CUSTOMER SEGMENTATION USING DBSCAN ALGORITHM

A Project Report Submitted to

Rajiv Gandhi University of Knowledge Technologies

SRIKAKULAM

In partial fulfillment of the requirements for the

Award of the Degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

3rd year B. Tech 2nd semester

Submitted by

P. Prameela (S180048)

J. Sujay (S180626)

P. Kumari (S180192)

Under the Esteemed Guidance of

Mr. T. Narasimhappadu, M. Tech, (Ph. D),

Assistant Professor

Department of Computer Science and Engineering,

RGUKT-SRIKAKULAM, ETCHERLA.

CERTIFICATE

This is to certify that the thesis work entitled "Customer Segmentation using DBSCAN Algorithm" was successfully completed by P. Prameela (S180048), J. Sujay (S180626), and P. Kumari (S180192) in partial fulfilment of the requirements for the Minor Project in Computer Science and Engineering of Rajiv Gandhi University of Knowledge Technologies under my guidance and output of the work carried out is satisfactory.

The report has not been submitted previously in part or in full to this or any other University or Institution to award any degree or diploma.

Mr. T. Narasimhappadu, Asst. Prof	**Mr. N. Sesha Kumar, Asst. Prof**
Project Internal Guide,	**Head of the Department,**
Department of CSE,	**Department of CSE,**
RGUKT- SKLM	**RGUKT- SKLM**

BONAFIDE CERTIFICATE

Certified that this project work titled "Customer Segmentation using DBSCAN Algorithm" is the Bonafide work of P. Prameela (S180048), J. Sujay (S180626) and P. Kumari (S180192), who carried out the work under my supervision, and submitted in partial fulfillment of the requirements for the award of the degree, BACHELOR OF TECHNOLOGY, during the year 2022-2023.

Mr. T. Narasimhappadu, Asst. Prof
Project Internal Guide,
Department of CSE,
RGUKT-SKLM

Mr. N. Sesha Kumar, Asst. Prof
Head of the Department,
Department of CSE,
RGUKT-SKLM

DECLARATION

We hereby declare that this thesis entitled “Customer Segmentation using DBSCAN Algorithm” is carried out by us during the academic year 2022-2023 in partial fulfilment of the requirements for the Minor Project in **Computer Science and Engineering.**

We further declare that this dissertation has not been submitted elsewhere for any Degree. The matter embodied in this dissertation report has not been submitted elsewhere for any other degree.

Date:

Place:

Project Associate

P. Prameela (S180048)

J. Sujay (S180626)

P. Kumari (S180192)

ACKNOWLEDGEMENT

We would like to articulate my profound gratitude and indebtedness to my project guide **Mr. T. Narasimhappadu (Asst. Prof),** Assistant Professor who has always been a constant motivation and guiding factor throughout the project time. It has been a great pleasure for me to get an opportunity to work under his guidance and complete the thesis work successfully.

We wish to extend my sincere thanks to **Mr. N. Sesha Kumar (Asst. Prof)**, Head of the Computer Science and Engineering Department, for his constant encouragement throughout the project. We also grateful to other members of the department; without their support my work would have been carried out so successfully. We thank one and all who have rendered help to me directly or indirectly in the completion of our thesis work.

Project Associate

P. Prameela (S180048)

J. Sujay (S180626)

P. Kumari (S180192)

ABSTRACT

This project aims to perform customer segmentation based on the products they have bought and their shopping behavior in a mall. The objective of this project is to identify distinct customer groups and understand their characteristics to develop targeted marketing strategies and increase business income. Applying DBSCAN algorithm as one of the density based algorithms results in a meaningful customer segmentation. It aims to create a relationship with the most profitable customers by designing the most appropriate marketing strategy. The results obtained from this project can help the mall to personalize marketing efforts and optimize promotional campaigns to target specific customer groups effectively. In addition, this project proposes to offer personalized incentives, the mall aims to increase customer loyalty, attract new customers, and ultimately increase revenue.

Keywords: DBSCAN, Marketing strategies.

