Clustering Consumers: A Deep Dive into Customer Segmentation Strategies

Fahmid Bin Kibria[1*] and Moin Mostakim[1]

*Corresponding author(s). E-mail(s): fahmid.bin.kibria@g.bracu.ac.bd;
Contributing authors: mostakim@bracu.ac.bd;

Abstract

Customer segmentation plays a vital role in modern marketing strategies by enabling businesses to understand and target distinct groups of consumers based on their behaviors and preferences. This study explores an unsupervised learning approach to customer segmentation using clustering techniques such as K-Means algorithm along with dimensionality reduction techniques such as Principal Component Analysis (PCA) and feature transformation strategies. The objective is to discover meaningful customer profiles from unlabeled transactional and demographic data, helping organizations to personalize services and optimize marketing efforts. An in-depth analysis of the wholesale customers dataset is performed, which focuses on features such as "Channel" and "Region". The proposed method effectively uncovers hidden patterns within the customer base, offering insights that can support targeted campaigns and customer relationship management. Challenges such as choosing the optimal number of clusters, handling feature scaling, and assessing cluster validity in an unsupervised context are also addressed. Our results demonstrate the potential of neural-inspired clustering models in enhancing business intelligence and data-driven decision-making. The performance of the segmentation is evaluated by comparing it with traditional clustering methods using internal validation metrics such as the silhouette score.

Keywords: Customer Segmentation . Clustering . K-Means . PCA . Decision-making

1 Introduction

In the rapidly evolving landscape of data-driven business intelligence, customer segmentation has emerged as a crucial strategy for understanding and targeting diverse customer groups. It enables businesses to tailor marketing strategies, improve customer service, and enhance product recommendations by dividing the customer base into homogeneous segments based on common attributes such as purchasing behavior, demographics, and lifestyle preferences. However, as the volume and complexity of customer data continue to grow, traditional segmentation techniques fall short of delivering actionable insights.

To address this challenge, unsupervised machine learning techniques have gained prominence for their ability to uncover hidden structures within data without the need for predefined labels. These methods are particularly useful in domains where the nature of customer groups is unknown in advance. Among these, K-Means clustering is one of the most popular algorithms due to its simplicity, scalability, and effectiveness in partitioning data into distinct clusters. It operates by minimizing intra-cluster variance, offering an efficient way to group customers based on feature similarity. Despite its wide adoption, K-Means assumes spherical clusters and requires prior knowledge of the number of clusters, which can limit its performance in real-world, high-dimensional datasets.

To enhance the quality of clustering and overcome the limitations of high-dimensional data, dimensionality reduction and feature transformation techniques such as Principal Component Analysis (PCA) are frequently employed. PCA transforms the original features into a new set of orthogonal components that capture the most significant variance in the data, enabling more efficient and meaningful clustering. This preprocessing step is particularly effective in eliminating redundancy, reducing noise, and improving the performance of clustering algorithms like K-Means.

In addition to PCA, advanced feature transformation strategies including normalization, standardization, and encoding of categorical variables play a critical role in preparing the data for unsupervised learning. These transformations ensure that all features contribute equally to the distance metrics used in clustering, which is essential for producing balanced and interpretable clusters.

In this study, we propose a robust framework for customer segmentation by integrating feature engineering, dimensionality reduction using PCA, and K-Means clustering. We perform in-depth preprocessing of a real-world retail dataset, transforming and reducing the data into a form suitable for effective unsupervised learning. Our approach is designed to optimize the segmentation process, uncover meaningful patterns, and provide interpretable insights that can support business decision-making. The key contributions of this paper are:

- A systematic application of unsupervised learning techniques for customer segmentation using K-Means clustering.
- The use of Principal Component Analysis (PCA) for dimensionality reduction to enhance clustering quality and visualization.
- An evaluation of feature transformation strategies including normalization, scaling, and encoding to improve data quality.

- A comparative analysis of clustering performance using internal validation metrics such as the Silhouette Score.
- A discussion of challenges in unsupervised learning, such as determining the optimal number of clusters and interpreting unlabeled data.

The rest of the paper is organized as follows: Section 2 presents a review of related works in customer segmentation using unsupervised learning. Section 3 describes the dataset and preprocessing techniques. Section 4 outlines the proposed methodology. Section 5 presents experimental results and performance analysis. Section 6 brings the end to this paper along with prior discussion on limitations, challenges, and future directions.

2 Related Works

Customer segmentation has long been a core area of interest in data-driven marketing and customer relationship management. Traditional segmentation methods, such as demographic-based or rule-based classification, often fall short in capturing the complex, nonlinear relationships present in real-world consumer data. To overcome these limitations, unsupervised machine learning techniques have gained popularity due to their ability to uncover hidden patterns without the need for labeled data.

