KLE TECH.
KLE Technological University
Creating Value
Leveraging Knowledge
Department of MCA

A Project Work on

LEVERAGING CLUSTERING ALGORITHMS FOR OPTIMIZED CUSTOMER SEGMENTATION IN MARKETING CAMPAIGNS

Presenter:
Jaideep S(174)

Presenter:
Christopher(175)

Presenter:
Nayeem P(146)

Presenter:
Sai Nithin(171)

Under the guidance of:
Prof. Rashmi Benni

STRUCTURE OF PRESENTATION

1. Introduction
2. Literature Survey
3. Research Gap
4. Problem Statement
5. Objectives and Motivation
6. Software and Hardware Requirements
7. Functional and Non-Functional Requirements
8. Methodology

INTRODUCTION

Nowadays, **Marketing and E-commerce** field is having rapid growth and it’s highly competitive, Most of businesses face challenges in catering to diverse and ever-growing customers base.

Customer Segmentation approach can help a marketing team of an organization’s marketing campaigns which effectively target the needs and preferences of a specific set of customers, which leads to higher engagement and conversion rates, which will eventually helps in growth of organization.

INTRODUCTION

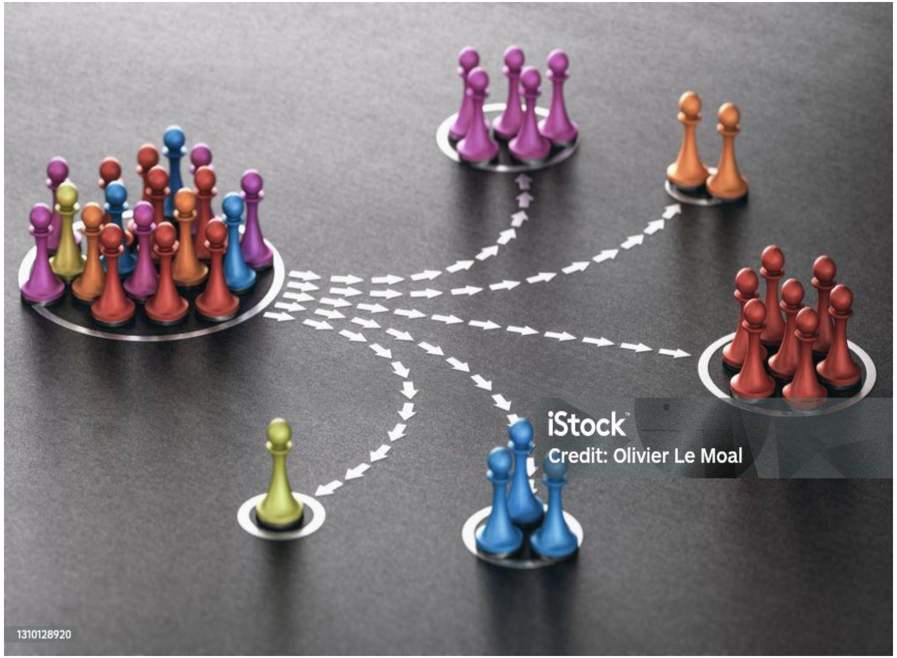
Case Study: As the Ecommerce industry experiencing rapid growth, the need of customer segmentation is becoming more critical and everyone has a different demands and choices.
Ex: Amazon, A global E-commerce platform.
With the help of **Machine Learning** and **Big Data Analytics**, We can process the large amount of customer data in real-time creating segments of customers, So we can Adjust our marketing strategies to maximize **conversion** and **profits**, leading in a competitive market, suggesting and providing customers the right products, at the right time.

CUSTOMER SEGMENTATION:

Customer Segmentation is the process of dividing a large and vast customer base into distinct groups or segments based on characteristics. These can include demographic data (age, gender, income), behavioral data (purchase history, website interactions), and psychographic data (lifestyle, interests).

The goal behind segmentation is to tailor marketing strategies to meet the specific needs and preferences of different customer bases. This allows businesses to create more effective and personalized and effective marketing campaigns, driving better customer satisfaction.

CUSTOMER SEGMENTATION:



MARKETING CAMPAIGNS:

A **Marketing campaign** is a series of strategic efforts to promote a organization's services, or products on television, print and online platforms for achieving goals or objectives.

In Segmented marketing, businesses use customer insights for creating customized offers, messages, and promotions which a dedicated for specific groups. This leads to improved conversion rates, higher engagement and better overall marketing performance.

EXAMPLE: Television advertisement, Printed hoardings , Social media marketing campaigns etc..

KLE TECH. KLE Technological University
Creating Value Leveraging Knowledge

MARKETING CAMPAIGNS:

MACHINE LEARNING:

Machine Learning is a subset of **artificial intelligence** that focuses on building models that learn from data and make decisions or predictions. It can easily detect patterns, cluster data and make predictions without any human intelligence.

We are using **unsupervised machine learning algorithms** such as **clustering algorithms** (K-Means, DBSCAN, Agglomerative).

How ML helps in customer segmentation:

1. Automation of segmentation
2. Handling large data volumes
3. Real-Time data processing
4. Personalized marketing campaigns
5. Predictive customer behavior

NEED IN MODERN MARKETING:

1. Nowadays due to high competition and data-driven world, organizations wants to know their customers in deeper level to stay ahead from others. For this tradition "**one-size-fits-all**" marketing approach are not effective.
2. Companies are encountering difficulties in understanding large, diverse customer bases without proper segmentation and wasting their ad spend and not getting much impressive results and fails to reach to right audience.
3. Through segmentation using ML and BDA allows companies to effectively allocate resources and design customized and personalized marketing strategies for different set of customers or audience.

TYPES OF CUSTOMER SEGMENTATION:

1. **Behavioral Segmentation**: Based on customer actions such as purchase frequency, brand loyalty and product range.
2. **Demographic Segmentation**: Based on factors like age, gender, income, and education level.
3. **Geographical Segmentation**: Based on the location of customer, such as country, city or pincode.
4. **Psychographic Segmentation**: Based on lifestyle, interest, attitudes, and opinions.

EXAMPLES OF APPLICATION IN MARKETING CAMPAIGN:

- **E-commerce** : Segmenting Customers based on browsing and purchase behaviors allows for personalized product recommendations and dynamic discounts.
- **Retail** : Retailers can use this to design loyalty programs that cater to different customer groups based on their buying habits and preferences.
- **Financial Services** : Banks can segment customers based on transactions patterns and demographic data to offer customized financial services.

LITERATURE SURVEY

1. "Customer Segmentation Model Using K-means Clustering on E-commerce".

Authors: Anshika Agrawal, Puneet Kaur, Monika Singh

Submitted to: IEEE Access

DOI: 10.1109/ICSCDS56580.2023.10105070

This Paper focuses on segmenting customers for e-commerce businesses to improve marketing strategies. This explores how the K-means clustering algorithm can be applied to e-commerce data, such as browsing behavior, purchase history and demographics, to know distinct customer groups with similar purchasing patterns.

This Solution allows businesses to alter their marketing campaigns based on specific needs of every customer segment, betterment of customer satisfaction and increasing sales.

2. "Customer Segmentation of Indian Restaurants on the basis of demographic locations using Machine Learning"**.**

Authors: Rishi Gupta Akash Verma Hari On Topal

Submitted to: IEEE Access

DOI: 10.1109/ICTAI53825.2021.9673153

This Paper focuses on segmenting customers of an indian restaurant based on demographic locations.This Highlights how various ML techniques, such as Clustering algorithms, can be applied to analyze demographic data like age, gender, income and mainly the location where they live to create a specific customer base.

By segmenting customers indian restaurants can know their customer base, optimize menu items, and tailor their marketing campaigns to increase revenue.

3. "Optimizing Customer Segmentation through Machine Learning"**.**

Authors: Anitha julian Hariprasath S R

Submitted to: IEEE Access

DOI: 10.1109/IC2PCT60090.2024.10486699

This Paper focuses on improving customer segmentation using advanced machine learning techniques. This explores how machine learning algorithms, such as decision trees, and neural networks, can be used to create more accurate customer segmentation model based on behavioral and demographic data.

It convey us need of optimization in customer segmentation for businesses so they can understand better their customer's needs and all.

4. “Customer Segmentation for Telecommunication using Machine Learning”**.**

Authors: Haitham H.Mahmoud A.Taufiq Asyhari

Submitted to: Springer Access

DOI: 10.1007/978-981-97-5489-2_13

This Paper discusses the application of machine learning techniques to segment customers in the telecommunication industry. This explores how algorithms can be used to analyze customer data, such as usage patterns, billing information, and service preferences, to identify specific customer bases.

This helps them to better understand customer behaviour, predict churn, and offer services needed by them from which company will improve customer satisfaction.

5. "Comparative Analysis of Machine Learning Models for Customer Segmentation"**.**

Authors: parmeshwara joga B.Harshini Rashmi sahay

Submitted to: Springer Access

DOI: 10.1007/978-3-031-35510-3_6

This Paper evaluates various machine learning techniques like clustering algorithms and others for Customer Segmentation.

It highlights the performance,strengths, and limitations of each model offering insights into which algorithms are more effective for customer segmentation.

This analysis enables businesses to improve their marketing strategies and customer engagement by identifying optimal approaches for grouping customers based on their behavior and preferences.

6. "E-commerce Customer Segmentation by Unsupervised Learning"**.**

Authors: S.M.hemadharshini R.Kanchana Devi S.Rajkumari R.Adline Freeda

Submitted to: Springer Access

DOI: 10.1007/978-981-99-8628-6_25

This Paper explores the use of unsupervised machine learning techniques, particularly K-means and Gaussian Mixture Models (GMM), for segmenting e-commerce customers based on purchasing behaviors. This segmentation enables businesses to tailor marketing strategies, improve customer engagement, and enhance decision-making in customer relationship management by gaining insights into customer preferences.

7. "Mall Customer Segmentation Using K-means Clustering"**.**

Authors: Ashwani Gurleen kaur Lekha rani

Submitted to: Springer Access

DOI: 10.1007/978-981-99-6553-3_35

This research paper outlines a methodology for segmenting mall customers using the K-Means clustering algorithm. The study aims to enhance marketing strategies and customer satisfaction by identifying distinct customer profiles through demographic and purchasing behavior data. By effectively categorizing customers into meaningful clusters, the K-Means algorithm facilitates targeted marketing efforts. The findings highlight the significance of data-driven approaches in retail for optimizing marketing strategies and improving overall customer experiences.

9. "Customer Segmentation Hyperparameter Tuning using Machine Learning "**.**

Authors: C Ganesh S Boomikha M Sneha

Submitted to: Springer Access

DOI: 10.1109/ICSSEECC61126.2024.10649417

This Paper focuses on improving customer segmentation by using Machine Learning along with Hyperparameter Tuning. It explores how adjusting hyperparameters in models can enhance the accuracy and efficiency of customer segmentation.

10. "Customer Segmentation in Retail using Machine Learning ".

Authors: Harshini Thatarvarti Joshna Rangala

Submitted to: Springer Access

DOI: 10.1109/I2CT57861.2023.10126155

This Paper focuses on improving customer segmentation by using Machine Learning in retail industry so that they can can analyze customer data, including purchasing habits, preferences, and demographics, to create meaningful customer groups.

It enable retailers to tailor marketing campaigns, product recommendations, and promotions to meet the specific needs of different customer types.

RESEARCH GAP:

Recent Studies focus on ML for segmentation but fail to improve flexibility of big data to handle real-time, large scale customer data and works on static rule based approaches and scale up for large datasets.

Our Project can fulfil this gap by developing a machine learning based segmentation model that can scale using big data tools and provide insights for marketing strategies.

PROBLEM STATEMENT:

"To design and develop a model to improve accuracy of customer segmentation model leveraging three different clustering algorithms (K-means, Agglomerative, DBSCAN) for marketing campaigns based on different types, leading to better engagement, conversion rates and sales".

MOTIVATION:

1. Growing consumer demand for personalized experiences
2. Competitive advantage to stay ahead in market
3. Growing business and Making profits

OBJECTIVES:

1. Implement a Machine Learning-based segmentation model to identify distinct customer bases based on behavioral and demographic data.
2. Utilize Big Data Analytics to ensure the model scales for large datasets and provides real-time insights.
3. Evaluate the effectiveness of the model in optimizing marketing campaigns and improving business outcomes, such as customer satisfaction and revenue.
4. Compare and Analyze three different clustering algorithms on customer segmentation model for optimizing accuracy.

FUNCTIONAL REQUIREMENTS:

1. System shall be able to process input from predictive performance of classification algorithms
2. System shall be able to optimize the solution for different datasets
3. System shall be able to provide good technique to analyse segmentation of customers.
4. Identify the best possible classification model
5. Perform result analysis

NON-FUNCTIONAL REQUIREMENTS:

1. **Compatibility**: The application should work on any machine with the required configuration.
2. **Availability**: The application should be available all the time.
3. **Performance**: The application must provide high performance.
4. **Efficiency**: The application must have good final test accuracy after training the model.
5. **Reliability**: The model should work for any computer-related component(software, or hardware, or a network) that consistently performs according to its specifications.