CONTENTS

CERTIFICATE ii
BONAFIDE CERTIFICATE iii
DECLARATION iv
ACKNOWLEDGEMENT v
ABSTRACT vi
CONTENTS vii
LIST OF FIGURES viii
CHAPTER- 1 1
1. INTRODUCTION 1
1.1 Introduction 1
1.2 Problem Statement 1
1.3 Objectives 2
1.4 Goals 2
1.5 Scope 2
1.6 Applications 3
1.7 Limitations 3
CHAPTER- 2 4
LITERATURE SURVEY 4
2.1 Collecting Information 4
2.2 Study 4
2.3 Summary 5
CHAPTER- 3 6
SYSTEM ANALYSIS 6
3.1 Existing System 6
3.2 Disadvantages 6
3.3 Proposed System 7
3.4 Advantages 7
3.5 System Requirements 8
CHAPTER- 4 9
METHODOLOGY 9
4.1 Data Collection 9

4.2 Exploratory Data Analysis 9
4.3 Algorithm Used 11
CHAPTER- 5 13
IMPLEMENTATION 13
5.1 Results and Analysis 13
CONCLUSION 15
SOURCE CODE 16
APPENDIX 22
REFERENCES 22

LIST OF FIGURES

Fig. 1. Histogram - Male vs Female customers. 9
Fig. 2 Distribution plot of Annual Income 10
Fig. 3. Distribution Plot of Age 10
Fig. 4. Distribution plot of Spending Score 11
Fig. 5. Clusters created by DBSCAN 12
Fig. 6. Clusters based on Spending Score and Annual income 13
Fig. 7. Clusters based on Spending Score and Age 14

CHAPTER- 1
1. INTRODUCTION

1.1 Introduction

Customer segmentation is a vital strategy that businesses use to divide their customer base into distinct groups or segments. By grouping customers with similar characteristics together, businesses can better understand their customers and tailor their marketing efforts to meet their unique needs and preferences. The primary goal of customer segmentation is to gain insights into different customer groups and develop targeted marketing strategies. This approach allows businesses to personalize their messaging, products, and services, resulting in increased customer satisfaction and engagement. Customer segmentation offers several advantages. It enables businesses to allocate their resources effectively by focusing on high-value customer segments. It also enhances customer retention and loyalty by providing personalized experiences that resonate with specific customer groups. Moreover, customer segmentation gives businesses a competitive edge by enabling them to differentiate their offerings in the market.

1.2 Problem Statement

In today's highly competitive business landscape, understanding customers and delivering personalized experiences is critical for success. One key challenge faced by businesses is effectively segmenting their customer base to tailor marketing campaigns and strategies for maximum impact. The primary issue is the lack of a comprehensive understanding of customers' diverse characteristics, needs, and preferences. Without proper segmentation, businesses struggle to identify distinct customer groups and fail to deliver targeted messages, products, and promotions. The solution to this problem would enable businesses to gain deep insights into customer segments, understand their unique characteristics and preferences, and design targeted marketing strategies.

1.3 Objectives

- To provide an overview of customer segmentation and its significance in marketing strategy optimization.
- To explain the different methods and techniques used for customer segmentation, such as demographic, behavioural, and psychographic segmentation.
- To showcase the practical implementation of customer segmentation using real-world examples and case studies.
- To explain how DBSCAN Algorithm plays a crucial role in customer segmentation.

1.4 Goals

- Understand the concept of customer segmentation
- To evaluate the impact of customer segmentation on marketing campaign performance, customer engagement, and business outcomes.
- Demonstrate the benefits of customer segmentation
- Showcase real-world applications and case studies

1.5 Scope

The scope of this project is to implement customer segmentation techniques for marketing strategy optimization. It includes gathering and preparing customer data, applying segmentation algorithms, and analyzing the results to gain insights into customer behaviour and preferences. The scope encompasses evaluating the impact of customer segmentation on marketing campaign effectiveness and key performance indicators such as customer acquisition and retention. Additionally, the project aims to provide practical recommendations for businesses on effectively implementing customer segmentation strategies. While related areas such as data pre-processing and algorithm selection may be touched upon, the primary focus is on the implementation and evaluation of customer segmentation techniques within the context of marketing strategy.

1.6 Applications

- Product Development and Customization
- Customer Retention and Loyalty Programs
- Customer Service and Support
- Targeted Marketing Campaigns
- Market Expansion and Target Market Identification
- Enhance Customer experiences

1.7 Limitations

- Dynamic Nature of Customers
- Data Availability and Quality
- Overlapping Segment Characteristics
- Cost and Complexity
- Ethical Considerations

CHAPTER- 2
LITERATURE SURVEY

2.1 Collecting Information

We have gathered information from various online resources and previous models which are existing. Collected required data sheet which contains major attributes in it from online resources.