Among the most widely used clustering algorithms, K-Means has been extensively applied for customer segmentation tasks due to its simplicity and efficiency [1]. However, its reliance on spherical cluster assumptions and sensitivity to initial centroids limits its performance on high-dimensional or non-convex datasets. Hierarchical clustering and DBSCAN have also been explored for segmentation, offering alternative advantages such as noise handling and flexible cluster shapes [2].

To enhance segmentation quality, researchers have turned to Self-Organizing Maps (SOM), a type of unsupervised neural network that preserves topological properties of data while reducing dimensionality [3]. SOMs have been effectively applied to visualize and cluster customer profiles, especially when interpretability is crucial [4]. Studies have shown that SOM-based segmentation provides more intuitive cluster structures and improved stability compared to K-Means [5].

Recent works have also explored hybrid approaches, combining SOM with other methods such as KNN, Fuzzy C-Means, or PCA, to improve clustering precision and robustness [6]. These models leverage the strength of SOM in feature abstraction and the classification power of algorithms like KNN for refined segmentation. In parallel, deep learning-based approaches using autoencoders and deep clustering networks have emerged, providing competitive performance on large, complex datasets [7].

Sarvari et al. [8] employed hierarchical clustering and K-Means on retail customer data, emphasizing that proper feature selection significantly influences clustering performance. Zhang et al. [9] combined autoencoders with K-Means clustering for customer segmentation, achieving better performance on high-dimensional e-commerce datasets compared to standalone methods. Gouda et al. [10] implemented a hybrid SOM and K-Nearest Neighbors (KNN) approach to analyze customer buying behavior. The study highlighted how SOM can structure the data while KNN enhances classification post-segmentation, improving interpretability and downstream decision-making.

Despite these advancements, there remains a research gap in systematically comparing neural-based clustering approaches like SOM with classical techniques under practical marketing scenarios. Our work addresses this by building a SOM-based segmentation model and evaluating it against conventional clustering methods. We also explore the use of K-means along with PCA and feature transformation techniques to simulate post-segmentation classification, thereby enhancing both visualization and interoperability.

3 Dataset and Exploratory Data Analysis

3.1 Dataset Description

The dataset used in this study is the Wholesale Customers [11] Dataset, which is publicly available from the UCI Machine Learning Repository. It contains data on customer spending patterns across various product categories and is commonly used for market segmentation and clustering analysis. The dataset provides a realistic basis for unsupervised learning, as it includes no predefined labels or classes. Table 1 shows the features of the dataset and its descriptions.

Table 1: Dataset features

Feature	**Type**	**Description**
Channel	Categorical (1: Horeca, 2: Retail)	Type of customer: Horeca (Hotel/Restaurant/Café) or Retail.
Region	Categorical (1: Lisbon, 2: Oporto, 3: Other)	Geographical region of the customer.
Fresh	Continuous	Annual spending (in monetary units) on fresh products.
Milk	Continuous	Annual spending on milk products.
Grocery	Continuous	Annual spending on grocery items.
Frozen	Continuous	Annual spending on frozen products.
Detergents_Paper	Continuous	Annual spending on detergents and paper products.
Delicassen	Continuous	Annual spending on delicatessen items.

3.2 Exploratory Data Analysis (EDA)

For data exploration, I have selected 3 samples. For the first sample, as illustrated in Fig. 1, we could see that the purchase cost of “Fresh” and “Frozen” are way less than the sample mean. Also, the purchase cost of other categories is also less than the population mean. The establishment could be of a small cafe or hotel. For the second sample, as illustrated in Fig. 2, we could see that the purchase cost of “Fresh” is way low than the population mean. This establishment has a purchase cost approximately equal to population mean for “Milk”, “Grocery” and “Frozen”. This indicates that this could be a supermarket or restaurant. For the third sample, as illustrated in Fig. 3, we could see that the purchase cost of all the categories is almost equal to the population mean except for “Fresh” for which the purchase cost is very high. This

