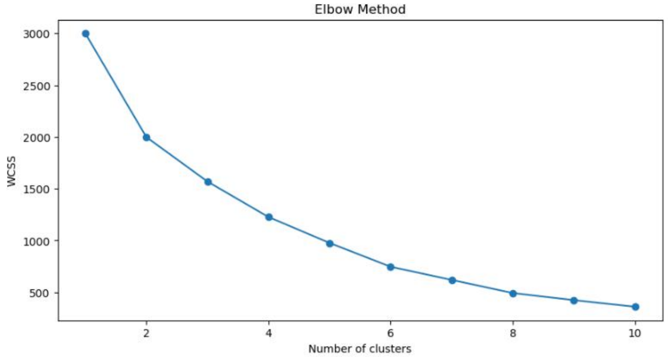
Variable Name	Description	Unique Values
Transaction ID	Serial Number of transactions during the period	1000
Date	The date when transactions were generated	354
Customer ID	A unique 6-digit number assigned to each transaction	1000
Gender	Gender of customer	2
Age	Age of customer	47
Product Category	Category of Product sold to different Customer ID's	3
Price	Product Price Per Unit	4
Quantity	Quantity of units sold per transaction.	5
Total Amount	Total spent money per transaction	18

The data initially does not have the column Total Amount. It was calculated as a product of Price and Quantity. No missing values, special values, or negative values were found in the data suggesting no transactions were canceled. Further, variables age, product category, and gender were identified as relevant and were taken into consideration for the remaining process. The raw data only consists of variables that describe historical transactions, these cannot be used as final inputs for clustering algorithms. Thus, variables were standardized in positive values between zero and one which gives consistency for clustering analysis.

CLUSTERING

The above-mentioned variables are the ones used for the clustering algorithm. In this section, our analysis, the K-means clustering technique has been used. Selecting the optimal number of clusters becomes an important part of the process. A lower number of clusters can lead to low precision and poor decision-making, whereas a higher number of clusters and lead to overfitting and complexity in interpretation. Good clustering can be classified as ones in which there is minimum within-cluster variation. The study has used the 'Elbow method' to identify the number of clusters by reducing the within-cluster sum of squares. The method assigns centroids to each point and then it iterates the process and finds the optimal centroid which reduces its within cluster sum of squares. Figure 1, displays within the cluster sum of squares for each k. The decision to select the optimal number of clusters is based on the slope of Figure 1.

Figure 1: The Elbow Method



The information in figure 1 suggests the number of clusters to be 3, suggesting it is the point to which the within-cluster sum of squares falls drastically and then it starts falling at a decreasing rate. Table 1 shows the number of Customer IDs in each cluster.

Table 1: Number of Customers in each cluster

Cluster	Number of customers
0	349
1	344
2	307

PRODUCT ANALYSIS

The data was uniformly distributed on gender accounting for 490 transactions being carried out by males, and the remaining transactions being carried out by female counterparts by female counterparts. The average spending of males was 455.4286, which of the female counterparts amounted to 456.54. Table 2 shows transactions made by customers based on product category.

Table: 2: Transactions Based on Product Category

Product category	Transactions (Male)	Transactions (Female)
Beauty	307	166
Clothing	94	174
Electronics	89	170
Total	490	510

Figure 2: Product Category by Gender

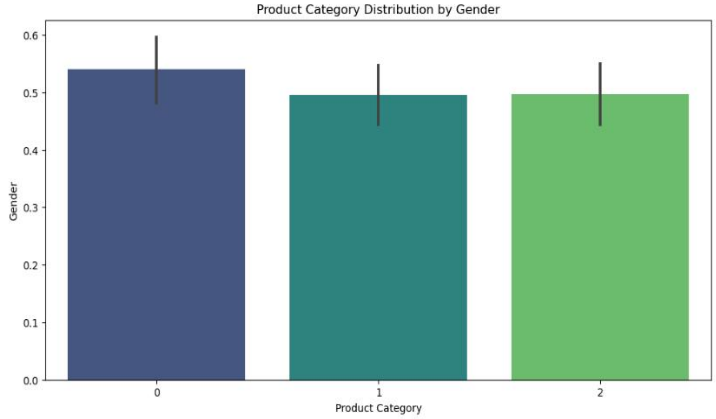
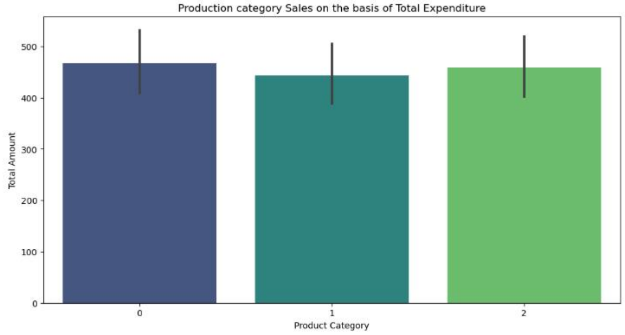