2.2 Study

We have studied IEEE papers and online resources to gain insights into customer segmentation and its application in marketing strategy optimization. Two IEEE papers we have referred to are:

"Maximizing Strategy in Customer Segmentation Using Different Clustering Techniques": This paper focuses on the application of various clustering techniques in customer segmentation to maximize marketing strategy effectiveness. It explores different clustering algorithms, such as k-means, hierarchical clustering, and DBSCAN, and evaluates their performance in segmenting customers based on their characteristics and behaviours. The paper provides valuable insights into the benefits and challenges of using different clustering techniques for customer segmentation.

"A Cluster-based Analysis for Targeting Potential Customers in a Real-world Marketing System": This paper presents a cluster-based analysis approach for identifying potential customers in a real-world marketing system. It emphasizes the importance of segmentation in targeting customers with specific characteristics and preferences. The paper proposes a method that combines clustering algorithms with predictive analytics to identify high-potential customer segments for targeted marketing campaigns. It showcases the practical application of customer segmentation for improving marketing outcomes.

In addition to these IEEE papers, we have also referred to various online resources, including industry reports, research articles, and reputable websites. These resources

have provided us with a comprehensive understanding of the concepts, techniques, and best practices related to customer segmentation.

2.3 Summary

Our project focuses on Customer Segmentation and its impact on optimizing marketing strategies. Through a comprehensive literature survey of IEEE papers and online resources, we gained valuable insights into customer segmentation and its applications. We identified major benefits, such as targeted marketing campaigns, product development, customer retention, and pricing strategies. The project aims to provide practical recommendations for businesses on effectively implementing customer segmentation. Regular updates and adaptations to segmentation approaches are emphasized.

Overall, our project aims to demonstrate how customer segmentation enhances marketing strategies, improves customer satisfaction, and drives business growth. By leveraging customer segmentation effectively, businesses can gain a competitive advantage and build long-term customer relationships.

CHAPTER- 3
SYSTEM ANALYSIS

3.1 Existing System

- One of the most common and widely used systems for customer segmentation is RFM (Recency, Frequency, and Monetary) analysis.
- This method segments customers based on three key metrics: how recently they have made a purchase (Recency), how frequently they make purchases (Frequency), and how much money they spend (Monetary).
- By analysing these three factors, businesses can identify their most valuable customers and tailor their marketing strategies to suit their needs.

3.2 Disadvantages

While RFM analysis is a useful and straightforward method for customer segmentation, it does have some limitations. Here are a few disadvantages of RFM analysis:

1. Limited Attribute Consideration: RFM analysis primarily focuses on Recency, Frequency, and Monetary Value as the key factors for segmentation. It may overlook other important customer attributes, such as demographics, preferences, or behavioral patterns, which can provide deeper insights into customer behavior and preferences.
2. Lack of Flexibility: RFM analysis relies on predefined scoring criteria and segment boundaries. It lacks the adaptability and flexibility
3. Inability to Handle Non-linear Relationships: RFM analysis assumes linear relationships between the RFM components and customer behavior. However, customer behavior is often influenced by complex and non-linear factors.
4. Handling Noisy Data and Outliers: RFM analysis may struggle with handling noisy or incomplete data and outliers.

3.3 Proposed System

- The proposed system for customer segmentation is the use of Machine Learning algorithms to analyse customer data and identify patterns that might not be immediately apparent using traditional methods.
- For example, a business might use clustering algorithms to group customers based on their purchasing behaviour and other data points. The data points are grouped into clusters based on the density of the points will help the company understand how the customers differ.
- By offering personalized incentives, the company aims to increase customer loyalty, attract new customers, and ultimately increase revenue.

3.4 Advantages

- Enhanced Accuracy and Predictive Power: ML algorithms can uncover complex patterns and relationships within customer data that may not be easily identifiable through RFM analysis.
- Increased Flexibility and Adaptability: ML algorithms can handle a wide range of data types, attributes, and variables.ML models can incorporate multiple factors beyond RFM, such as demographic data, browsing behavior, social media interactions, and more, leading to more comprehensive and nuanced customer segments.
- Personalized Recommendations and Experiences: ML algorithms excel at personalization by leveraging individual-level data. They can analyze customer preferences, purchase history, and behavior to deliver tailored recommendations
- Identification of Non-linear Relationships: ML techniques can capture non-linear relationships and interactions among variables. This allows for a more accurate understanding of customer behavior and the identification of hidden patterns that influence purchasing decisions.
- Handling Large and Complex Datasets: ML algorithms are designed to handle large volumes of data efficiently.

- Anomaly and Outlier Detection: ML algorithms can identify anomalies or outliers in customer behavior that may require special attention.