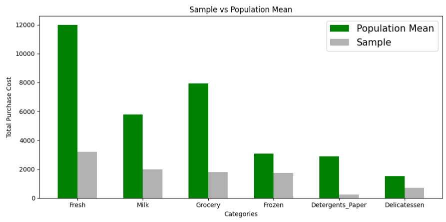

Fig. 1: Plot data of sample 1 w.r.t population mean

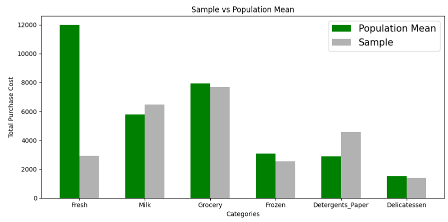

Fig. 2: Plot data of sample-2 w.r.t population mean

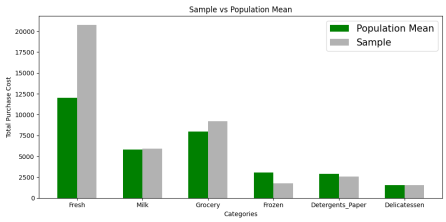

Fig. 3: Plot data of sample-3 w.r.t population mean

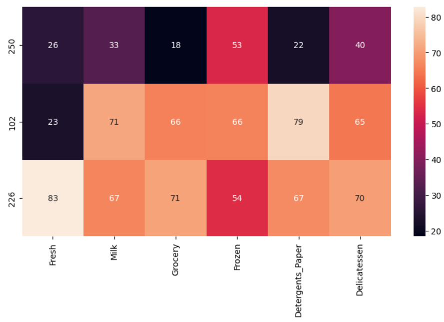

Fig. 4: Percentile Heatmap of the samples

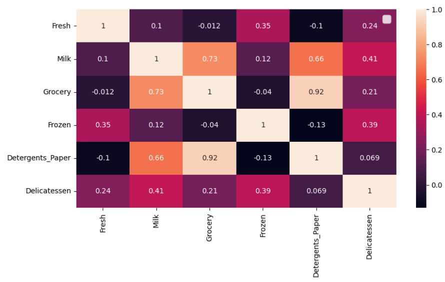

Fig. 5: Correlation Heatmap of the features

establishment could be a retailer. Moreover, a visual heatmap, as shown in Fig. 4, is used to describe this pattern. The correlation heatmap between the features, visualized in Fig. 5, indicates that "Detergents_Paper" has a high correlation with "Grocery" (score = 0.92) and "Milk" (score = 0.66) and so is the high predictability score.

To get a better understanding of the dataset, we can construct a scatter distribution (as in Fig. 6) of each of the six product features present in the data. If you found that the feature you attempted to predict is relevant for identifying a specific customer, then the scatter plot may not show any correlation between that feature and the others. Conversely, if you believe that feature is not relevant for identifying a specific customer, the scatter plot might show a correlation between that feature and another feature in the data.