Figure 3: Distribution of Total spending on product category

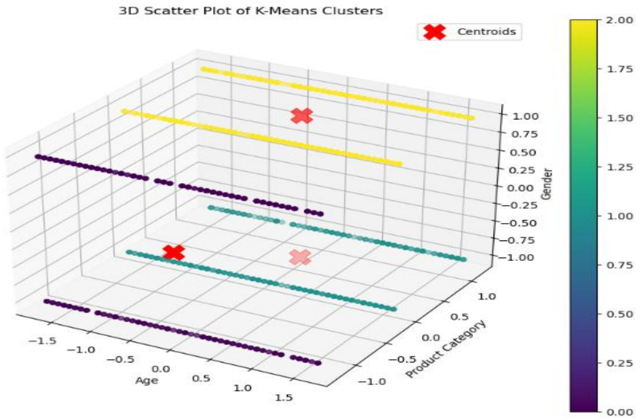


In Figure 2, Gender 0 is referred to as male, and gender 1 as female. Product category 0 is referred to as Beauty, 1 as Clothing, and 2 as Electronics. The uniform distribution gives us insight that Beauty Products are preferred more by Females whereas, Clothing and Electronics are preferred more by Males. Similarly, In Figure 3, it can be seen that Customer spending is uniform in Beauty and Electronic products, and clothing products though have a slightly lower average spending.

CLUSTER INTERPRETATION

As mentioned above the dataset has been scaled to achieve consistency. Figure 4 shows the graph of K-Means clusters. The data points are scaled from -1.5 to 1.5. For Centroid 0, we can see the features of Cluster 0 as having

Figure 4: 3D scatter plot of K-means clusters



Uniform behaviour in choosing product category the product category beauty, a slight oreintation of male gender is shown in this gender rather than females, age group of this clutser consists more of young customers. The cluster 2, having centroid 1 as its feature centroid, consist more of average age group, customers of this cluster prefer more of clothing and electronic products. Average gender of this cluser if of female counterparts. The interpretation of cluster 3, having its feature centroid as 2, shows inclination of customers oriented towards colthing and electronic products for an average of 43.76 years, the average gender of this cluster is Female.

Table 3: Centroid attributes of clusters

Centroid	Age	Product Category	Gender	Cluster
0	30.37	Beauty	0.54	1
1	41.84	Clothing and Electronics	0:1	2
2	43.76	Clothing and Electronics	1:0	3

DIMENSIONALITY REDUCTION

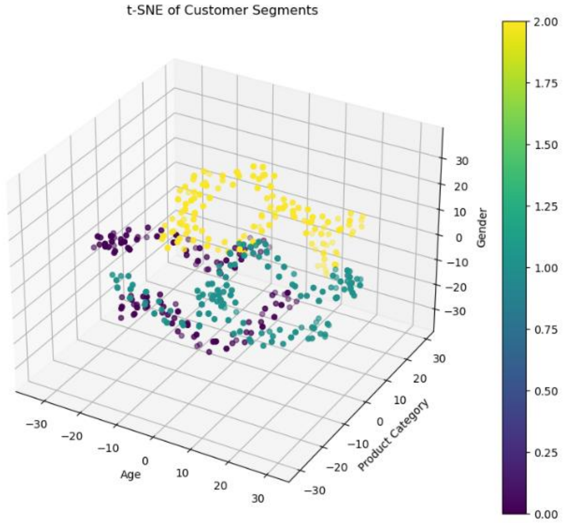
The applicability of dimensionality reduction has been assessed through the application of evaluation parameters. Figure- 5, shows the principal components, and Figure 6, shows the graphical representation of T- Stochastic Embedded Neighbors. Silhouette Score, Calinski-Harabasz Index and Davies-Bouldin Index. Silhouette scores measure the similarity of a data point with its own cluster in comparison to other clusters. It measures cohesion within the cluster and separation from other clusters. Calinski-Harabasz Index evaluates cluster quality by taking the ratios between cluster dispersion and within cluster dispersion. It assesses separation and compactness among clusters. Davies-Bouldin Index measures the similarity ratio of a cluster to its most similar cluster. A Silhouette score close to 1 is considered as ideal, Davies-Mouldin Index reflects before performance for values close to 0. In the case of the Calinski-Harbasz Index higher score reflects a better. On a comparative evaluation of the performance matrix in Table 4, it can be analyzed that K-Means provides optimal values of all three evaluation methods representing compact and well-separated cluster quality. The application of dimensionality reduction methods reduces cluster quality significantly suggesting the inapplicability of Dimensionality reduction techniques for the dataset.