3.5 System Requirements

- Software Requirements- Python IDLE/Google Colab/Jupyter Notebook.
- ML libraries and Segmentation Algorithms.
- Operating System- Windows, MAC OS.
- Hardware requirements- RAM 4GB or more.

CHAPTER- 4
METHODOLOGY

4.1 Data Collection

The dataset we have taken is from a mall authority which summarizes the behavior of 200 active mall customers over the last 3 months. The dataset is from Kaggle. Features include in the dataset are:

- CustomerID: Customer's unique ID.
- Gender – categorical-binary: Gender of the customer (Male & Female).
- Age - numerical: Customer's age.
- Annual Income (k$) - numerical: Annual income of the customer.
- Spending Score (1-100) - numerical: A score (out of 100) given to a customer by mall authorities based on money spent and behavior.

4.2 Exploratory Data Analysis

This section examines the dataset statistically. It's a crucial step in any analysis because it clarifies the data. The dataset contains no missing data. This section covers distributions.

- Distributions: Here, in Fig. 1. we will explore numerical variable distributions. Data will be stratified by gender, the only categorical variable.
- In Fig. 2. Distribution plot with respect to Annual Income (k$) is shown.

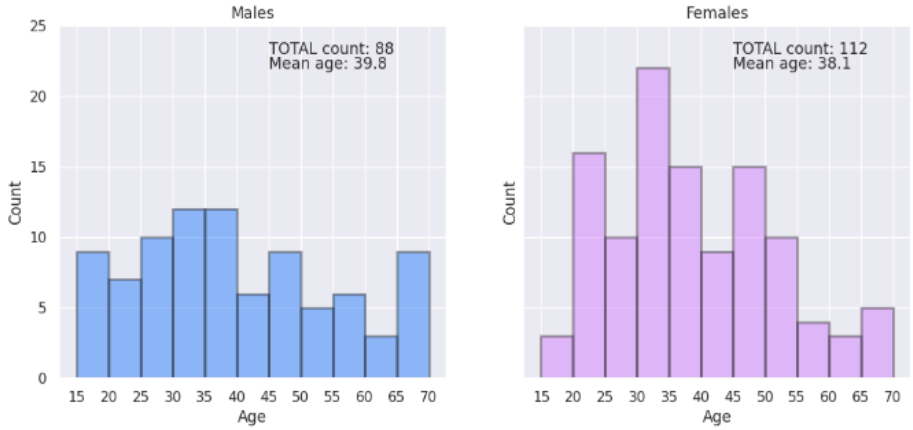

Fig. 1. Histogram - Male vs Female customers.

- Here most of the annual income falls between 50k and 80k.

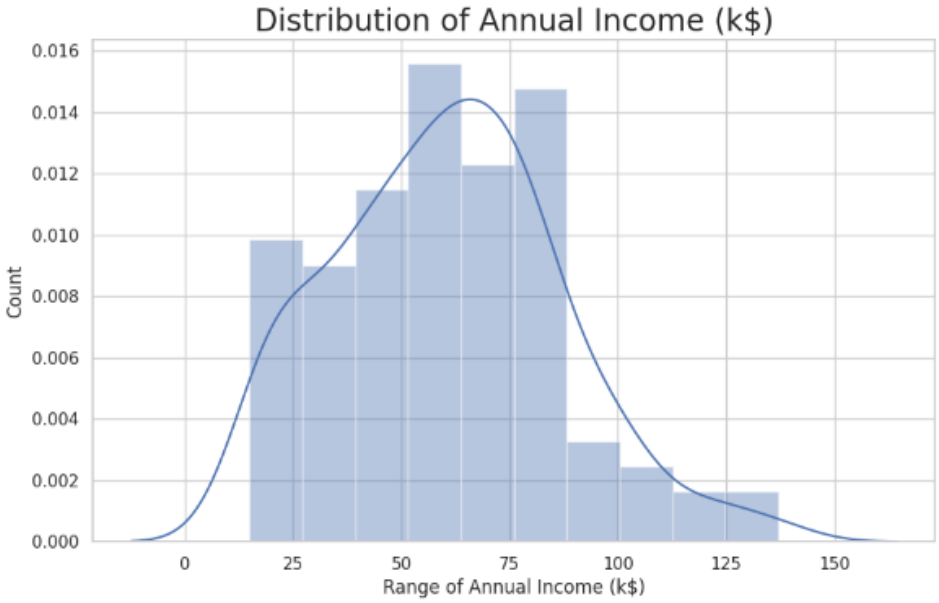

Fig. 2 Distribution plot of Annual Income

- In Fig. 3. Distribution plot based on Age is shown.
- There are customers of wide variety of ages.
- Age group near 30-40 have the highest density.