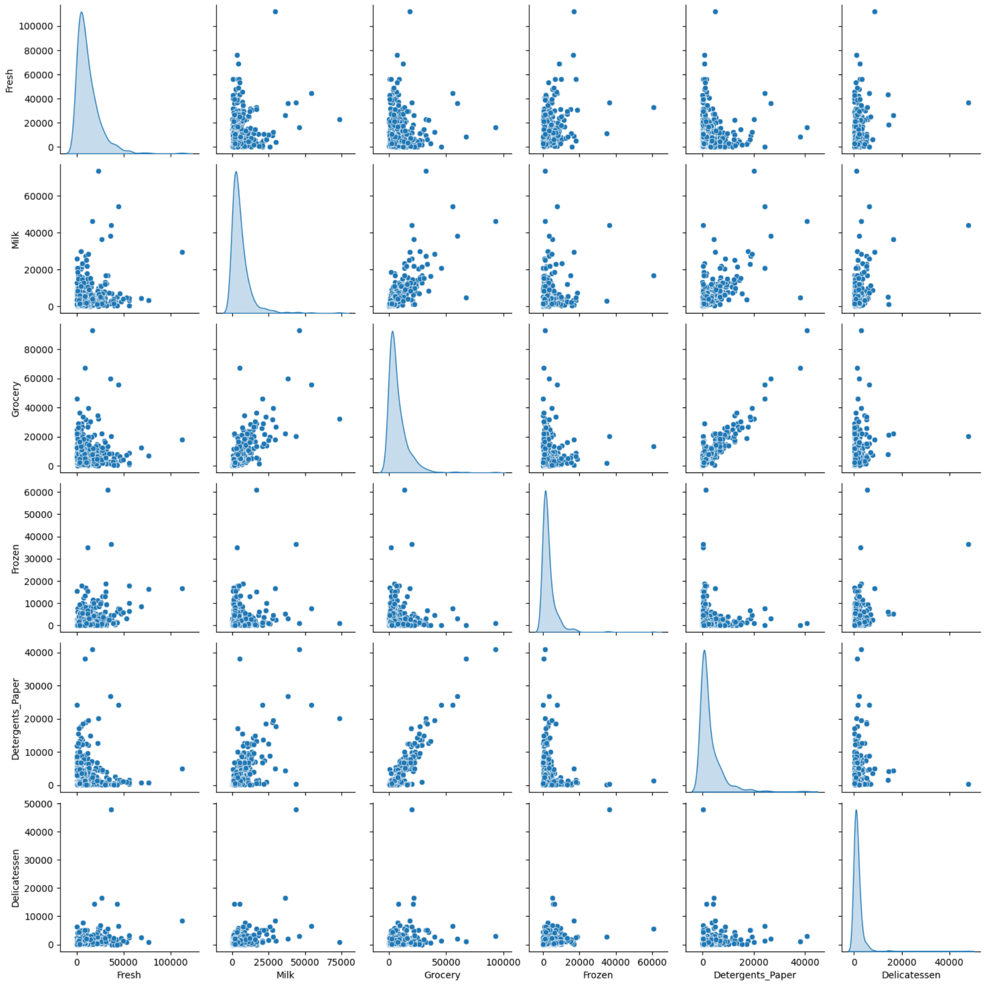

Fig. 6: Feature Correlation Scatter Distribution Plots

3.3 Dataset Pre-processing

In this section, we will preprocess the data to create a better representation of customers by performing a scaling on the data and detecting (and optionally removing) outliers. Preprocessing data is often times a critical step in assuring that results you obtain from your analysis are significant and meaningful. If data is not normally distributed, especially if the mean and median vary significantly (indicating a large skew), it is most often appropriate to apply a non-linear scaling — particularly for financial data. One way to achieve this scaling is by using a Box-Cox test, which calculates the best power transformation of the data that reduces skewness. A simpler approach which can work in most cases would be applying the natural logarithm, as in Fig. 7. After applying a natural logarithm scaling to the data, the distribution of each feature

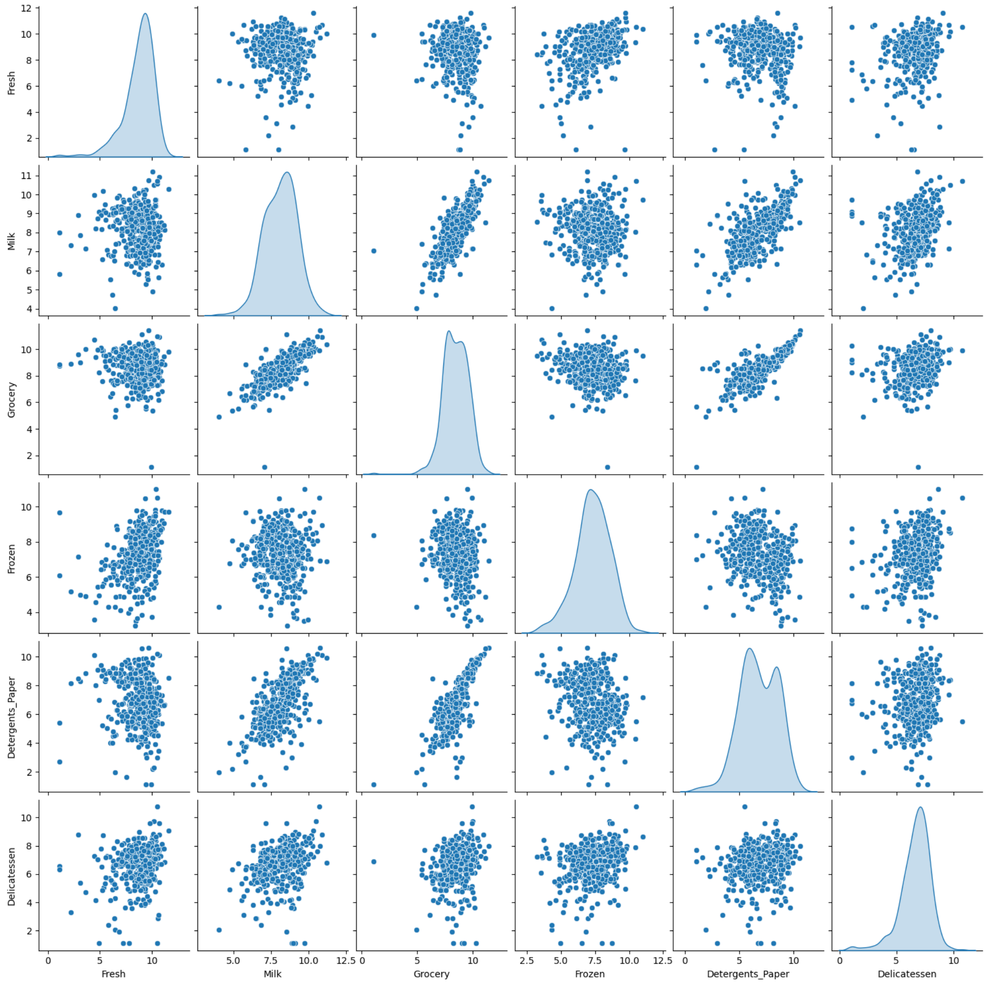

Fig. 7: Feature Correlation Scatter Distribution Plots after Log scaling

should appear much more normal. For any pairs of features you may have identified earlier as being correlated, observe here whether that correlation is still present (and whether it is now stronger or weaker than before).