SOFTWARE AND HARDWARE REQUIREMENTS:

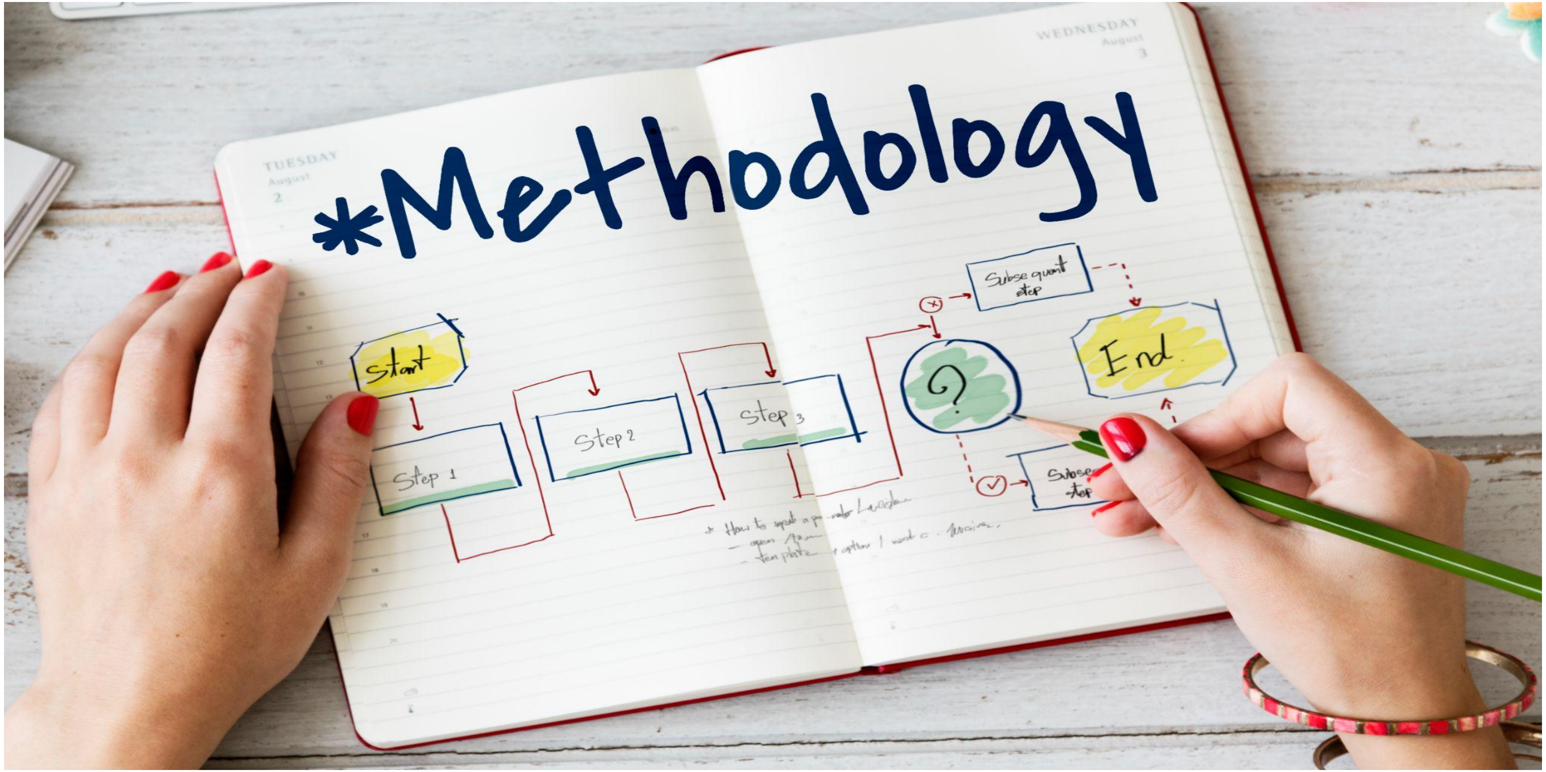
SOFTWARE REQUIREMENTS:

1. Colab / Jupyter Notebook
2. Python 3.10 / 3.11
3. Data Visualization: POWER BI
4. OS-Windows

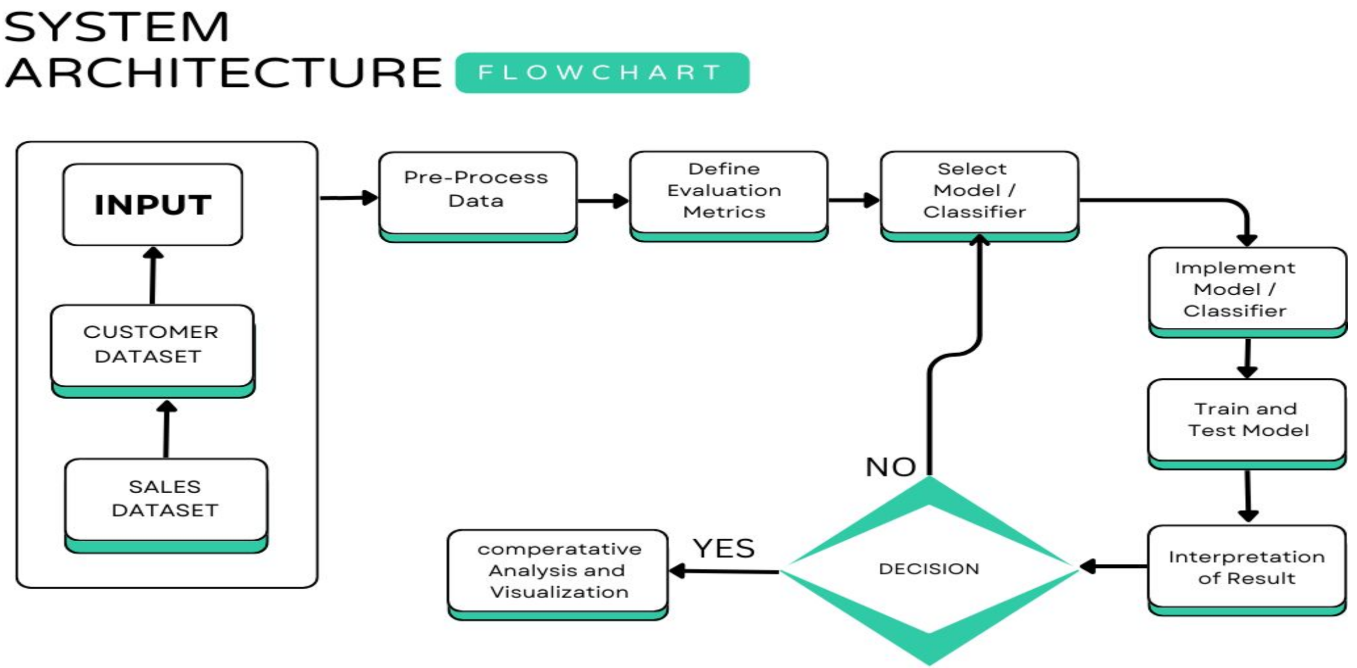
HARDWARE REQUIREMENTS:

1. 8GB RAM
2. Ryzen 5600H Processor

METHODOLOGY:



METHODOLOGY:

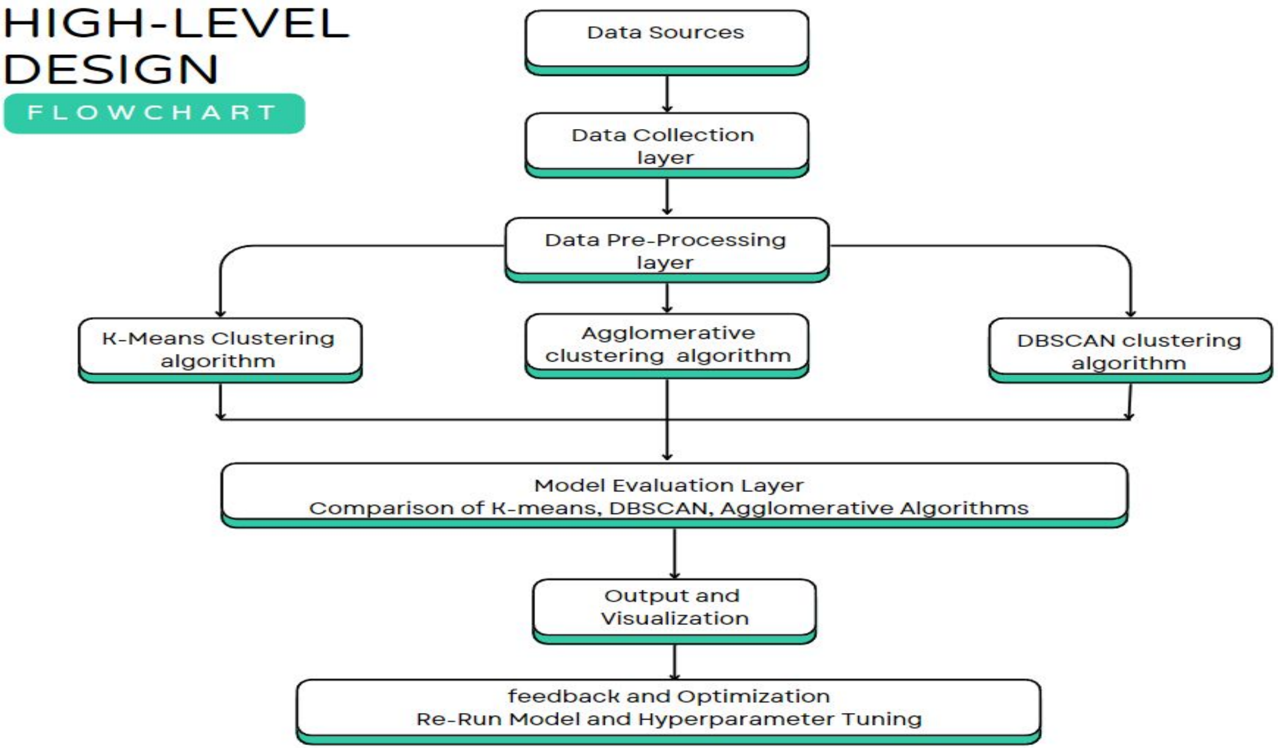


SYSTEM ARCHITECTURE:

The diagram outlines

1. Gather raw customer and sales data
2. Clean and normalize data to remove inconsistencies
3. Set metrics like accuracy, precision, or recall to access model performance
4. Choose suitable ML models
5. Build and configure the selected classifier models
6. Train the model using a training dataset and validate it with test dataset
7. Analyze the model's output to understand customer segments
8. Compare different model results and visualize key insights with graphs

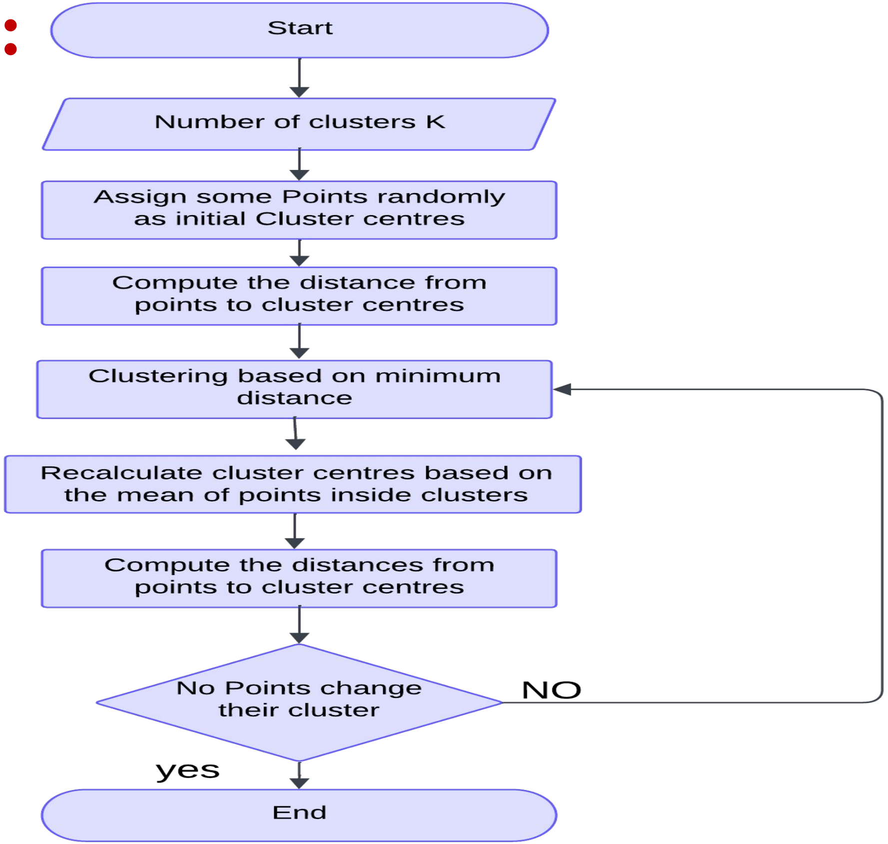
METHODOLOGY



ALGORITHMS:

1. K-Means Clustering
2. Agglomerative Clustering
3. DBSCAN Clustering

K-MEANS CLUSTERING:



K-MEANS CLUSTERING WORKING:

1. Select the number of clusters(k)
2. Randomly initialize k centroids
3. Assign customers to the nearest centroid
4. Recompute centroids based on customer segments
5. Repeat until convergence, forming customer segments

MATHEMATICAL EQUATION:

- Objective Function (cost function): The K-means algorithm aims to minimize the within-cluster sum of squares(WCSS)

Where:

K is the number of clusters, x represents data point in cluster Ci, μi is the centroid of cluster Ci, // x−μi // is the euclidean distance between the data point and centroid.