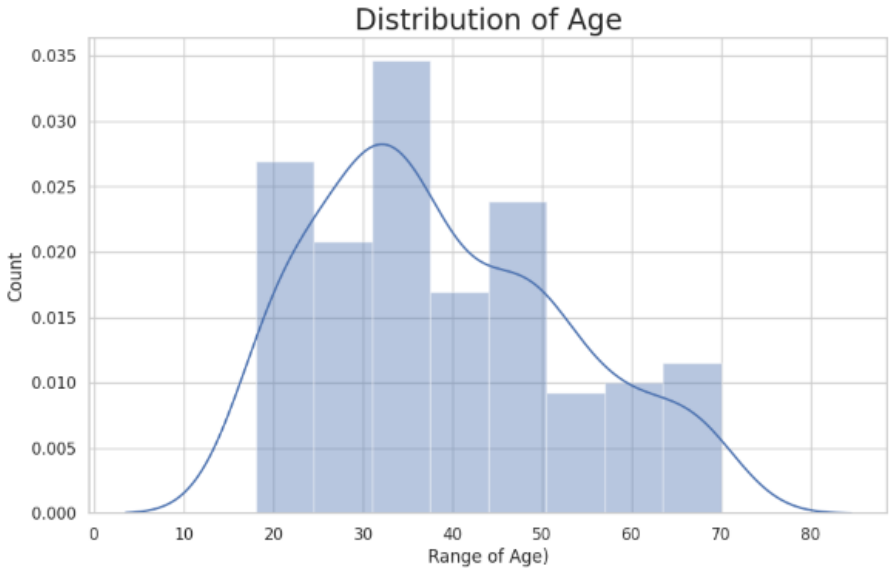

Fig. 3. Distribution Plot of Age

- In Fig. 4. Distribution plot based on spending score is shown.
- The maximum Spending Score is in the range of 40 to 60.

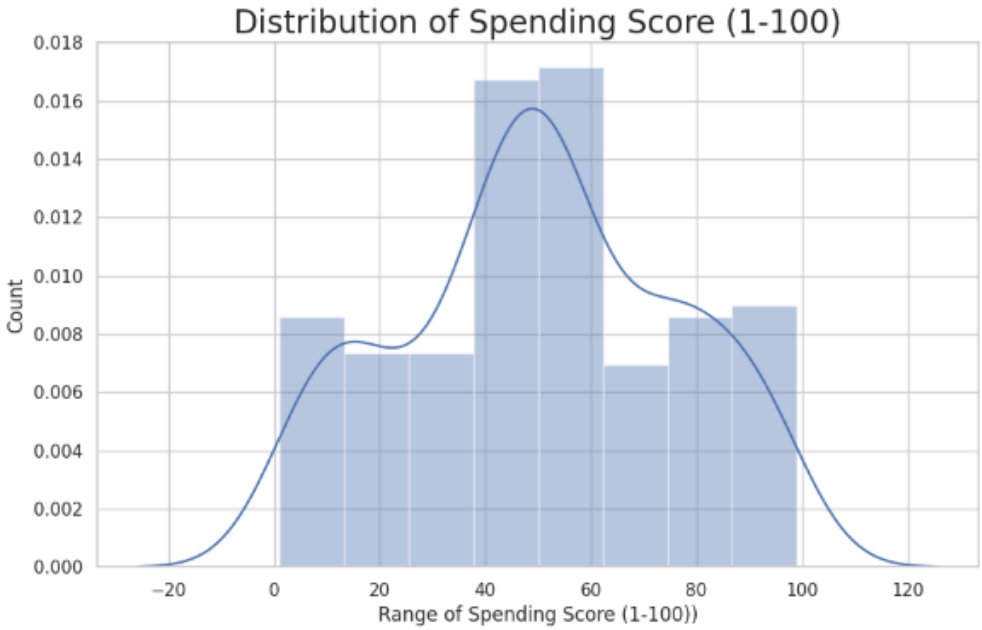

Fig. 4. Distribution plot of Spending Score

Analyzing the distributions of attributes provides valuable insights into the characteristics and behaviors of the customer base. It helps in understanding the composition of different segments, identifying patterns, and guiding decision-making processes for targeted marketing strategies, product development, and customer engagement initiatives.