Detecting outliers in the data is extremely important in the data preprocessing step of any analysis. The presence of outliers can often skew results which take into consideration these data points. There are many "rules of thumb" for what constitutes an outlier in a dataset. Here, we will use "Tukey's Method" for identifying outliers: An outlier step is calculated as 1.5 times the interquartile range (IQR). A data point with a feature that is beyond an outlier step outside of the IQR for that feature is considered abnormal. 5 data points are considered outliers for more than one feature based on the above analysis. As these data points are considered outliers for multiple features we could remove them. The above 5 data points that were outliers for more than one feature were removed from the data set so that we do not build a clustering model on the skewed data. We clearly know that clustering algorithms are sensitive to noisy data and outliers. As we use SSE to choose the cluster center, outiers will pull the center towards them and the cluster center will not be a proper representation of the data.

4 Methodology

The proposed methodology for customer segmentation follows a systematic pipeline consisting of data preprocessing, dimensionality reduction using Principal Component Analysis (PCA), and clustering using the K-Means algorithm. Section 4.1 explains the rationale behind the use of PCA, its mathematical foundations, and how it integrates with the clustering process. In section 4.2, the procedure of K-Means clustering is explained with relevant mathematical formulae.

4.1 Principal Component Analysis

Principal Component Analysis (PCA) is a widely used linear dimensionality reduction technique that transforms a dataset with possibly correlated variables into a set of linearly uncorrelated components known as principal components [12]. The goal of PCA is to reduce the number of input variables while retaining the maximum variance present in the original data. This not only simplifies the dataset but also improves the performance and visualization of unsupervised learning algorithms such as K-Means.

Mathematically, PCA works by computing the eigenvectors and eigenvalues of the covariance matrix of the standardized data. The eigenvectors define the directions (principal components) in which the data varies the most, while the eigenvalues indicate the amount of variance explained by each component. The components are ranked in descending order of their eigenvalues, and a subset of top components is selected to represent the data in reduced dimensions. Let X be the standardized data matrix. PCA proceeds through the following steps:
1) Compute the covariance matrix $\Sigma : \Sigma = \frac{1}{n-1} X^T X$

2) Compute the eigenvectors and eigenvalues of $\Sigma : \Sigma \mathbf{v}_i = \lambda_i \mathbf{v}_i \quad \text{for} \quad \text{i} = 1, 2, \ldots, \text{d}$

3) Sort the eigenvectors by decreasing eigenvalues and select the top k components.

4) Project the data onto the new feature space:

$$Z = XW$$

where $\text{W} = [\text{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$

In our study, PCA was applied after feature scaling [13], and the number of components was selected such that at least 90% of the total variance in the original data was retained. This dimensionality reduction step not only improved clustering quality but also mitigated the "curse of dimensionality," reduced noise, and made it feasible to visualize the clusters in two or three dimensions.

PCA is especially advantageous in datasets like the Wholesale Customers Dataset, where some features exhibit high skewness and multicollinearity. By transforming these features into uncorrelated components, PCA enables more meaningful cluster formation, as clustering algorithms such as K-Means are sensitive to feature scaling and correlation [14].

4.2 K-means clustering

K-Means is a partition-based unsupervised learning algorithm widely used in customer segmentation due to its simplicity, scalability, and efficiency. It aims to divide a dataset into k non-overlapping, spherical clusters where each data point belongs to the cluster with the nearest mean (centroid). The resulting clusters help uncover hidden customer patterns based on purchasing behavior and preferences.

It operates by minimizing the intra-cluster variance, or equivalently, the within-cluster sum of squares (WCSS). Due to its simplicity, scalability, and efficiency, K-Means is commonly used in customer segmentation tasks [1]. The pseudocode 1 of the K-Means algorithm is illustrated below.

Algorithm 1 K-Means Clustering Algorithm

Require: Dataset $X = \{x_1, x_2, \ldots, x_n\}$, Number of clusters k
Ensure: Cluster assignments $C = \{C_1, C_2, \ldots, C_k\}$ and centroids $\mu = \{\mu_1, \mu_2, \ldots, \mu_k\}$

Initialize k centroids $\mu_1, \mu_2, \ldots, \mu_k$ randomly or using k-means++
repeat
    Assignment Step:
    for each data point $x_i \in X$ **do**
        Assign x_i to the nearest centroid:

$$j = \arg\min_{l} \|x_i - \mu_l\|^2$$

        Assign x_i to cluster C_j
    end for
    Update Step:
    for each cluster C_j **do**
        Update centroid μ_j:

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

    end for
until cluster assignments no longer change or convergence criteria is met **return** Final clusters C and centroids μ

Since K-Means performs best on isotropic, low-dimensional, and uncorrelated data, Principal Component Analysis (PCA) was applied beforehand to reduce dimensionality and eliminate feature collinearity. This preprocessing step enhances cluster separability and speeds up convergence while enabling 2D or 3D visualization of the results.

5 Results and Analysis

5.1 Experimental Environment

To maintain experimental consistency, the models were trained in an identical setup using Google Colab's resources, including 24 GB of RAM and an NVIDIA Tesla T4 GPU. For building the models, we have used the Scikit-Learn Python library.