Table 4: Evaluation scores of Clustering

Evaluation Matrices	K-Means	PCA	T-SNE
Silhouette Score	0.58	0.34	0.25
Calinski-Harabasz Index	3901.57	454.03	305.77
Davies-Bouldin Index	0.52	1.19	1.48

The results from the evaluation matrices further emphasize the superiority of the K-Means algorithm over dimensionality reduction techniques like PCA and T-SNE for this dataset. K-Means achieved the highest Silhouette Score, indicating strong cohesion within clusters and clear separation between them. Additionally, the Calinski-Harabasz Index for K-Means was much higher, reinforcing the idea that it produces well-formed and tightly grouped clusters. On the other hand, both PCA and T-SNE had lower scores across all metrics, indicating that using dimensionality reduction techniques led to a decrease in the quality of the clusters. These findings suggest that while dimensionality reduction techniques can be useful in certain scenarios, they may not be suitable for maintaining cluster integrity in this particular case.

Figure 5: t-SNE of customer segments



HIEREARCHIAL CLUSTERING

Hierarchical clustering can be classified as one of the Unsupervised Machine Learning technique for grouping the pair of objects in the clusters. It configures database in dendrogram. Dendrogram is a graphical representation of the merge done during clustering to signify the number of clusters and the steps taken to attain them. This method is particularly useful and helpful in determining the hierarchy and relationships within a dataset when the number of clusters too is not given. Types of Hierarchical Clustering Agglomerative (Bottom-Up Method): In this method each data point is set as its own cluster initially. This progressively combines clusters based on certain measures such as distance or how similar the clusters are until one cluster containing all data points is formed. This method is the most widely used because it is easier to implement. Divisive (Top-Down Method): This method begins with all data housed in one cluster. Clusters are unsystematically split in smaller ones using dissimilarity. This method is not used commonly due to being very computationally intensive.

Linkage Criteria:

Single Linkage: the distance between the outer most points of the two clusters, which is the minimum distance in this context.

Complete Linkage: considers the maximum distance of the outermost points of the two clusters.

Average Linkage: distance, rather the average distance between all the points in each of the two clusters rather than just two pointed outer limits.

Ward's Method: This is the most commonly used method that aims to minimize the variance in a cluster.

Since classification of customers and distinct purchase patterns were important, hierarchical clustering was used to examine the customer retail sales data. The data was the first organized in a way that clusters could be formed around every customer purchasing. To determine the degree of similarity, or dissimilarity between customers, a distance matrix was defined based on the Euclidean distance, although other distance measures may be deployed. This distance relationship matrix is the starting point for hierarchical clustering, a process in which data that are more similar to one another are placed in a cluster in sequence until there are no more clusters.

In our case, Ward's linkage method which aims at reducing the within-cluster variance during the cluster merging stage was used. This approach guarantees compact, well-separated clusters. The resulting dendrogram showed a series of steps taken during the processes of cluster formation, by illustrating how various clusters were formed through an iterative merging process. A single customer was positioned at each edge of the dendrogram, with the branches connecting customers or clusters based on how alike they were. As we ascended the dendrogram, the top clusters became flatter and less precise, indicating the sequential development of the hierarchy. For the proper identification of the best number of clusters, the largest vertical lines in the dendrogram were studied, due to the fact that such lines signify an important difference in the degree of similarity of the united groups. Thus, by "cutting" the dendrogram at an appropriate height, we were able to translate the data into a finite number of clusters. For example, in our analysis, cutting the dendrogram at a threshold with a certain level led to the emergence of three distinct clusters. There is a consistent and uniform customer behavior in each cluster. These clusters were interpreted as follows:

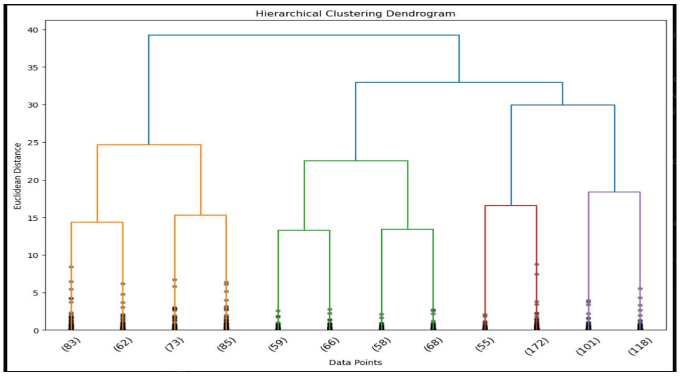
Cluster 1: These are customers who made lower value but higher number of transactions. Such customers could have been frequent buyers of the firm but such firm has had modest earnings from them.

Cluster 2: They are categorized as low number of transactions but high value transactions. Such a group is an attractive target market for marketing campaigns since such customers are able to generate much revenue for the firm even if they are few.

Cluster 3: These are customers who buy in average transaction values at infrequent periods which could be occasional buyers of the firm.

Such may be the characteristics of customers who have a generalist or less precise purchasing patterns. The clustering results offered useful interpretation on the customer segmentation. People who shop frequently may be offered loyalty programs; high value customers may be persuaded to do more frequent shopping with special offers, while people who sometimes shop would be targeted with mass advertising campaigns to make them come back and shop again.

Fig6: Hierearchial clustering Dendrogram



The dendrogram offers a thorough picture of how hierarchical clustering works on the dataset. It shows how data points form groups based on how alike they are. This process uses the Ward linkage method and measures distance with the Euclidean metric. On the x-axis, we can see clusters of data points. Numbers like 83, 62, and 73 tell us how many points are in the smallest clusters. At the bottom of the tree, each data point starts alone. As we look up, points that are alike join into bigger groups. The y-axis shows the Euclidean distance where these joins happen. Lower distances mean the clusters are more alike, while higher distances show they're less similar. For example, smaller clusters with 85 and 73 points join at lower distances. This suggests their features (like Age, Product Category, Gender, and Total Amount) are more alike. On the other hand, large blue and purple groups join at higher distances. This means their traits are quite different from each other.

The dendrogram's merging process shows how the dataset's structure unfolds in a hierarchy. When the Euclidean distance reaches about 20, the dendrogram points to five separate clusters: orange, green, red, blue, and purple groups. Clusters near the x-axis at the bottom have more in common. The orange and green clusters, for example contain customers with similar spending habits or demographic traits. In contrast bigger clusters like the red group, with 172 data points, stand for a more varied customer segment. The blue and purple clusters join at a higher Euclidean distance, which means their features vary more and their customer profiles are less alike. For businesses, the dendrogram helps to group customers by spotting those with similar buying behaviors, demographics, or product. Smaller clusters, like the orange group, might represent specific customer groups with unique needs. Larger more diverse clusters such as the blue and purple groups, stand for wider market segments.

VI. CONCLUSION

This study shows how unsupervised machine learning effectively uncovers valuable, previously unknown information in retail customer transaction data. We used K-means clustering to divide our customers into three groups with different ages, genders, and favorite products. This approach revealed purchasing trends beyond simple demographics, giving us insights into traditional methods. The Elbow Method helped us find the best number of groups, showing how important it is to choose the right settings in unsupervised machine learning. These customer groups are perfect for creating focused marketing and personalized services for the companies. A detailed understanding of cluster analysis improves marketing by enabling personalized messages, product recommendations, and promotions that are tailored specifically to target each group's preferences. The study rigorously evaluated the impact of dimensionality reduction techniques (PCA and t-SNE) on the quality of the resulting clusters. The findings indicate that, for this particular dataset, dimensionality reduction techniques did not improve the clustering results, as measured by the Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index. This highlights the crucial point that the application of dimensionality reduction is not universally beneficial and should be carefully considered based on the specific dataset and the goals of the analysis. A blanket application of dimensionality reduction can degrade the quality of the results, as demonstrated here. In conclusion, the research highlights the importance of using customer data to understand their behaviors and identify patterns. By identifying the customer groups and their shopping patterns, businesses can gain a competitive advantage. The insights gained from this study can guide decisions on inventory management, product development, pricing strategies, and overall business planning, ultimately driving better performance. It also shows how unsupervised machine-learning techniques can make a real difference in the retail industry. By using clustering techniques and carefully analyzing the results, businesses can understand their customers in a better way and run more effective marketing campaigns while building stronger relationships and even boosting their profitability. The findings emphasize the value of a data-driven approach and demonstrate how advanced analytical techniques can uncover meaningful insights hidden within complex and high-dimensional datasets.