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

MATHEMATICAL EQUATION:

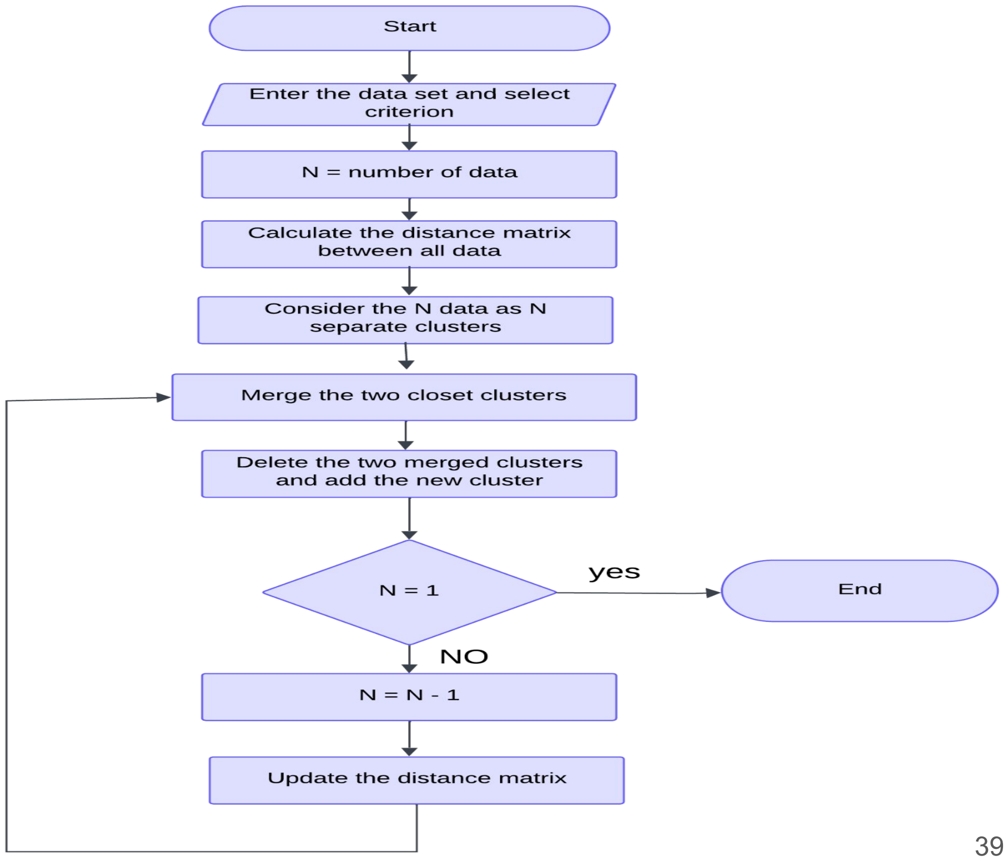
- Centroid Calculation: The Centroid of each cluster is updated by calculating the mean of all points in the cluster.

Where:

∣Ci∣ is the number of data points in cluster Ci, x represents the data points assigned to the cluster.

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$$

AGGLOMERATIVE CLUSTERING:



AGGLOMERATIVE CLUSTERING WORKING:

1. Start with each customer as an individual cluster
2. Merge the closest cluster based on distance metrics
3. Repeat until desired number of clusters is reached
4. Resulting clusters form a hierarchical structure for customer segmentation

MATHEMATICAL EQUATION:

Single Linkage (Minimum distance between points in two clusters):

$$dmin(Ci,Cj)=x\in Ci,y\in Cjmin||x-y||$$

Complete Linkage (Maximum distance between points in two clusters):

$$dmax(Ci,Cj)=x\in Ci,y\in Cjmax||x-y||$$

Average Linkage (Average distance between points in two clusters):

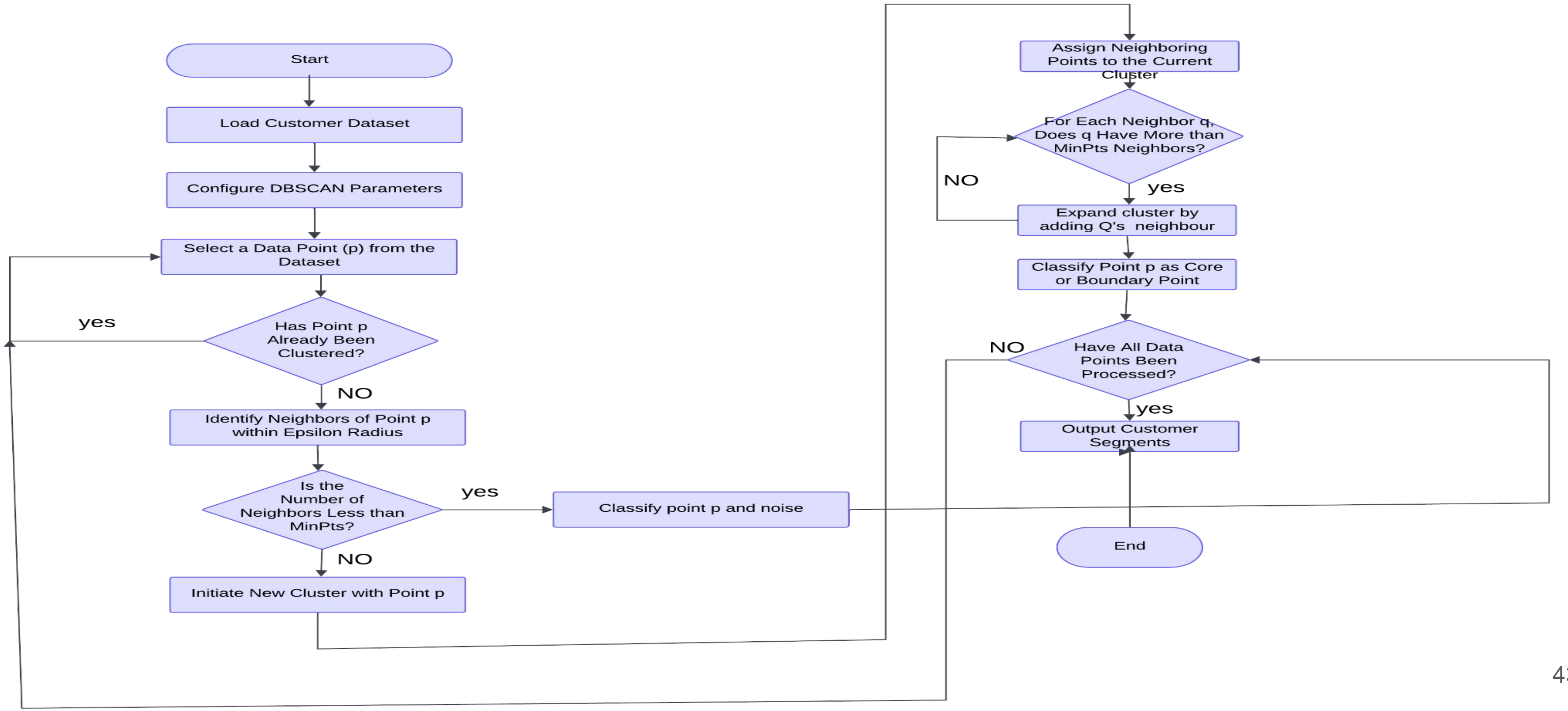
$$davg(Ci,Cj)=|Ci|\cdot|Cj|1x\in Ci\sum y\in Cj\sum||x-y||$$

MATHEMATICAL EQUATION:

where:

- Ci and Cj are two clusters.
- x and y are data points in clusters Ci and Cj, respectively.
- $\| x-y \|$ is the Euclidean distance between two data points.

DBSCAN CLUSTERING:



DBSCAN CLUSTERING WORKING:

1. Set parameters: Epsilon (ε) and MinPts
2. Identify core points with MinPts within ε radius
3. Expand clusters by adding density-reachable customers
4. Mark outliers as noise; final clusters represent customer segments

MATHEMATICAL EQUATION:

Core Definitions:

Epsilon (ε): Radius of the neighborhood.

MinPts: Minimum number of points required to form a dense region.

Core Point: A point *p* is a core point if at least MinPts points are within its ε-neighborhood, defined as:

$$|N\epsilon(p)| \geq \text{MinPts}$$

Where:

Nϵ(p) is the set of points within ε distance of point p:

$$N\epsilon(p)=\{q \in D : \text{dist}(p,q) \leq \epsilon\}$$

MATHEMATICAL EQUATION:

Density-Reachability:

A point q is **directly density-reachable** from p if:

$$q \in N\epsilon(p) \text{and} |N\epsilon(p)| \geq \text{MinPts}$$

DATASET DESCRIPTION:

Collected Dataset:

No. of Class Variables: 21

Dataset Size: 7044

Dataset Source:

Train:

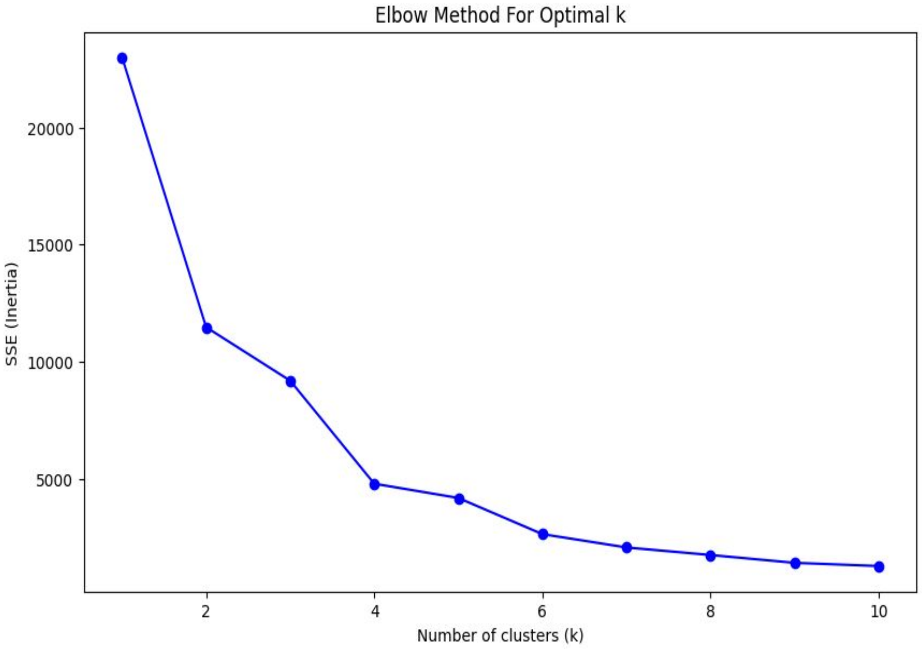
Test:

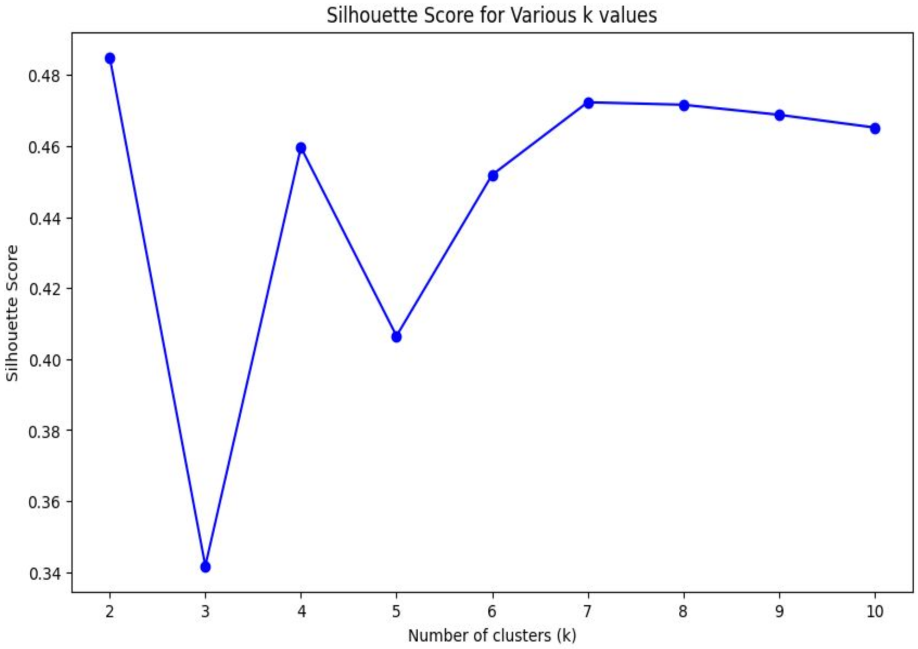
TEST CASE 01:

Metrics	K-Means Clustering	Agglomerative Clustering	DBSCAN Clustering
No. of Components	2	2	2
Silhouette Score	0.484	0.543	0.454
Calinski-Harabasz Index	7068.432	8134.948	2936.447
Davies-Bouldin Score	0.846	0.714	0.999
Avg Intra-Cluster Distance	1.140	0.954	1.328