5.2 Experiment Results

5.2.1 Feature Transform and PCA

Now that the data has been scaled to a more normal distribution and has had any necessary outliers removed, we can now apply PCA to the "good" data to discover which dimensions about the data best maximize the variance of features involved. In addition to finding these dimensions, PCA will also report the explained variance ratio of each dimension — how much variance within the data is explained by that dimension alone, as shown in Fig. 8.

Note that a component (dimension) from PCA can be considered a new "feature" of the space, however it is a composition of the original features present in the data. The 1st Principal Component (PC1) captures high variance in "Detergents_Paper" followed by "Grocery" and "Milk". This first PC explains a strong correlation between these features as was observed earlier. An increase in PC1 is associated with a high purchase cost for "Detergents_Paper" followed by "Grocery" and "Milk", and this could be a "Retailer".

The 2nd PC captures high variance in "Fresh" followed by "Frozen" and "Delicatessen". An increase in PC2 is associated with a high purchase cost for "Fresh" followed by "Frozen" and "Delicatessen", and this could be a "Cafe/Restaurant".

The 3rd PC captures high variance in "Fresh" followed by "Delicatessen". An increase in PC3 is associated with a high purchase cost for "Fresh" followed by "Delicatessen", and this could be a "Cafe/Restaurant".

The 4th PC captures high variance in "Frozen" followed by "Delicatessen". An increase in PC4 is associated with an high purchase cost for "Frozen" followed by "Delicatessen", and this could be a "Cafe/Restaurant".

5.2.2 Dimensionality Reduction with PCA

When using principal component analysis, one of the main goals is to reduce the dimensionality of the data — in effect, reducing the complexity of the problem. Dimensionality reduction comes at a cost: Fewer dimensions used implies less of the total

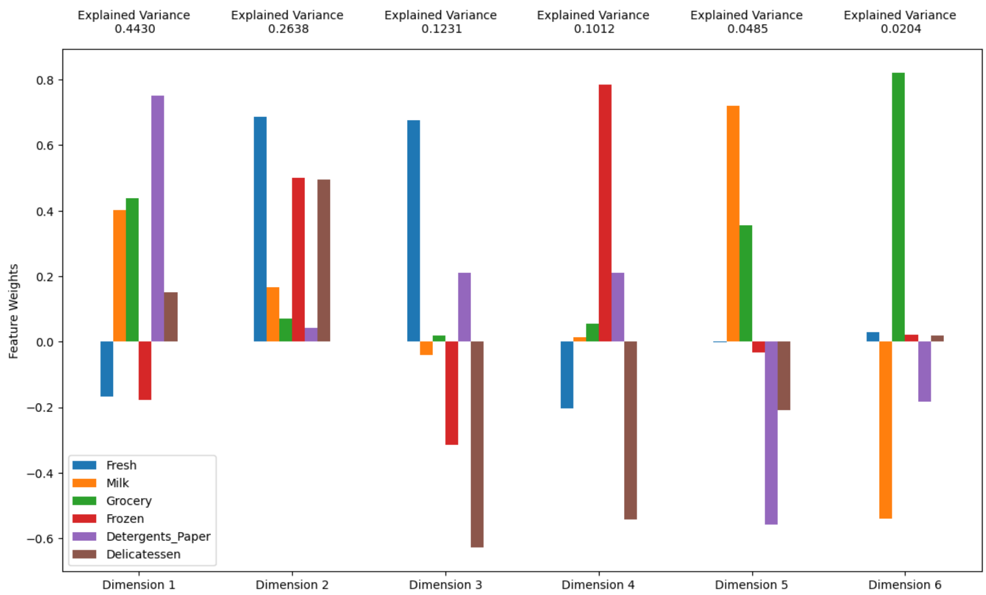

Fig. 8: PCA plot considering all dimensions/components

variance in the data is being explained. Because of this, the cumulative explained variance ratio is extremely important for knowing how many dimensions are necessary for the problem. Additionally, if a significant amount of variance is explained by only two or three dimensions, the reduced data can be visualized afterwards, which is for this experiment's case. It can be inferred that the values for the first two dimensions remain unchanged when compared to a PCA transformation in six dimensions.

A biplot is a scatterplot where each data point is represented by its scores along the principal components. The axes are the principal components (in this case Dimension 1 and Dimension 2), as pictured in Fig. 9. In addition, the biplot shows the projection of the original features along the components. A biplot can help us interpret the reduced dimensions of the data, and discover relationships between the principal components and original features.

Once we have the original feature projections (in red), it is easier to interpret the relative position of each data point in the scatterplot. For instance, a point the lower right corner of the figure will likely correspond to a customer that spends a lot on "Milk", "Grocery" and "Detergents_Paper", but not so much on the other product categories. From visualization (Fig. 9), we can see that for Dimension 1, "Detergents_Paper", "Grocery" and "Milk" are strongly correlated on the negative side. For the Dimension 2, "Fresh", "Frozen" and "Delicatessen" are strongly correlated in the negative direction.