VII. REFERENCES

[1] Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In M. W. Berry, A. Mohamed, & B. W. Yap (Eds.), *Supervised and Unsupervised Learning for Data Science* (pp. 3–21). Springer International Publishing. https://doi.org/10.1007/978-3-030-22475-2_1

[2] Alzahrani, L. A. (2021, July 15). *Customer Segmentation: Unsupervised Machine Learning Algorithms In Python*. Medium. https://towardsdatascience.com/customer-segmentation-unsupervised-machine-learning-algorithms-in-python-3ae4d6cfd41d

[3] Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation using Clustering Ensemble and Spectral Clustering—ProQuest. (n.d.). Retrieved November 17, 2024, from https://www.proquest.com/openview/bf1aad25aa493cea5049413829c034cc/1?pq-origsite=gscholar&cbl=5444811

[4] Cazares, S. M., Parrish, E. M., Holzer, J. R., & Moeller, R. A. (2020). *Supervised versus Unsupervised Learning* (IDA Ideas (Podcast Transcript)— Weaponized Tweets:, pp. 3–4). Institute for Defense Analyses. https://www.jstor.org/stable/resrep36544.6

[5] Dellot, B., & Balaram, B. (2018). Machine Learning. *RSA Journal*, *164*(3 (5575)), 44–47.

[6] Ebrahimi, P., Basirat, M., Yousefi, A., Nekmahmud, M., Gholampour, A., & Fekete-Farkas, M. (2022). Social Networks Marketing and Consumer Purchase Behavior: The Combination of SEM and Unsupervised Machine Learning Approaches. *Big Data and Cognitive Computing*, *6*(2), Article 2. https://doi.org/10.3390/bdcc6020035

[7] E-commerce Customer Segmentation via Unsupervised Machine Learning | The 2nd International Conference on Computing and Data Science. (n.d.). Retrieved November 23, 2024, from https://dl.acm.org/doi/10.1145/3448734.3450775

[8] Finesso, R., Spessa, E., & Venditti, M. (2016). An Unsupervised Machine-Learning Technique for the Definition of a Rule-Based Control Strategy in a Complex HEV. *SAE International Journal of Alternative Powertrains*, *5*(2), 308–327.

[9] Gopal, A. C., & Jacob, L. (2022). Customer Behavior Analysis Using Unsupervised Clustering and Profiling: A Machine Learning Approach. *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2075–2078. https://doi.org/10.1109/ICACITE53722.2022.9823646

[10] Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn: Journal of Organizational Computing and Electronic Commerce: Vol 19, No 2. (n.d.). Retrieved November 17, 2024, from https://www.tandfonline.com/doi/abs/10.1080/10919390902821291

[11] Katyayan, A., Bokhare, A., Gupta, R., Kumari, S., & Pardeshi, T. (2022). Analysis of Unsupervised Machine Learning Techniques for Customer Segmentation. In J. I.-Z. Chen, H. Wang, K.-L. Du, & V. Suma (Eds.), *Machine Learning and Autonomous Systems* (pp. 483–498). Springer Nature. https://doi.org/10.1007/978-981-16-7996-4_35

[12] Khanum, M., Mahboob, T., Imtiaz, W., Abdul Ghafoor, H., & Sehar, R. (2015). A Survey on Unsupervised Machine Learning Algorithms for Automation, Classification and Maintenance. *International Journal of Computer Applications*, *119*(13), 34–39. https://doi.org/10.5120/21131-4058

[13] Krishan, R. (2023). Customer Behaviour Analysis Using Machine Learning Algorithms. In *Digital Transformation, Strategic Resilience, Cyber Security and Risk Management* (world; Vol. 111B, pp. 133–142). Emerald Publishing Limited. https://doi.org/10.1108/S1569-37592023000111B009

[14] Kuiper, E., Constantinides, E., Vries, S., Muster, R., & Metzner, F. (2022). A Framework of Unsupervised Machine Learning Algorithms for User Profiling A Framework of Unsupervised Machine Learning Algorithms for User Profiling.