4.3 Algorithm Used

We used DBSCAN algorithm for customer segmentation. DBSCAN stands for Density-Based Spatial Clustering of Applications with Noise. It groups ‘densely grouped’ data points into a single cluster. It can identify clusters in large spatial datasets by looking at the local density of the data points. The most exciting feature of DBSCAN clustering is that it is robust to outliers. It also does not require the number of clusters to be told beforehand, unlike K-Means, where we have to specify the number of centroids.

DBSCAN requires only two parameters: epsilon and minPoints. **Epsilon** is the radius of the circle to be created around each data point to check the density and **minPoints** is the minimum number of data points required inside that circle for that data point to be classified as a **Core** point.

DBSCAN creates a circle of epsilon radius around every data point and classifies them into **Core** point, **Border** point, and **Noise**. A data point is a **Core** point if the circle around it contains at least 'minPoints' number of points. If the number of points is less than minPoints, then it is classified as **Border** Point, and if there are no other data points around any data point within epsilon radius, then it treated as **Noise**. Fig. 5. Shows the illustration of DBSCAN clusters.

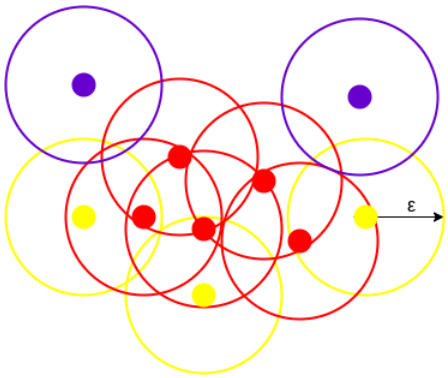

Fig. 5. Clusters created by DBSCAN

CHAPTER- 5
IMPLEMENTATION

5.1 Results and Analysis

The Fig. 6. shows the results of using the DBSCAN algorithm to segment customers based on their Spending score and Annual income. Four distinct clusters were identified, each representing a different customer segment. Outliers are also displayed separately. Annual income (k$) is taken on x-axis and Spending Score (1-100) on y-axis.

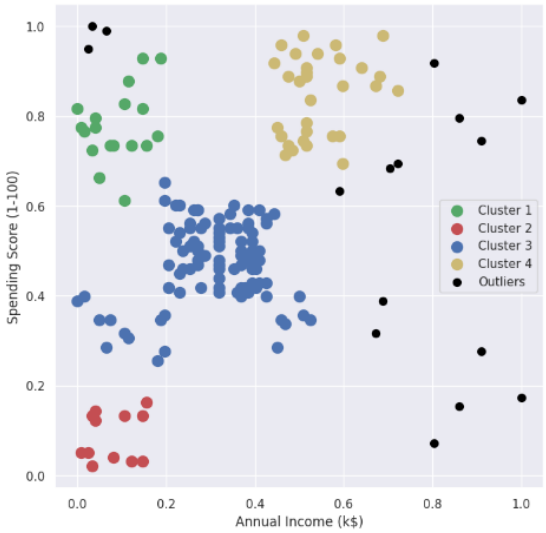

Fig. 6. Clusters based on Spending Score and Annual income

The Fig. 7. shows the results of using the DBSCAN algorithm to segment customers based on their **Spending score and Age**. Five distinct clusters were identified, each representing a different customer segment. Outliers are also displayed separately. Age is taken on x-axis and Spending Score (1-100) on y-axis.

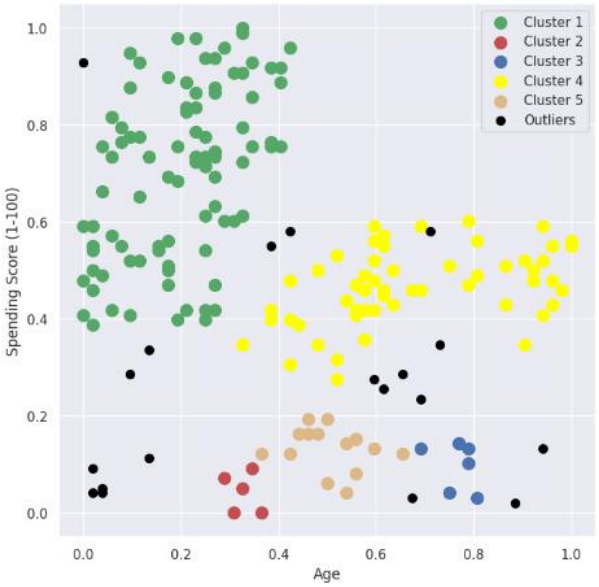

Fig. 7. Clusters based on Spending Score and Age

CONCLUSION

It’s not wise to serve all customers with the same product model, email, text message campaign, or ad. Customers have different needs. A one-size-for-all approach to business will generally result in less engagement, lower-click through rates, and ultimately fewer sales. Customer segmentation is the cure for this problem. Customer segmentation is well-liked because it makes marketing and sales more effective. This is so that you can have a better grasp on what your customers’ desires and needs are. This has even greater financial implications, and using efficient customer segmentation will help you raise client lifetime value. This implies that they will spend more money and stay longer. You can make customers more loyal by getting to know them better so you can target them better.