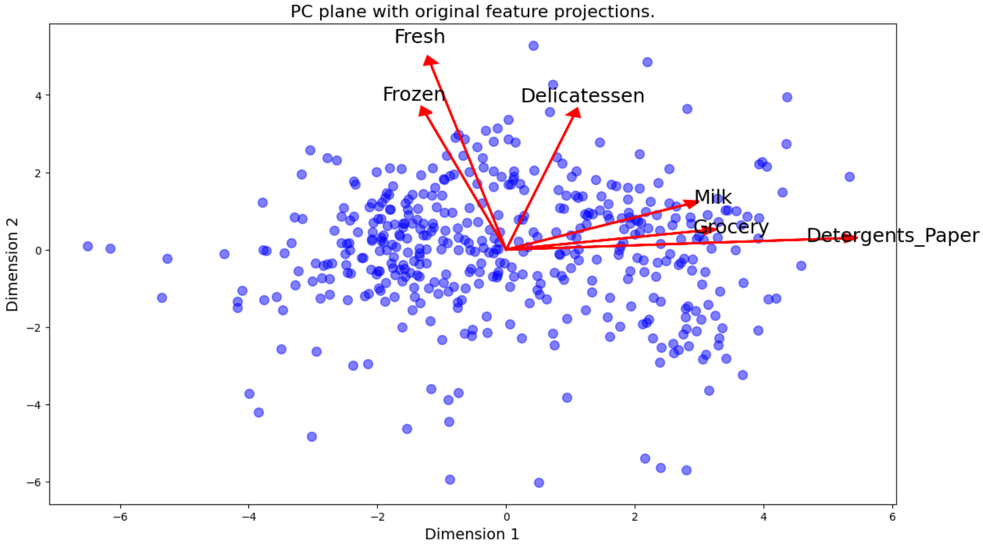

Fig. 9: Bi-plot showing PCA with 2 dimensions/components

Table 2: Silhouette Coefficients for Different Cluster Counts

Number of Clusters (k)	Silhouette Coefficient
2	0.4263
3	0.3352
4	0.3230
5	0.3044
6	0.3612
7	0.3293
8	0.3589
9	0.3698
10	0.3340
11	0.3260
12	0.3568
13	0.3577
14	0.3511

5.2.3 Customer Segmentation Analysis

Depending on the problem, the number of clusters that you expect to be in the data may already be known. When the number of clusters is not known a priori, there is no guarantee that a given number of clusters best segments the data, since it is unclear what structure exists in the data — if any. However, we can quantify the “goodness” of a clustering by calculating each data point’s silhouette coefficient. The silhouette coefficient for a data point measures how similar it is to its assigned cluster, scoring from -1 (dissimilar) to 1 (similar). Calculating the mean silhouette coefficient provides for a simple scoring method of a given clustering.

Fig. 10: Clusters

Table 2 shows the silhouette coefficient values computed for different cluster counts. From the above table, we see that the silhouette score is highest for 2 clusters. So, we will use K-means to divide into 2 clusters.

Each cluster (in Fig. 10) present in the visualization above has a central point. These centers (or means) are not specifically data points from the data, but rather the averages of all the data points predicted in the respective clusters. For the problem of creating customer segments, a cluster's center point corresponds to the average customer of that segment. Since the data is currently reduced in dimension and scaled by a logarithm, we can recover the representative customer spending from these data points by applying the inverse transformations.

It can be inferred from Table 3 that Segment 0 - these customers are purchasing less Milk, less Grocery and less Detergents_Paper than the population mean. This segment of customers tend to purchase more Frozen/Fresh and could be Restaurants or Hotels. Segment 1 - These customers are purchasing more Milk, more Grocery and more Detergents_Paper than the population mean. This segment of customers tend to purchase more Grocery/Milk/Detergents_Paper and could be Retailers or Supermarkets. Other observations include Segment 0 customers are purchasing more Fresh than Segment 1 customers. Segment 0 customers are purchasing more Frozen than Segment 1 customers. Segment 1 customers are purchasing more Detergents_Paper than Segment 0 customers. Segment 1 customers are purchasing more Delicatessen than Segment 0 customers.

Sample 0 - This customer purchased very less of "Milk", "Grocery", "Detergents_Paper" and definitely falls in Segment 0 (evident from Table 4). Sample 1 -

Table 3: Average Spending per Product Category by Customer Segment

Segment	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
Segment 0	4016.0	7869.0	12050.0	955.0	4526.0	1035.0
Segment 1	8878.0	1891.0	2469.0	2091.0	292.0	681.0

This customer purchased a lot of “Milk”, “Grocery”, “Detergents_Paper” and falls in Segment 1. Sample 2 - This customer purchased a lot of “Milk”, “Grocery”, “Detergents_Paper” and falls in Segment 1.