TEST CASE 01:

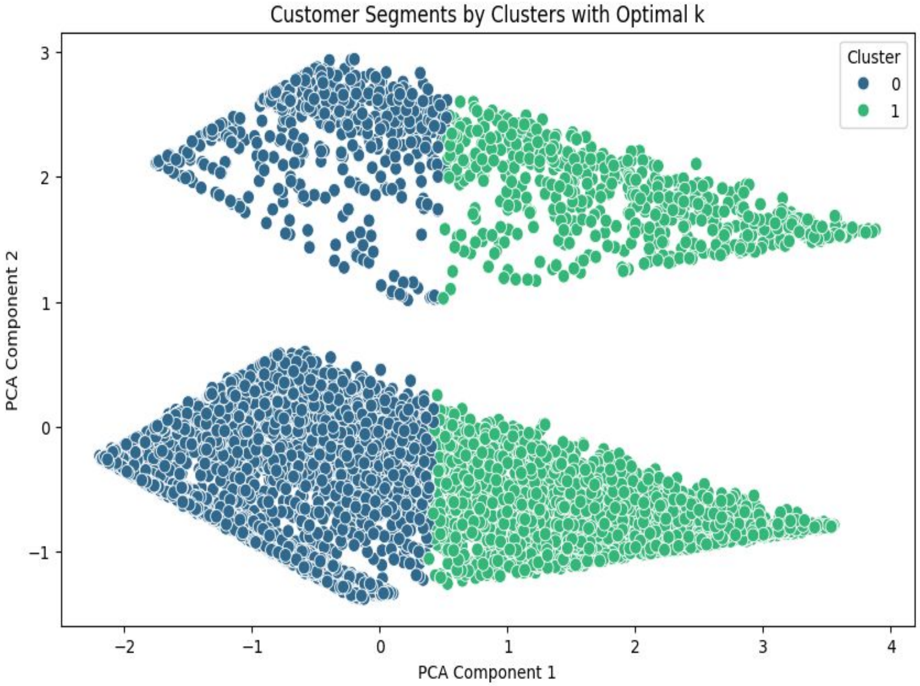
K-means Clustering:

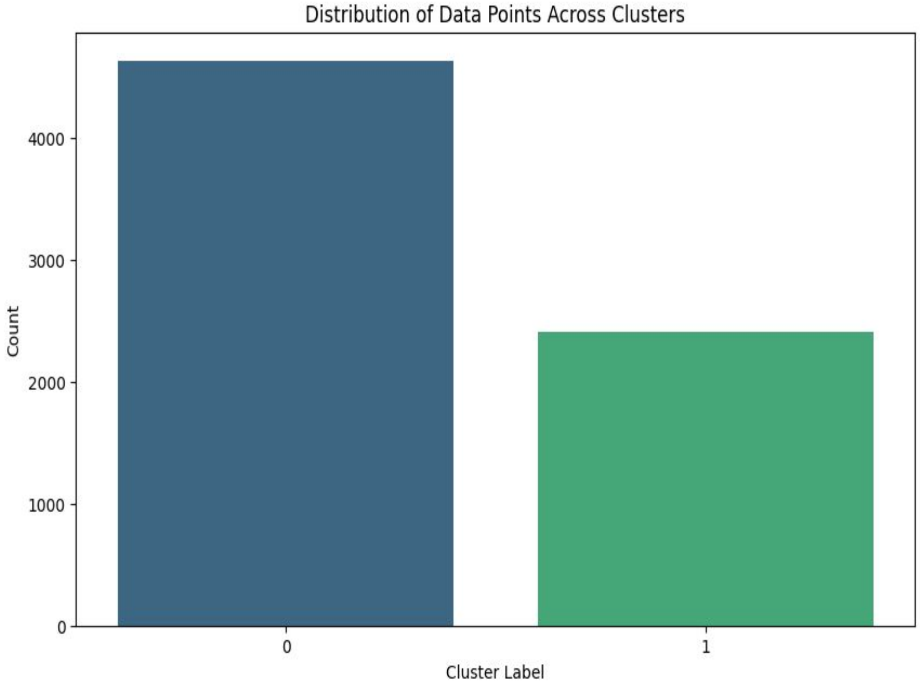




TEST CASE 01:

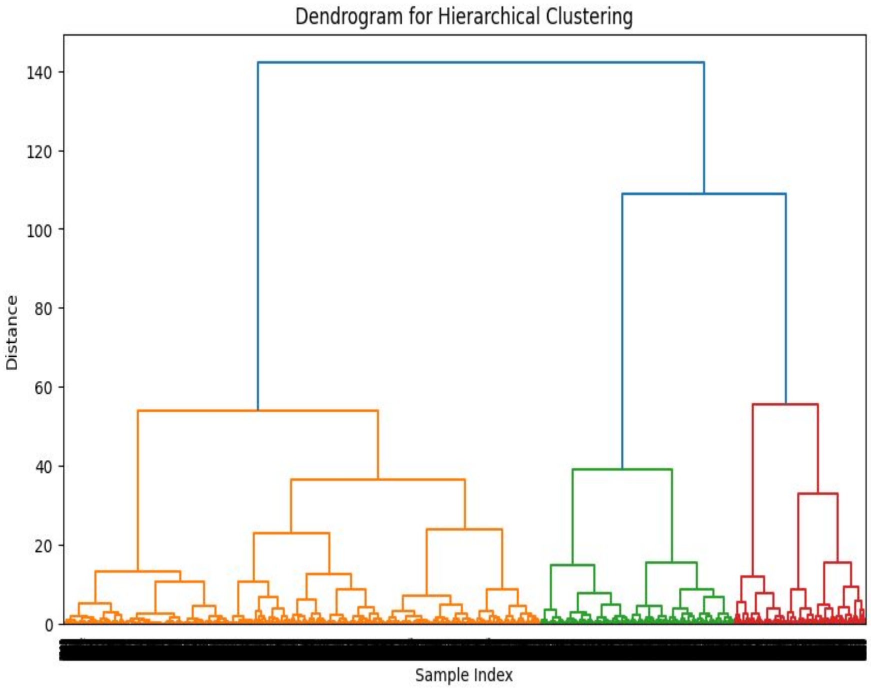
K-means Clustering:

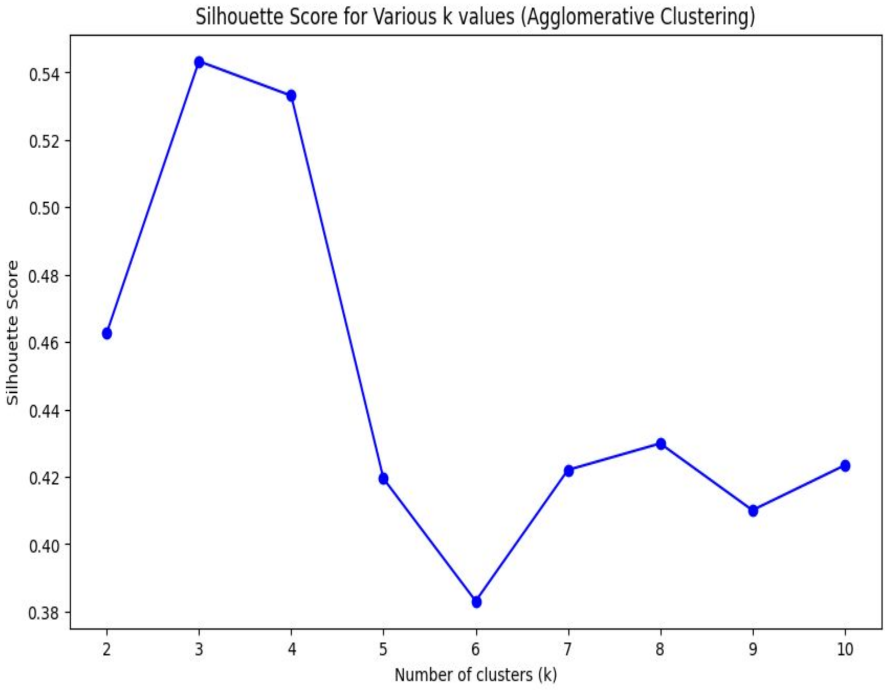




TEST CASE 01:

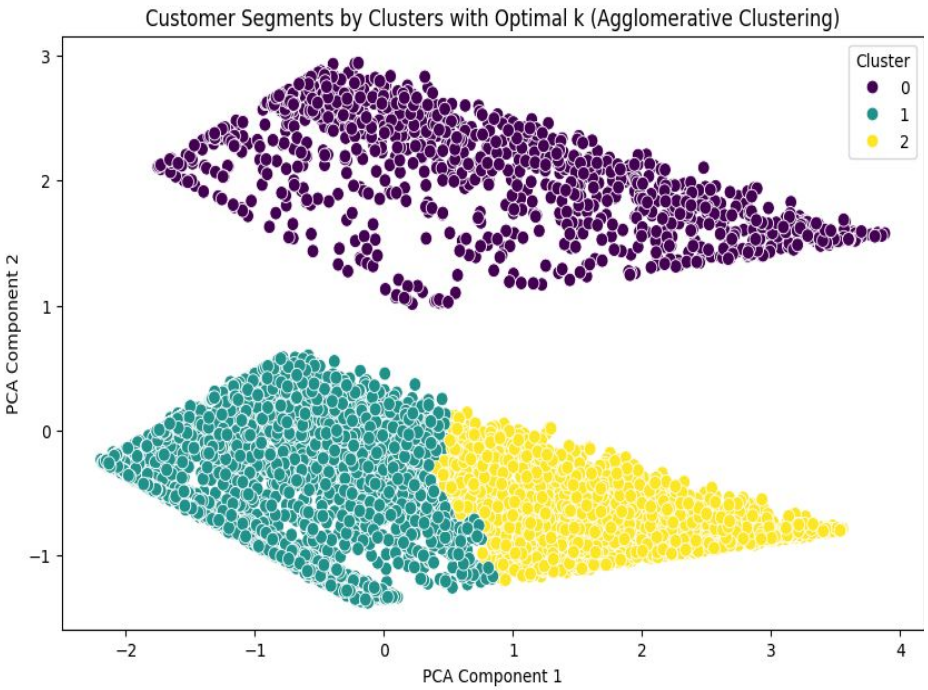
Agglomerative Clustering:

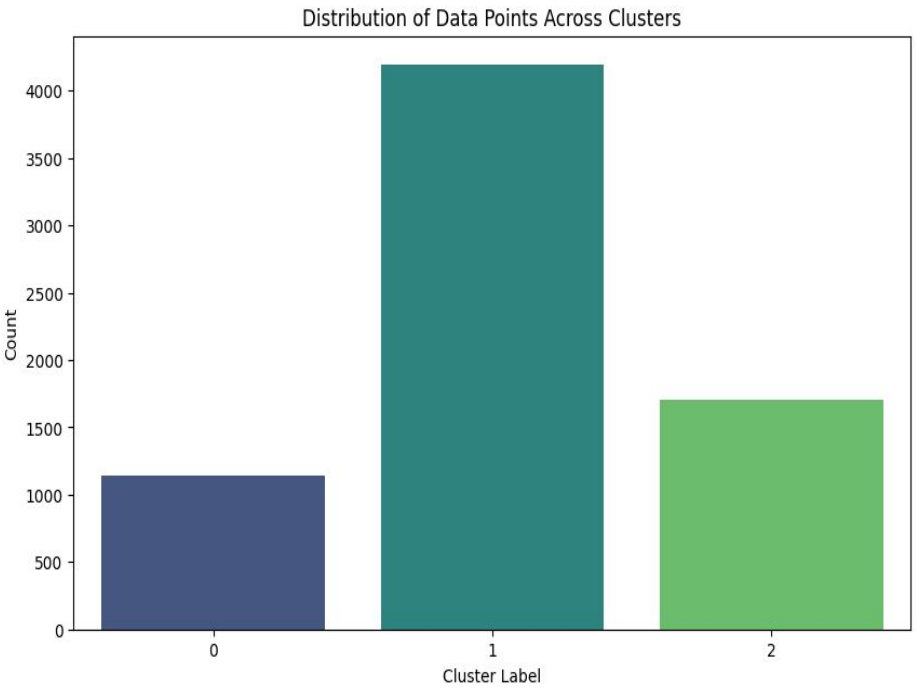




TEST CASE 01:

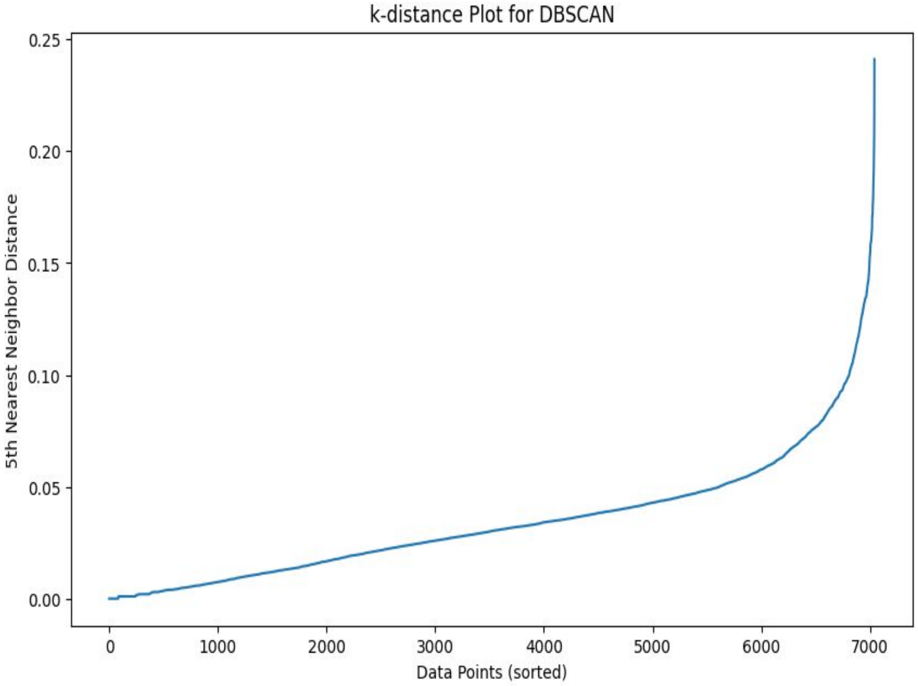
Agglomerative Clustering:

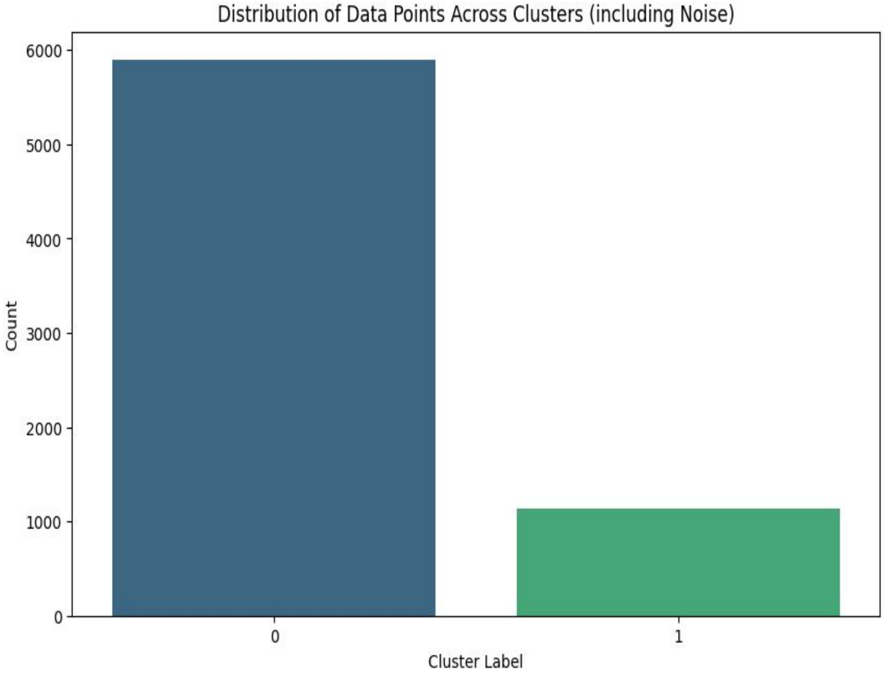




TEST CASE 01:

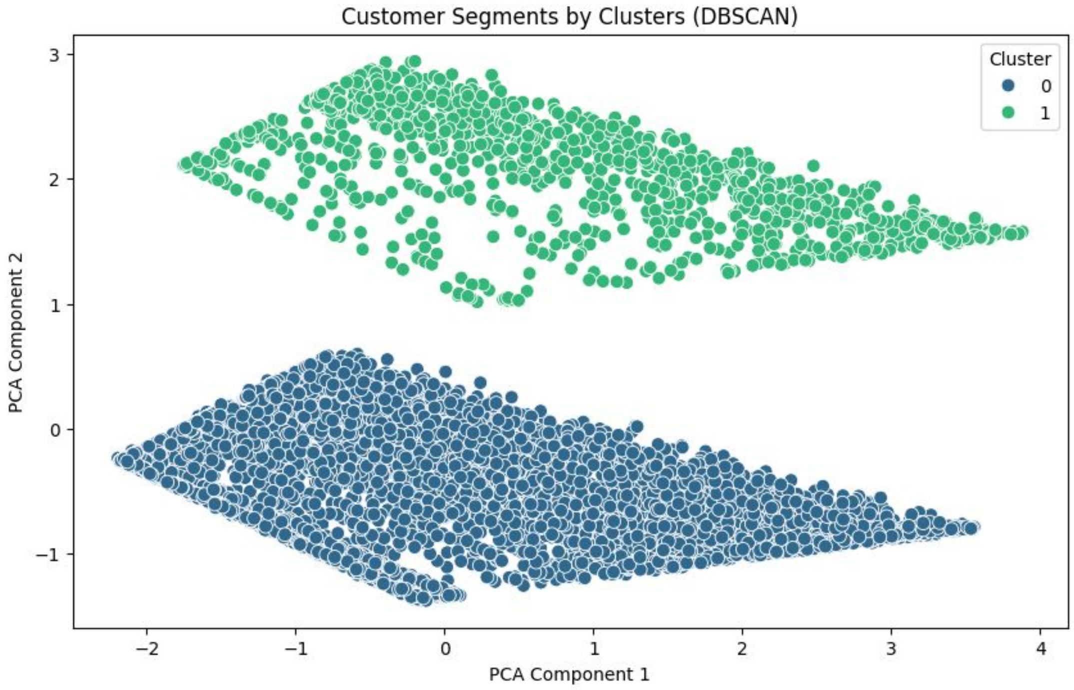
DBSCAN Clustering:





TEST CASE 01:

DBSCAN Clustering:

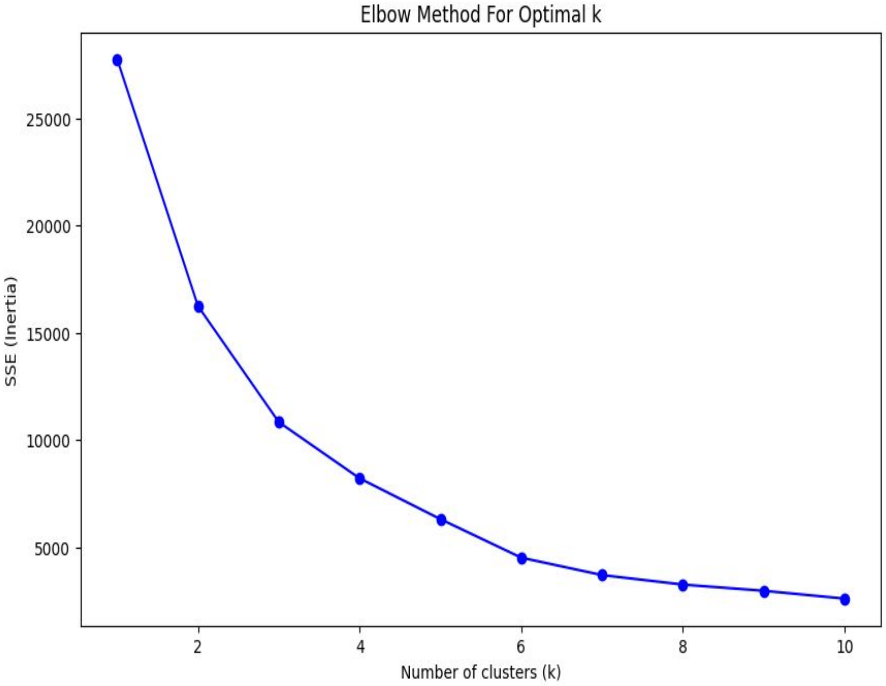


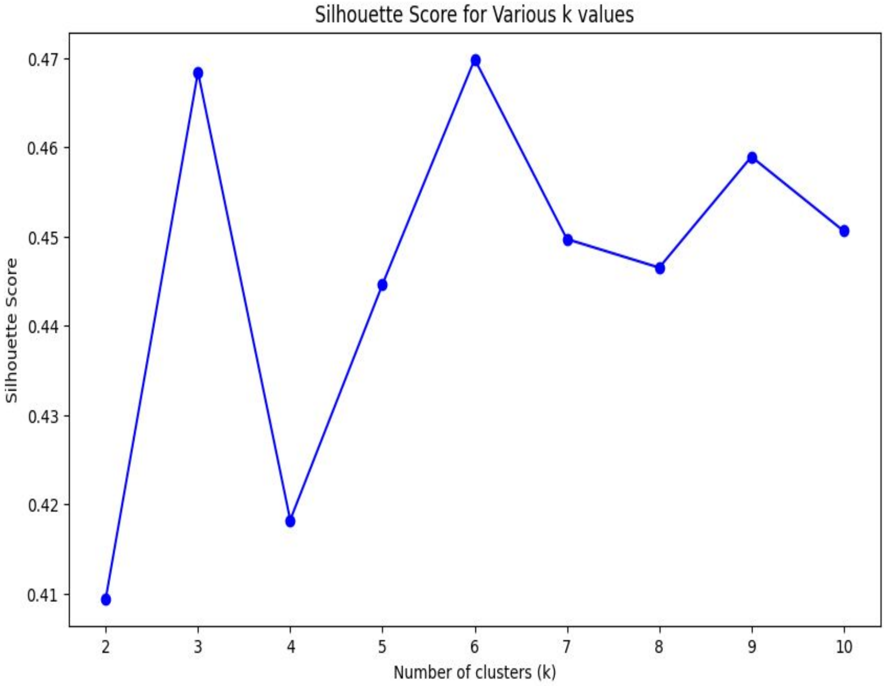
TEST CASE 02:

Metrics	K-Means Clustering	Agglomerative Clustering	DBSCAN Clustering
No. of Components	3	3	3
Silhouette Score	0.469	0.441	0.414
Calinski-Harabasz Index	7262.210	4561.212	2588.754
Davies-Bouldin Score	0.711	0.834	1.099
Avg Intra-Cluster Distance	0.767	1.137	1.535

TEST CASE 02:

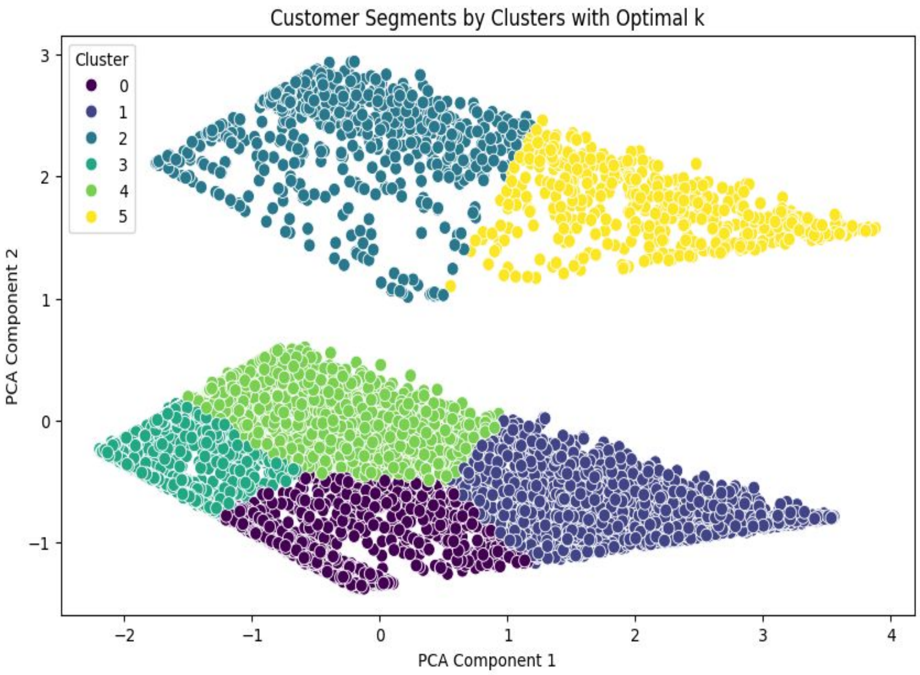
K-means Clustering:

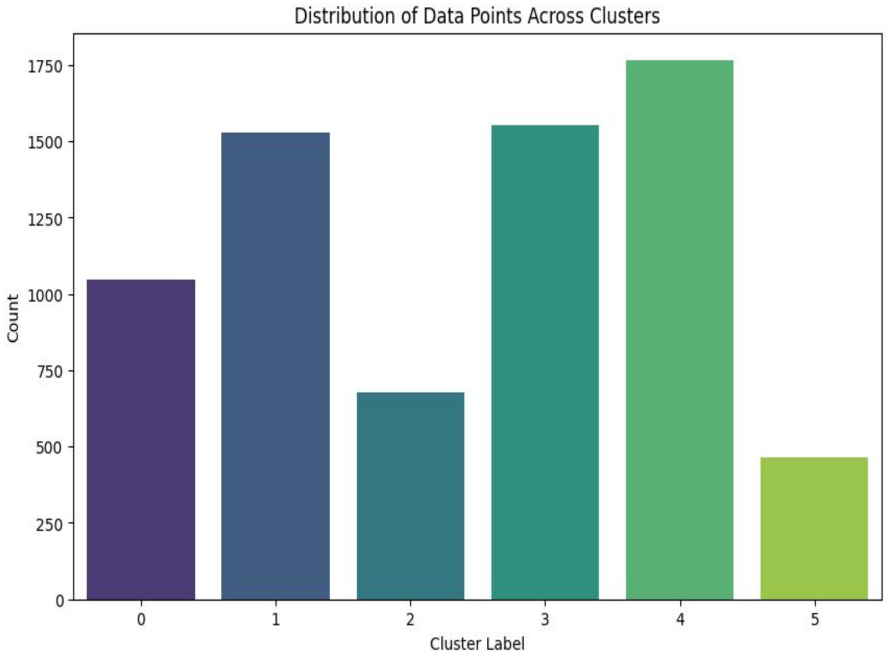




TEST CASE 02:

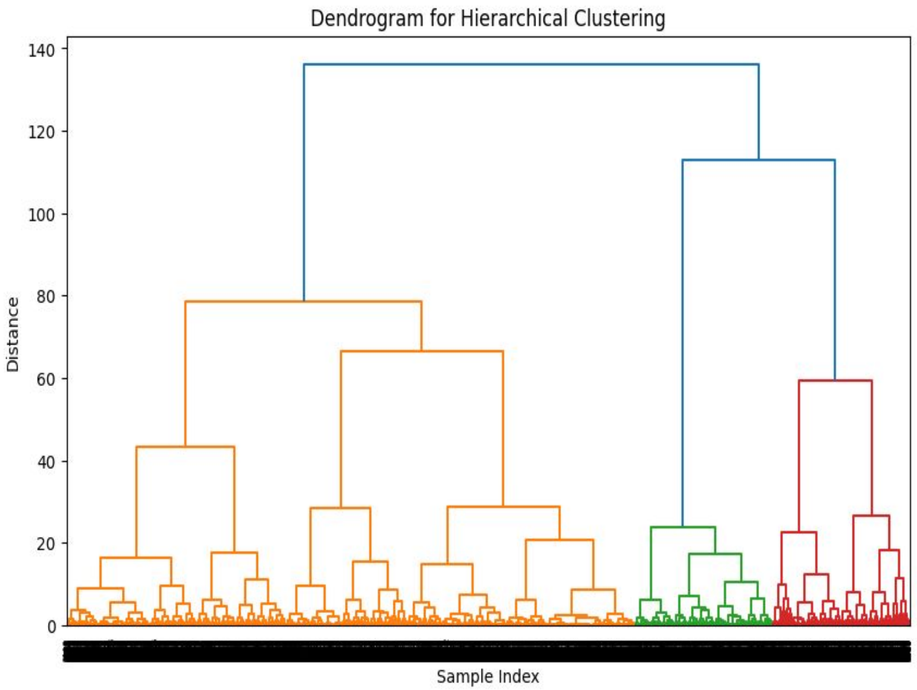
K-means Clustering:

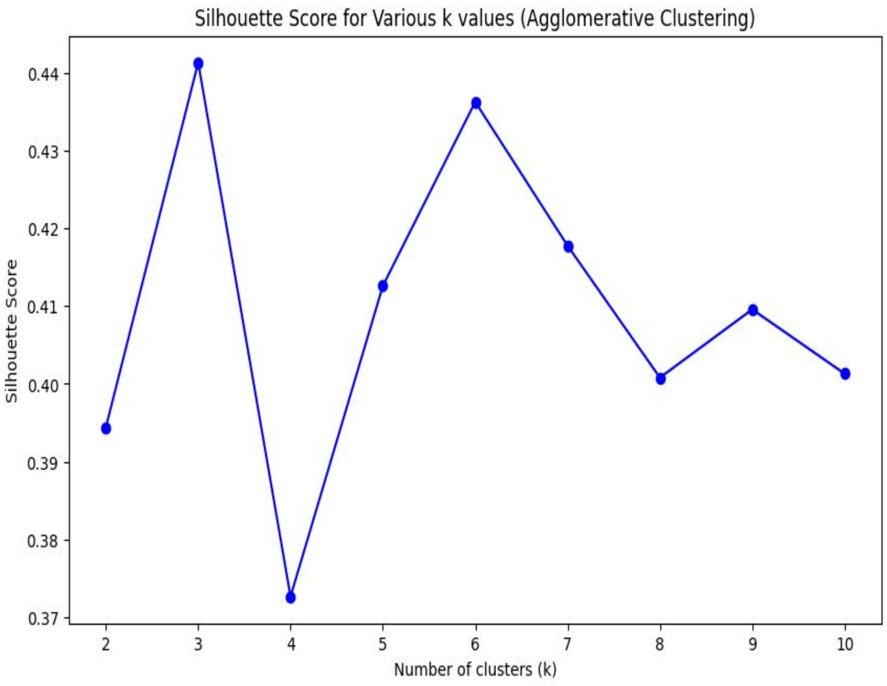




TEST CASE 02:

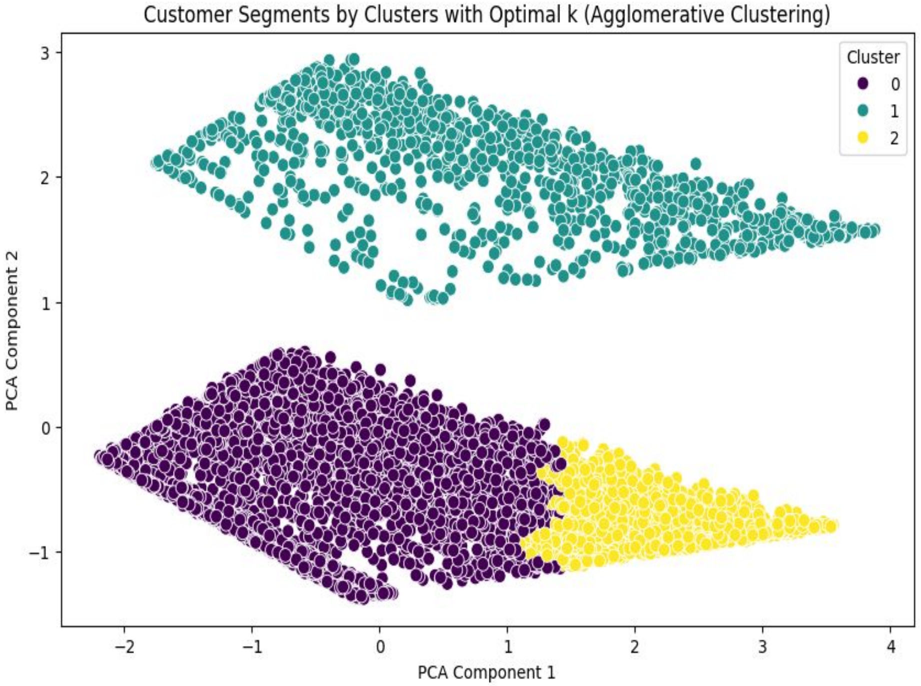
Agglomerative Clustering:

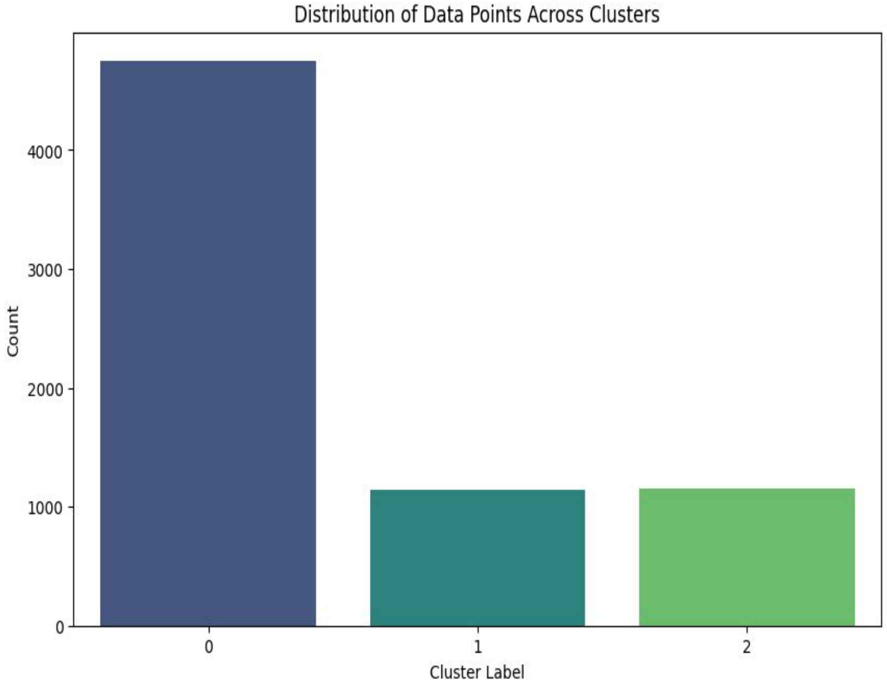




TEST CASE 02:

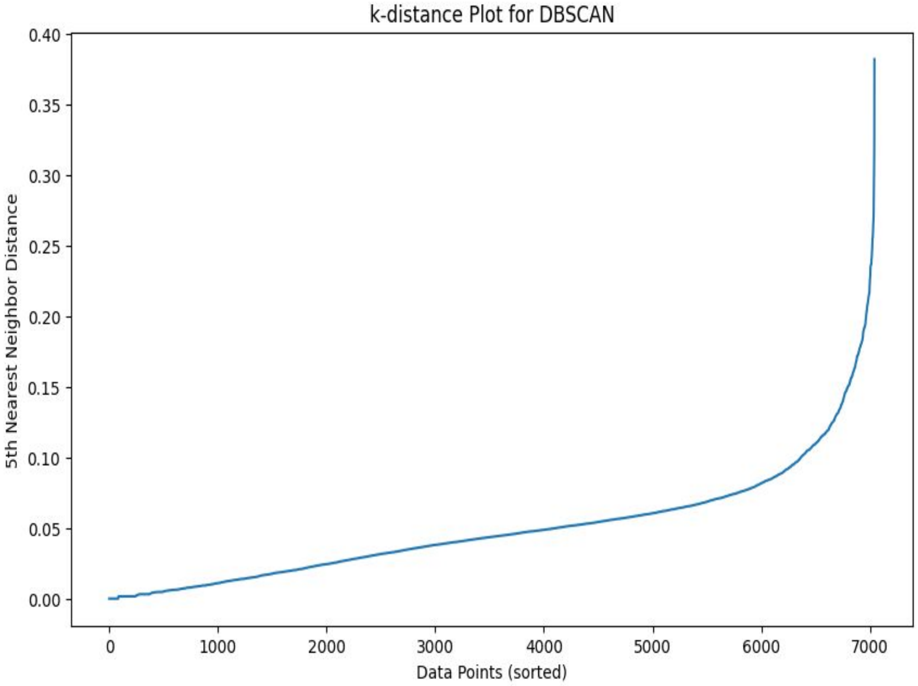
Agglomerative Clustering:

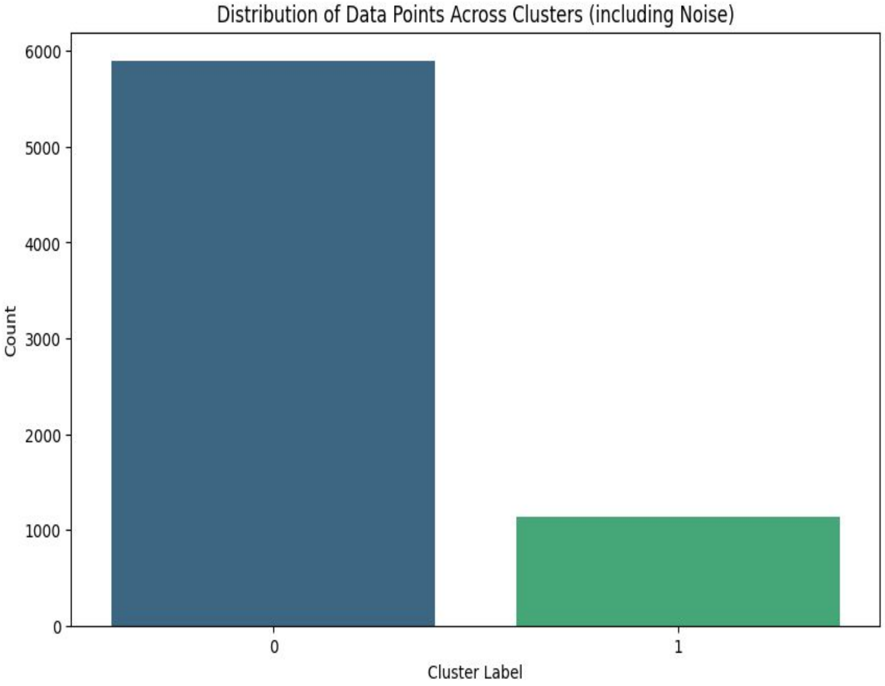




TEST CASE 02:

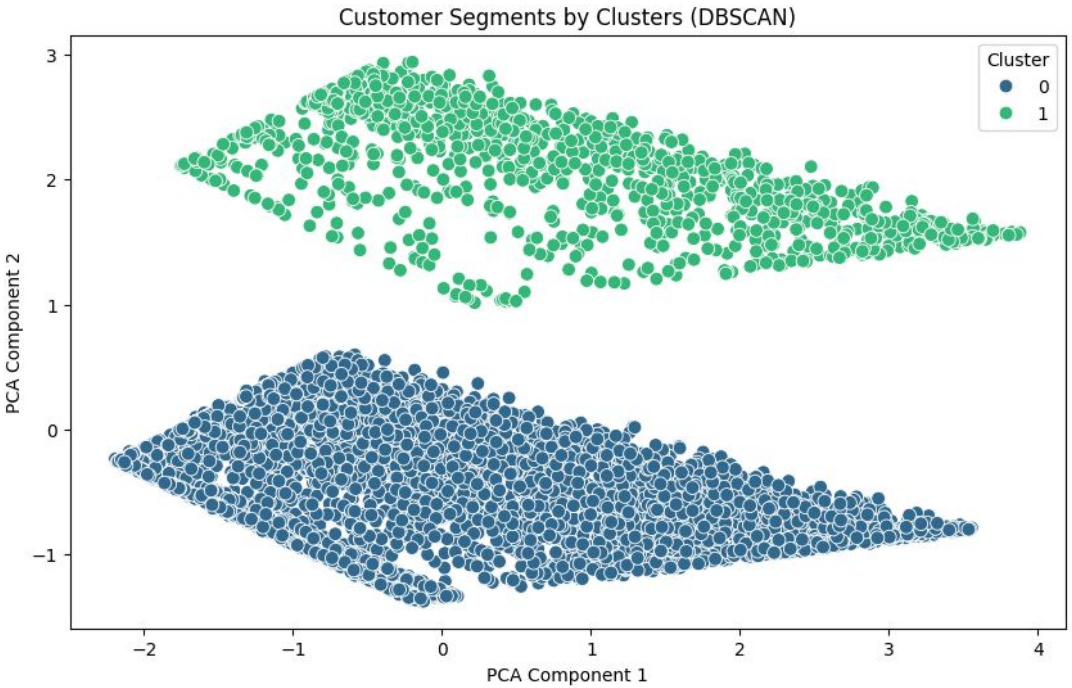
DBSCAN Clustering:





TEST CASE 02:

DBSCAN Clustering:

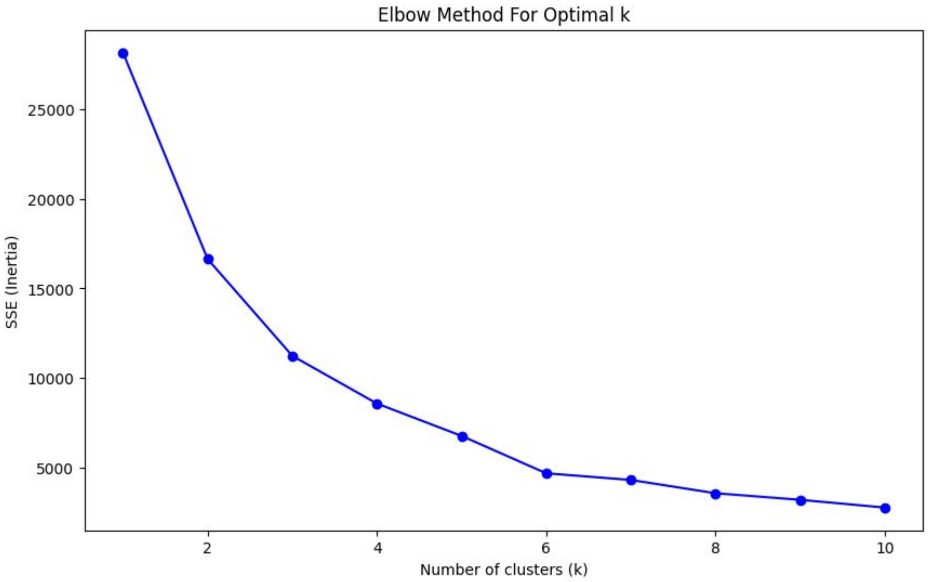


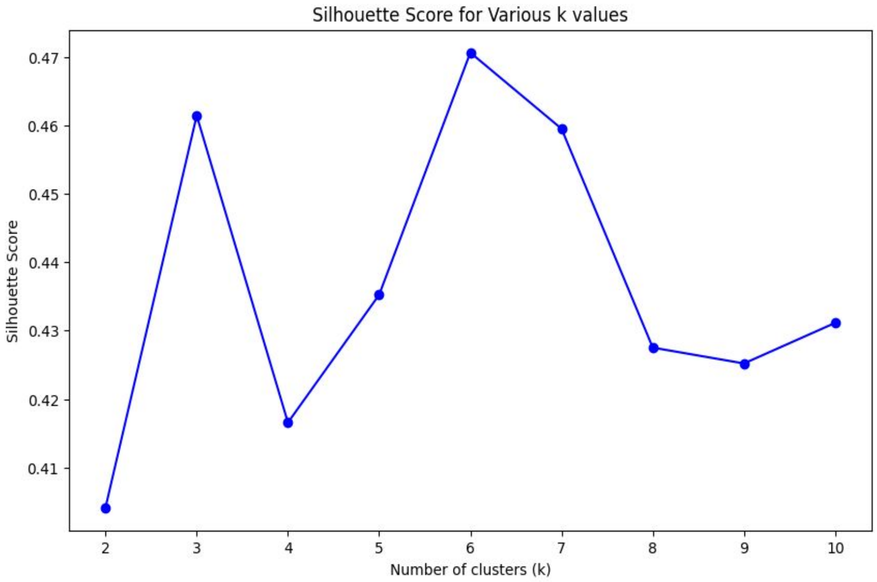
TEST CASE 03:

Metrics	K-Means Clustering	Agglomerative Clustering	DBSCAN Clustering
No. of Components	4	4	4
Silhouette Score	0.470	0.428	0.409
Calinski-Harabasz Index	7067.291	4938.750	2536.510
Davies-Bouldin Score	0.712	0.875	1.110
Avg Intra-Cluster Distance	0.782	1.238	1.549

TEST CASE 02:

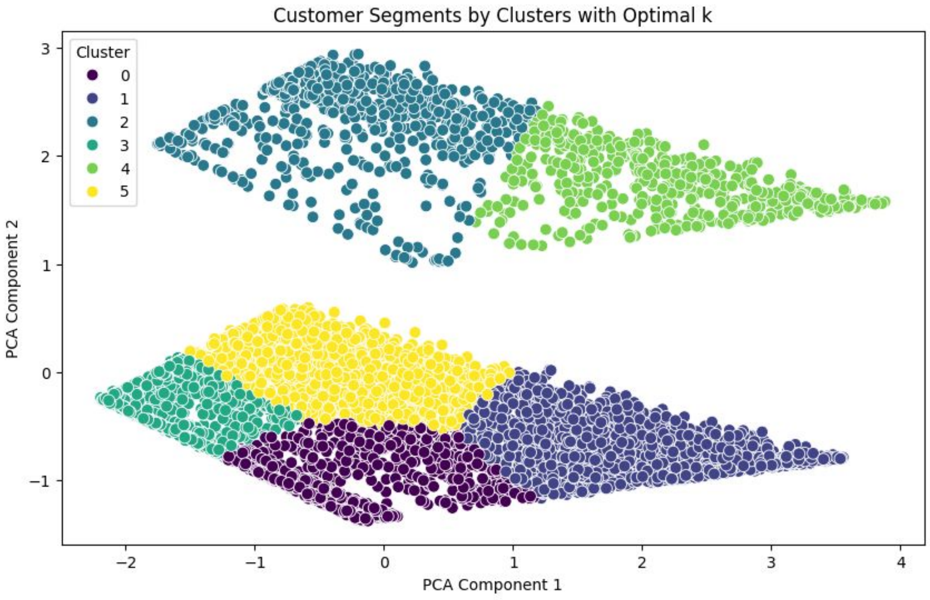
K-means Clustering:

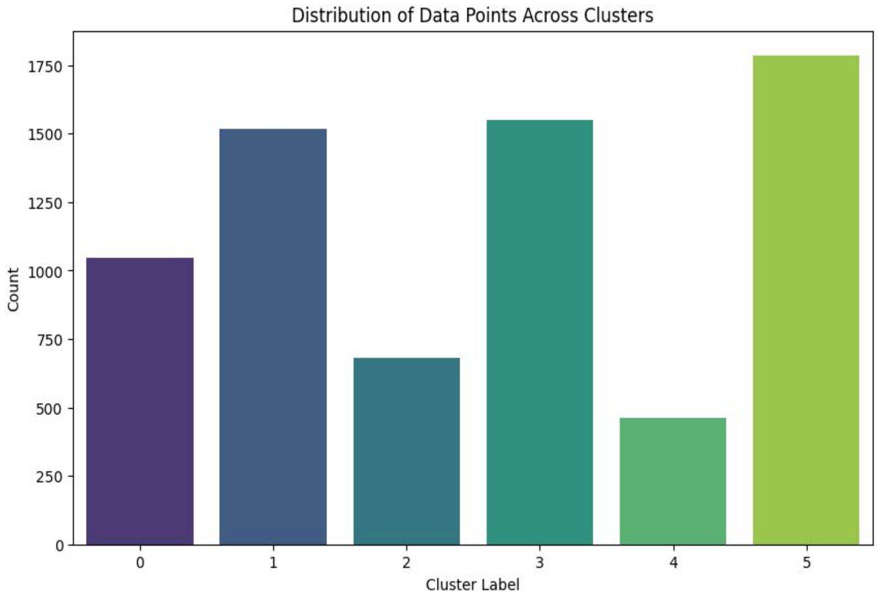




TEST CASE 02:

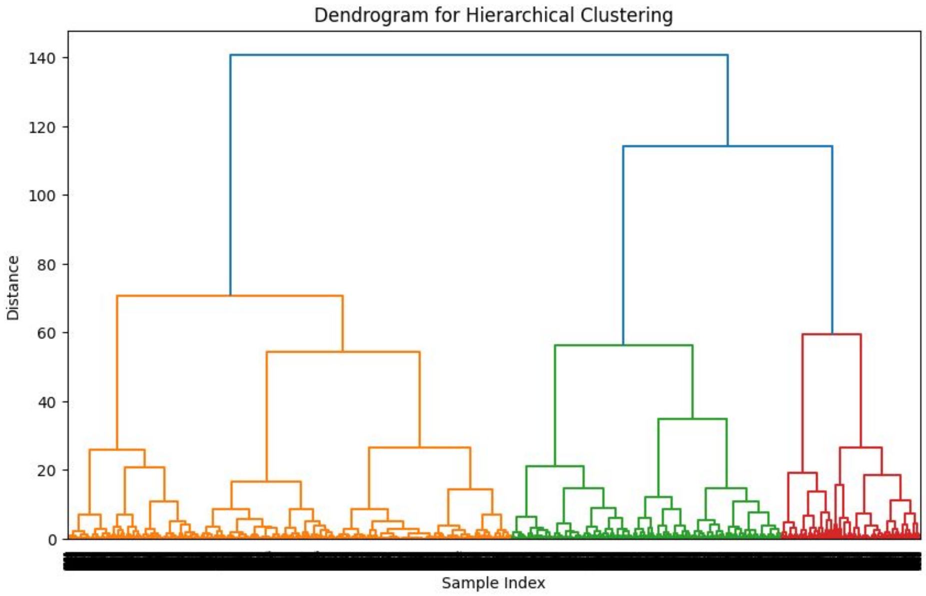
K-means Clustering:

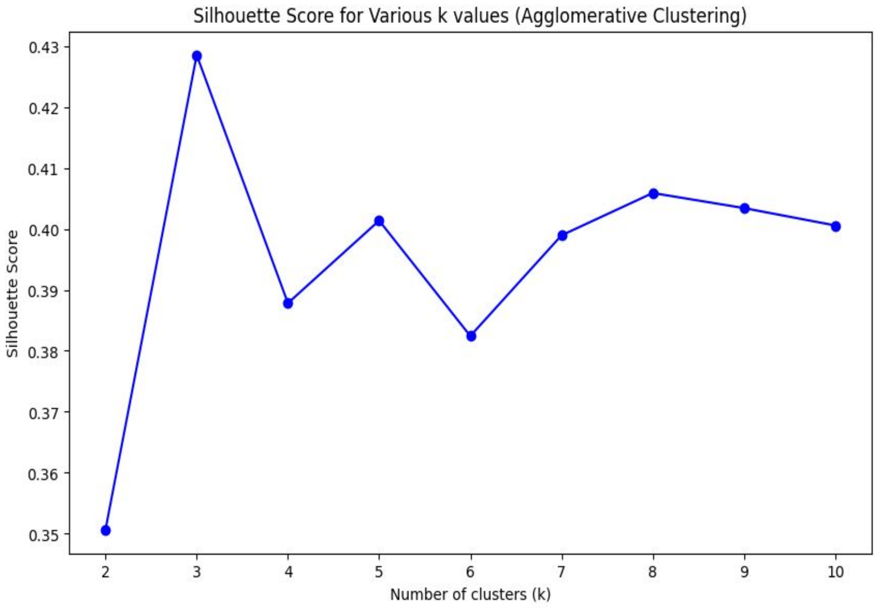




TEST CASE 02:

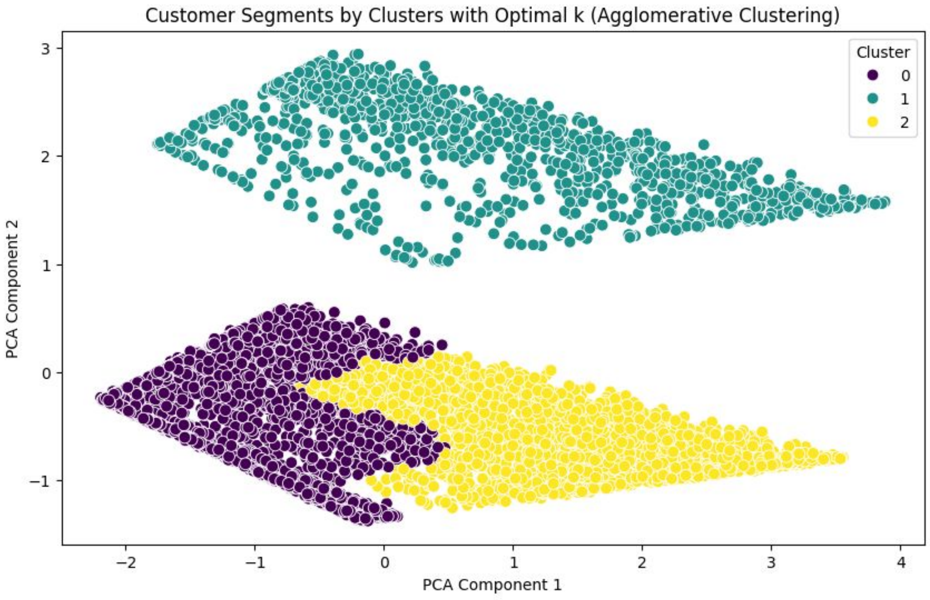
Agglomerative Clustering:

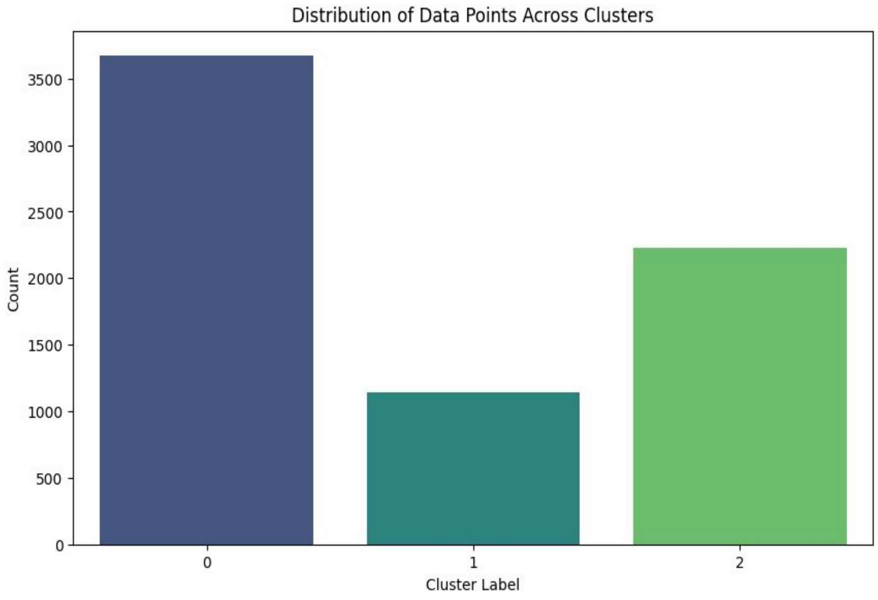




TEST CASE 02:

Agglomerative Clustering:

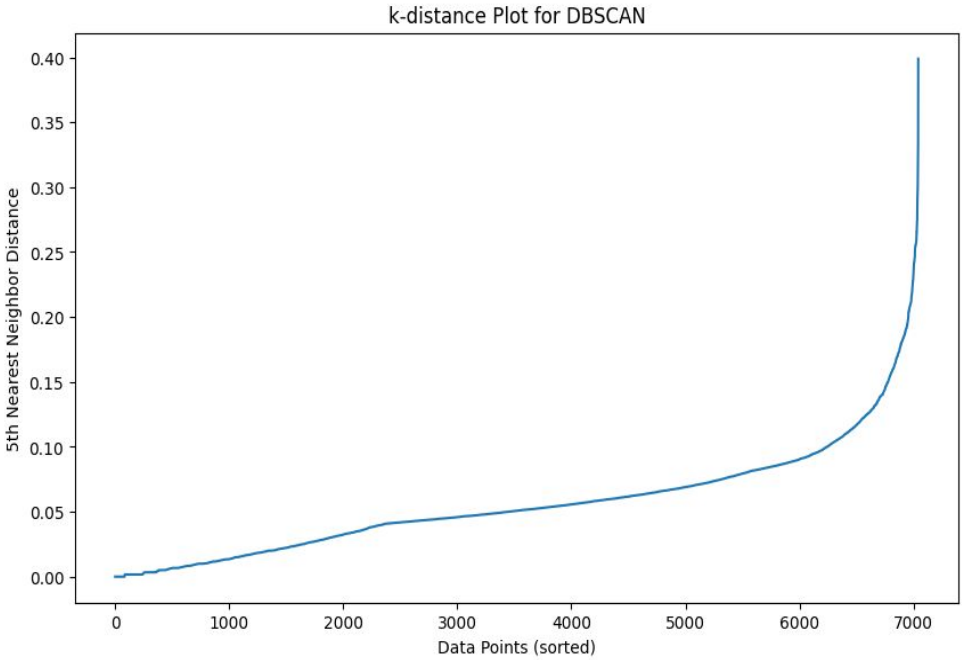


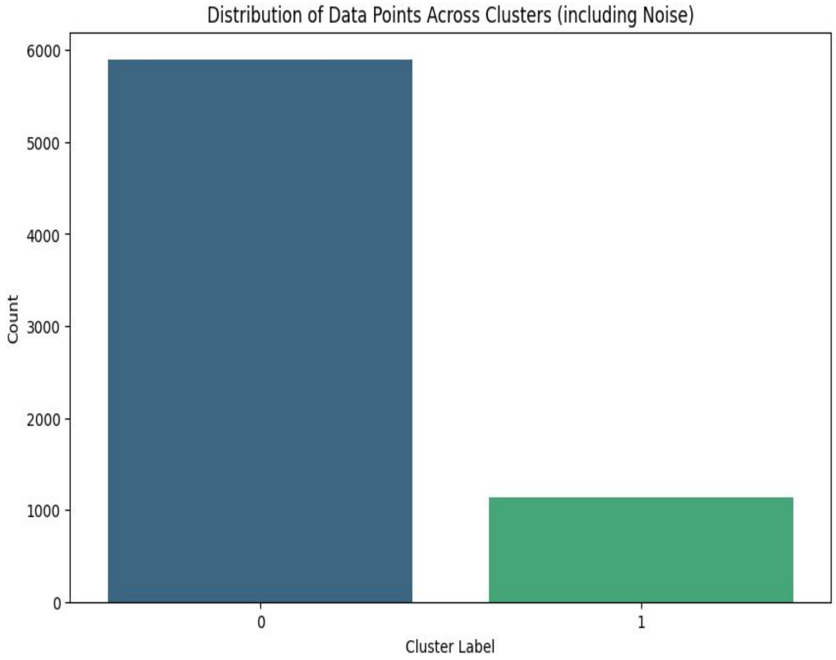


KLE TECH. KLE Technological University
Creating Value
Leveraging Knowledge

TEST CASE 02:

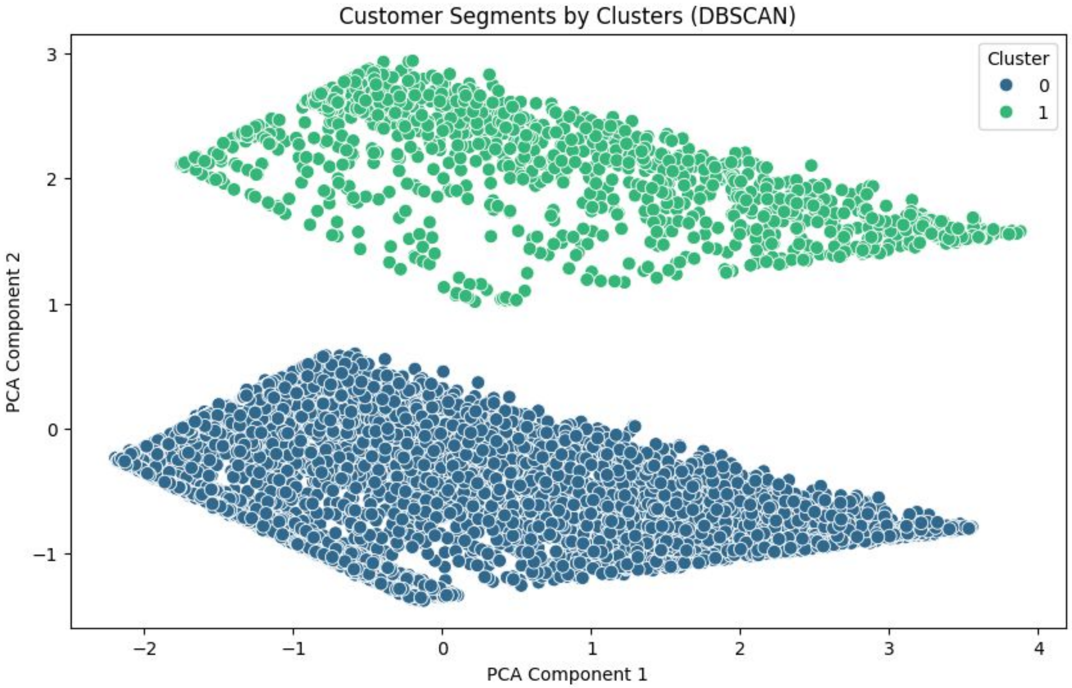
DBSCAN Clustering:





TEST CASE 02:

DBSCAN Clustering:



K-Means vs Agglomerative vs DBSCAN Clustering:

Aspect	K-Means	Agglomerative	DBSCAN
Basic Concept	Partitions data into k-clusters by minimizing within-cluster variance.	Builds a hierarchy of clusters by merging or splitting based on similarity.	Groups points in dense regions and treats sparse regions as noise.
Cluster Shape and Flexibility	Works best for spherical or circular clusters, struggles with irregular shapes.	Can model complex cluster shapes but computationally expensive for large data.	Handles clusters of arbitrary shapes; excels with non-convex clusters and noise.
Data Assumptions	Assumes clusters are spherical and similar in size.	Can handle clusters of varying sizes and densities.	Assumes clusters are dense regions separated by sparse regions, with no predefined structure.
Interpretability	Easy to interpret and visualize for low-dimensional data.	Produces a dendrogram for hierarchical relationships but can be complex for big data.	Flexible but harder to interpret due to parameter tuning and noise handling.

K-Means vs Agglomerative vs DBSCAN Clustering:

Aspect	K-Means	Agglomerative	DBSCAN
Scalability	Highly scalable; suitable for large datasets.	Less scalable for large datasets due to high computational cost of building hierarchies.	Moderately scalable; performance depends on data density and parameter settings.
Handling outliers	Poor handling of outliers; sensitive to noise in data.	Can capture outliers as separate clusters if hierarchical distances permit.	Excellent at identifying outliers as noise during clustering.
Application in marketing	Effective for segmenting customers into predefined groups (e.g., high, medium, low spenders).	Useful for identifying nested or hierarchical customer segments.	Ideal for detecting unique customer behaviors, niche markets, and noise (e.g., rare users).

CONCLUSION:

This project compared K-Means, Agglomerative Clustering, and DBSCAN for customer segmentation in marketing campaigns. **DBSCAN** outperformed the other algorithms in identifying complex, non-linear clusters and handling outliers, making it ideal for datasets with irregular cluster shapes and noise. K-Means was effective for segmenting customers into predefined groups but struggled with irregular cluster structures. Agglomerative Clustering demonstrated flexibility in hierarchical relationships but was computationally intensive for larger datasets. The analysis involved pre-processing customer data, applying clustering algorithms, evaluating performance using metrics like Silhouette Score and Davies-Bouldin Index, and visualizing results. Despite DBSCAN's sensitivity to parameter tuning, its ability to uncover nuanced customer patterns provides valuable insights for targeted marketing strategies.

FUTURE WORK:

Future research should focus on optimizing the computational efficiency of clustering algorithms, particularly DBSCAN, to handle large-scale customer datasets more effectively. Additionally, experimenting with hybrid clustering techniques by combining K-Means and DBSCAN could improve segmentation accuracy and robustness for marketing applications. Expanding the dataset to include more customer attributes such as geographic and behavioral data could yield deeper insights into segmentation patterns. Developing a real-time clustering system for dynamic customer segmentation in ongoing campaigns can enhance responsiveness and personalization. These efforts will further refine customer targeting strategies and support better decision-making for marketing teams.

ACKNOWLEDGEMENT:

I would like to express my sincere gratitude and a special thanks to my supervisor for their invaluable guidance and support throughout the process. In addition to that, I would like to thank my University and all those who contributed to the completion of the project and all the research papers. Lastly, I extend my appreciation to the co-authors for their insightful discussions and feedback.

THANK YOU…