SOURCE CODE

```
#importing necessary libraries
import pandas as pnd
import numpy as nmpy
import matplotlib.pyplot as matplotl
import seaborn as seabn
from sklearn.cluster import DBSCAN
from sklearn.cluster import KMeans
import warnings
warnings.filterwarnings('ignore')
mall_cust_data=pnd.read_csv('Mall_Customers.csv')
mall_cust_data.head()
mall_cust_data.describe()
mall_cust_data.shape
mall_cust_data.info()
mall_cust_data.isnull().sum()
# Distribution plot for Annual Income (k$)
matplotl.figure(figsize=(8,4))
seabn.set(style='darkgrid')
seabn.distplot(mall_cust_data['Annual Income (k$)'])
matplotl.title('Plot for Annual income',fontsize=20)
matplotl.xlabel('Annual income')
matplotl.ylabel('Count of Annual income')

"""In the above plot we can see that the income is mostly between 50k and 80k"""
# Distrbution plot for Age
matplotl.figure(figsize=(9,5))
seabn.set(style='darkgrid')
seabn.distplot(mall_cust_data['Age'])
matplotl.title('Plot for Age',fontsize=20)
matplotl.xlabel('Age')
```

```
matplotl.ylabel('Count of age ')
"""
*   Age group near 30-40 have the highest density
"""
# Distribution plot for Spending Score (1-100)
matplotl.figure(figsize=(9,6))
seabn.set(style='darkgrid')
seabn.distplot(mall_cust_data['Spending Score (1-100)'])
matplotl.title('Plot for Spending score (1-100)',fontsize=20)
matplotl.xlabel('Spending score (1-100)')
matplotl.ylabel('Count of Spending score')
"""*   We can see from the above plot that the maximum spending score is in the range of 40 to 60"""
#Bar plot for gender distribution
gender_data=mall_cust_data.Gender.value_counts()
seabn.set_style('darkgrid')
matplotl.figure(figsize=(9,4))
seabn.barplot(x=gender_data.index, y=gender_data.values)
gen_male_age = mall_cust_data[mall_cust_data['Gender']=='Male']['Age'] # subset with males age
gen_female_age = mall_cust_data[mall_cust_data['Gender']=='Female']['Age'] # subset with females age
gen_age_bins = range(15,75,5)
#hist plot for men
fig2, (ax1, ax2) = matplotl.subplots(1, 2, figsize=(10,5), sharey=True)
seabn.distplot(gen_male_age, bins=gen_age_bins, kde=False, color='blue', ax=ax1, hist_kws=dict(edgecolor="k", linewidth=3))
ax1.set_xticks(gen_age_bins)
ax1.set_ylim(top=25)
ax1.set_title('Males')
ax1.set_ylabel('Count')
ax1.text(38,18, "Total count: {}".format(gen_male_age.count()))
```

```
ax1.text(38,16, "Mean age: {:.1f}".format(gen_male_age.mean()))
#hist plot for women
seabn.distplot(gen_female_age, bins=gen_age_bins, kde=False, color='red', ax=ax2,
hist_kws=dict(edgecolor="k", linewidth=3))
ax2.set_xticks(gen_age_bins)
ax2.set_title('Females')
ax2.set_ylabel('Count')
ax2.text(38,18, "Total count: {}".format(gen_female_age.count()))
ax2.text(38,16, "Mean age: {:.1f}".format(gen_female_age.mean()))
matplotl.show()
"""# DBSCAN Algorithm
## Annual Income and Spending Score
"""
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
data_sp_an = mall_cust_data.iloc[:, [3,4]]
df_norm = scaler.fit_transform(data_sp_an)
DBS_clustering1 = DBSCAN(eps=0.09, min_samples=5).fit(df_norm)
DBSCAN_clustered1 = df_norm.copy()
labels1 = DBS_clustering1.labels_
labels1
unq_lbs, label_counts = nmpy.unique(labels1, return_counts=True)
# Now we can print the occurances for each label
for label, count in zip(unq_lbs, label_counts):
    print(f"Label {label}: {count} occurrences")
n_clusters1_ = len(set(labels1)) - (1 if -1 in labels1 else 0)
n_noise1_ = list(labels1).count(-1)
n_clusters1_
colors = ['g','r','b','y','burlywood','green', 'm', 'c']
matplotl.figure(figsize=(8,8))
for i in range(0 ,n_clusters1_ - 1):
```

```
    matplotl.scatter(df_norm[labels1 == i, 0], df_norm[labels1 == i, 1], s = 100, c = colors[i], label = 'Cluster ' + str(i + 1))