[15] Li, J., Pan, S., Huang, L., & Zhu*, X. (2019). A Machine Learning Based Method for Customer Behavior Prediction. *Tehnički Vjesnik*, *26*(6), 1670–1676. https://doi.org/10.17559/TV-20190603165825

[16] Naeem, S., Ali, A., Anam, S., & Ahmed, M. (2023). An Unsupervised Machine Learning Algorithms: Comprehensive Review. *IJCDS Journal*, *13*, 911–921. https://doi.org/10.12785/ijcds/130172

[17] Okonkwo, R., & Onasanya, A. E. (2022). *Predictive Analytics for Customer Behavior*. https://www.academia.edu/114857299/Predictive_Analytics_for_Customer_Behavior

[18] Orriols-Puig, A., Casillas, J., & Martınez-Lopez, F. J. (n.d.). Unsupervised Learning of Fuzzy Association Rules for Consumer Behavior Modeling.

[19] *(PDF) Predicting Consumer Behaviour in Digital Market: A Machine Learning Approach*. (n.d.). Retrieved November 24, 2024, from https://www.researchgate.net/publication/335149938_Predicting_Consumer_Behaviour_in_Digital_Market_A_Machine_Learning_Approach

[20] (PDF) PREDICTIVE ANALYTICS FOR CUSTOMER BEHAVIOUR: DEVELOPING A PREDICTIVE MODEL THAT ANALYZES CUSTOMER DATA TO FORECAST FUTURE BUYING TRENDS AND PREFERENCES, ENABLING SMALL BUSINESSES TO TAILOR THEIR

MARKETING AND PRODUCT STRATEGIES EFFECTIVELY. (n.d.). Retrieved November 17, 2024, from https://www.researchgate.net/publication/378176015_PREDICTIVE_ANALYTICS_FOR_CUSTOMER_BEHAVIOUR_DEVELOPING_A_PREDICTIVE_MODEL_THAT_ANALYZES_CUSTOMER_DATA_TO_FORECAST_FUTURE_BUYING_TRENDS_AND_PREFERENCES_ENABLING_SMALL_BUSINESSES_TO_TAILOR_THEIR_MARKETI

[21] Quynh, T. D., & Dung, H. T. T. (n.d.). Prediction of Customer Behavior using Machine Learning: A Case Study .

[22] Rachini, A., Fares, C., Assaf, M. A., & Jaber, M. M. (2024). Revolutionizing Business with AI: Unlocking Customer Insights Through Unsupervised and Supervised Learning for Behavior Prediction. In X.-S. Yang, S. Sherratt, N. Dey, & A. Joshi (Eds.), *Proceedings of Ninth International Congress on Information and Communication Technology* (pp. 483–493). Springer Nature. https://doi.org/10.1007/978-981-97-3299-9_39

[23] Shah, M. A., & Kumar, P. (2024). Leveraging Machine Learning Techniques to Project Customer Behaviour Through Predictive Analysis and Ethical Marketing. In S. S. Dadwal, H. Jahankhani, & K. Revett (Eds.), *Market Grooming* (pp. 121–138). Emerald Publishing Limited. https://doi.org/10.1108/978-1-83549-001-320241006

[24] Udayan, J. D., Moneesh, N., Vemulapalli, N. S., Pruthvi, P., & Sakhamuri, R. (2025). Application of Unsupervised Learning in Detecting Behavioral Patterns in E-commerce Customers. In A. Kumar, V. K. Gunjan, S. Senatore, & Y.-C. Hu (Eds.), *Proceedings of the 5th International Conference on Data Science, Machine Learning and Applications; Volume 1* (pp. 1208–1217). Springer Nature. https://doi.org/10.1007/978-981-97-8031-0_127

[25] Upreti, G., & Natarajan, A. K. (2024). Leveraging Unsupervised Machine Learning to Optimize Customer Segmentation and Product Recommendations for Increased Retail Profits. In *Intersection of AI and Business Intelligence in Data-Driven Decision-Making* (pp. 257–282). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-5288-5.ch009.