Table 4: Average Spending per Product Category of 3 Sample Customers

Samples	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
0	3191	1993	1799	1730	234	710
1	2932	6459	7677	2561	4573	1386
2	20782	5921	9212	1759	2568	1553

The predictions for each of the samples are mostly consistent with our observations. (As the initial predictions from EDA are not made for just 2 segments, there is a little difference with the above predictions that are made of just 2 cluster.

6 Conclusion, Limitations & Future Works

In this study, we explored unsupervised learning techniques for effective customer segmentation using the Wholesale Customers dataset. By applying Principal Component Analysis (PCA) for dimensionality reduction followed by K-Means clustering, we were able to reveal distinct patterns in customer purchasing behavior across key product categories such as Fresh, Milk, Grocery, and Detergents_Paper. PCA not only enhanced clustering efficiency but also enabled better visualization of customer groups, while K-Means provided interpretable and actionable segments.

Our analysis demonstrated that a smaller number of clusters (i.e., K=2) yielded highest silhouette coefficient, suggesting well-separated and compact customer groups. The segment profiles revealed interesting trends, such as some customer groups prioritizing fresh and frozen goods, while others showed high demand for detergents and groceries — insights that can be directly applied in retail marketing, product stocking, and personalized promotions. Despite promising results, this work presents several limitations:

- Sensitivity to Initialization: The K-Means algorithm can converge to local minima depending on centroid initialization. Though mitigated using k-means++, randomness still affects outcomes.
- Fixed Cluster Count: Determining the optimal number of clusters remains a subjective process, relying on heuristic methods like the Elbow method or Silhouette score.
- Assumption of Spherical Clusters: K-Means assumes clusters of similar size and shape, which may not reflect real-world customer behavior.

- Dataset Size and Diversity: The Wholesale Customers dataset is relatively small and lacks demographic features, limiting generalizability and richness of customer profiles.

To address these challenges and expand this research:

- Advanced Clustering Methods: Future studies could explore density-based (e.g., DBSCAN), hierarchical, or model-based clustering (e.g., Gaussian Mixture Models) to handle non-spherical, uneven clusters.
- Ensemble Approaches: Combining multiple clustering techniques or using consensus clustering may yield more robust segments.
- Feature Enrichment: Incorporating additional data such as geographic, demographic, or behavioral variables could provide a more holistic view of customers.
- Temporal Analysis: Extending the model to time-series or seasonal data could uncover evolving customer trends.
- Deployment in Business Tools: Integrating the segmentation model into recommendation systems or CRM tools would enhance its practical impact in real-world retail environments.

References

[1] MacQueen, J.: Some methods for classification and analysis of multivariate observations. (1967). https://api.semanticscholar.org/CorpusID:6278891

[2] Ram, A., Jalal, S., Jalal, A., Manoj, K.: A density based algorithm for discovering density varied clusters in large spatial databases. International Journal of Computer Applications **3** (2010) https://doi.org/10.5120/739-1038

[3] Kohonen, T.: The self-organizing map. Proceedings of the IEEE **78**(9), 1464–1480 (1990) https://doi.org/10.1109/5.58325

[4] Kuo, R.J., Ho, L.M., Hu, C.M.: Integration of self-organizing feature map and k-means algorithm for market segmentation. Computers Operations Research **29**, 1475–1493 (2002) https://doi.org/10.1016/S0305-0548(01)00043-0

[5] Vesanto, J., Alhoniemi, E.: Clustering of the self-organizing map. IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council **11**, 586–600 (2000) https://doi.org/10.1109/72.846731

[6] Dolnicar, S.: Beyond "commensense segmentation" - a systematics of segmentation approaches in tourism. Faculty of Commerce - Papers **42** (2004)

[7] Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis (2015)

[8] Saumendra, D., Nayak, J.: Customer Segmentation via Data Mining Techniques: State-of-the-Art Review, pp. 489–507 (2022). https://doi.org/10.1007/978-981-16-9447-9_38

[9] Lewaaelhamd, I.: Customer segmentation using machine learning model: An application of rfm analysis. Journal of Data Science and Intelligent Systems **2** (2023) https://doi.org/10.47852/bonviewJDSIS32021293

[10] Afzal, A., Khan, L., Hussain, M.Z., Hasan, M.Z., Mustafa, M., Khalid, A., Awan, R., Ashraf, M., Khan, Z., Javaid, A.: Customer segmentation using hierarchical clustering, pp. 1–6 (2024). https://doi.org/10.1109/I2CT61223.2024.10543349

[11] Cardoso, M.: Wholesale customers. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5030X (2013)

[12] Jolliffe, I.: Principal component analysis springer verlag (2002)

[13] Shlens, J.: A Tutorial on Principal Component Analysis (2014). https://arxiv.org/abs/1404.1100

[14] Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, (2012). https://doi.org/10.1016/C2009-0-61819-5