matplotl.scatter(df_norm[labels1 == -1, 0], df_norm[labels1 == -1, 1], s = 50, c = 'black', label = 'Outliers')
matplotl.xlabel('Annual Income (k$)')
matplotl.ylabel('Spending Score (1-100)')
matplotl.legend()
"""The above clusters can be explained as:
*   Cluster-1 --> low annual income and high spending score (Considerable)
*   Cluster-2 --> low annual income and low spending score (Sensible)
*   Cluster-3 --> medium annual income and medium spending score (Standard)
*   Cluster-4 --> high annual income and high spending score (Targeted)
"""
#Customer ids
customerIDs=mall_cust_data['CustomerID'].values
# Create a dictionary to store customer IDs for each label
label_to_customer_ids = {}
# Iterate through labels and customer IDs
for i in range(len(customerIDs)):
    customer_id = customerIDs[i]
    label = labels1[i]
    if label in label_to_customer_ids:
        label_to_customer_ids[label].append(customer_id)
    else:
        label_to_customer_ids[label] = [customer_id]
# Print the customer IDs for each label
for label, customer_ids in label_to_customer_ids.items():
    print(f"Label {label}: {customer_ids}")
"""## Spending Score and Age"""
df_age_score = mall_cust_data.iloc[:, [False, False, True, False, True]].values
scaler = MinMaxScaler()
df_age_score_scaled=scaler.fit_transform(df_age_score)
```

```
df_age_score = mall_cust_data.iloc[:, [False, False, True, False, True]]
df_norm_age = scaler.fit_transform(df_age_score)
DBS_clustering_age = DBSCAN(eps=0.08, min_samples=5).fit(df_norm_age)
DBSCAN_clustered_age = df_norm_age.copy()
labels_age = DBS_clustering_age.labels_
labels_age
unq_lbs_1, label_counts = np.unique(labels_age, return_counts=True)
# We can see the occurances of differnet labels
for label, count in zip(unq_lbs_1, label_counts):
    print(f"Label {label}: {count} occurrences")
n_clusters_age = len(set(labels_age)) - (1 if -1 in labels_age else 0)
n_noise_age_ = list(labels_age).count(-1)
n_clusters_age
colors = ['g','r','b','yellow','burlywood','green', 'm', 'c']
matplotl.figure(figsize=(8,8))
for i in range(0 ,n_clusters_age - 1):
    matplotl.scatter(df_norm_age[labels_age == i, 0], df_norm_age[labels_age == i, 1],
s = 90, c = colors[i], label = 'Cluster ' + str(i + 1))
matplotl.scatter(df_norm_age[labels_age == -1, 0], df_norm_age[labels_age == -1, 1],
s = 20, c = 'black', label = 'Outliers')
matplotl.xlabel('Age')
matplotl.ylabel('Spending Score (1-100)')
matplotl.legend()
# Create a dictionary to store customer IDs for each label
label_to_customer_ids = {}
# Iterating through labels and customer IDs
for i in range(len(customerIDs)):
    customer_id = customerIDs[i]
    label = labels_age[i]
    if label in label_to_customer_ids:
        label_to_customer_ids[label].append(customer_id)
    else:
```

```
        label_to_customer_ids[label] = [customer_id]
# Print the customer IDs for each label
for label, customer_ids in label_to_customer_ids.items():
    print(f"Label {label}: {customer_ids}")
```

APPENDIX

REFERENCES

- Vijay Choudhary, “Mall Customer Segmentation Data”, Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/vjchoudhary7/ customer-segmentation-tutorial-in-python?select=Mall Customers.csv.
- Sheikh Sharfuddin Mim, Doina Logofatu, “A Cluster-based Analysis for Targeting Potential Customers in a Real-world Marketing System”, 2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP),1-8, doi: 10.1109/ICCP56966.2022.10053985.
- Pavithra M, Ayushman Prashar, Abirami, “Maximizing Strategy in Customer Segmentation Using Different Clustering Techniques”, 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), 1-5, doi: 10.1109/SPICES52834.2022.9774200.