Project Report

On

Customer Segmentation

Submitted to D Y Patil International University, Akurdi, Pune in partial fulfilment of full-time degree

Master of Computer Applications

Submitted By:

Name: Yashshree Rasal   PRN:20230804001

Name: Shruti Hole   PRN:20230804017

Under the Guidance of

Ms. Priyanka Karale

School of Computer Science, Engineering and Applications

D Y Patil International University, Akurdi,Pune, INDIA, 411044

[Session 2023-24]

**D Y PATIL
INTERNATIONAL
UNIVERSITY**
AKURDI PUNE

CERTIFICATE

This report on (Customer Segmentation) is submitted for the partial fulfillment of project,which is part of the First Year Master of Computer Applications curriculum, under my supervision and guidance.

11/12/23

Ms. Priyanka Karale
(DYPIU Guide)

DECLARATION

I, hereby declare that the following Project which is being presented in the Project entitled as (Customer Segmentation) is an authentic documentation of my own original work to the best of my knowledge. The following Project and its report in part or whole, has not been presented or submitted by me for any purpose in any other institute or organization. Any contribution made to my work, with whom i have worked at D Y Patil International University, Akurdi, Pune, is explicitly acknowledged in the report.

Name:Yashshree Rasal   PRN No:20230804001   Signature :

Name:Shruti Hole   PRN No:20230804017   Signature :

ACKNOWLEDGEMENT

Name:Yashshree Rasal   PRN:20230804001
Name:Shruti Hole   PRN:20230804017

Abstract

A lot of customers buy products from the mall and to generate more revenue for the mall, the authorities need to attract these customers and for this large amount of capital is required. After the advertisement, the output is only around 30-40% . Hence customer segmentation comes into the picture.

Customer Segmentation is a popular application of unsupervised learning and by using this technique we'll only focus on the potential customers (customers whose probability of buying the product is very high). With this technique, the output will drastically increase to 90-95%Our project aims to build clusters of customers based on their Spending Score and Annual Income. The algorithm used in this project is K-means.

TABLE OF CONTENTS

List of Figures

1. INTRODUCTION

Customer segmentation is the process of dividing customers into groups based on their shared characteristics. By categorizing customers into meaningful segments, businesses can target their marketing campaigns more effectively, improve customer service, and develop new products and services that cater to specific customer needs.

Customer segmentation allows companies to understand their customer base better, personalize their offerings, and optimize their business strategies.

1.1. Background

1. **Diverse Customer Base:** Malls attract a varied customer demographic with distinct preferences.

2. **Data Abundance:** Malls collect extensive data, including transactions and foot traffic.

3. **Personalization Trend:** The trend towards personalized shopping experiences necessitates understanding individual customer needs.

4. **Marketing Efficiency:** Targeted marketing is more effective; segmentation allows tailored promotions.

5. **Competition in Retail:** Malls seek differentiation in a highly competitive retail landscape.

6. **Optimizing Tenant Mix:** Customer segmentation aids in optimizing tenant selection to meet diverse preferences.

7. **Resource Allocation:** Efficient allocation of resources is crucial for maximizing revenue and satisfaction.

8. **Seasonal and Trend Analysis:** Segmentation helps analyze seasonal shopping patterns and adapt to changing trends.

9. **Enhancing Customer Experience:** Understanding customer segments leads to a more tailored and enjoyable shopping experience.

10. **Operational Optimization:** K-means clustering optimizes operations based on customer demand patterns.

11. **Digital Transformation:** Malls leverage data analytics to enhance the in-person shopping experience amid digital trends.

1.2. Objectives

1. **Group Similar Customers:** Identify and group customers with similar characteristics and behaviors.

2. **Targeted Marketing:** Tailor marketing strategies to the specific needs and preferences of each customer segment.

3. **Improved Communication:** Enhance communication by using language and channels that resonate with each customer group.

4. **Enhance Product/Service Offerings:** Optimize products or services to better match the preferences of different customer segments.

5. **Optimize Pricing:** Set prices based on the varying sensitivities of different customer groups.

6. **Customer Retention:** Implement targeted strategies to retain valuable customers, especially those at risk of churning.

7. **Efficient Resource Allocation:** Allocate resources more efficiently by focusing on high-potential customer segments.

8. **Identify Market Opportunities:** Discover new market opportunities by understanding the unique needs of different customer segments.

9. **Data-Driven Decision-Making:** Make informed decisions based on insights into customer behavior and preferences.

10. **Continuous Improvement:** Regularly refine segmentation strategies based on evolving customer behaviors and market dynamics.

1.3. Purpose

The purpose of mall customer segmentation using k-means clustering is to better understand and categorize the diverse customer base within a mall. K-means clustering is a data analysis technique that groups similar data points together, and in the context of mall customer segmentation, it helps achieve several specific objectives:

1. **Targeted Marketing:** Tailor promotions and advertising to specific customer segments for more effective engagement.

2. **Optimize Operations:** Enhance store layouts, resource allocation, and inventory management based on customer preferences.

3. **Personalized Strategies:** Design promotions, loyalty programs, and pricing strategies customized for different customer segments.

4. **Enhanced Customer Experience:** Tailor services, amenities, and overall shopping experiences to meet the unique needs of various customer groups.

5. **Market Expansion and Innovation:** Identify opportunities to attract new stores or services by understanding and catering to the untapped needs of specific customer segments.

1.4. Scope

1. **Customer Loyalty Programs:** Design loyalty programs that cater to the preferences and behaviors of different customer segments, encouraging repeat visits and fostering long-term customer relationships.

2. **Enhanced Customer Service:** Provide targeted and personalized customer service by understanding the needs and expectations of different segments, leading to improved customer satisfaction.

3. **Market Basket Analysis:** Conduct market basket analysis within each segment to identify product associations and optimize product bundling, increasing overall sales and customer satisfaction.

4. **Dynamic Pricing Strategies:** Implement dynamic pricing strategies based on the price sensitivities of different customer segments, maximizing revenue and competitiveness.

5. **Expansion and Tenant Selection:** Use customer segmentation insights to guide decisions on expanding the mall and selecting tenants that align with the preferences of the local customer base, ensuring sustained foot traffic and success.

1.5. Applicability

1. **Targeted Marketing Campaigns:** K-means clustering helps identify distinct customer segments in a mall, allowing marketers to create tailored campaigns for specific groups based on their preferences and behaviors.

2. **Optimized Store Layouts:** Understanding different customer segments enables mall operators to optimize the placement of stores and attractions, improving the overall shopping experience and maximizing sales.

3. **Personalized Promotions:** By segmenting mall customers, businesses can offer personalized promotions and discounts to each group, increasing the likelihood of attracting and retaining customers.

4. **Enhanced Customer Experience:** Customer segmentation enables malls to provide a more personalized and enjoyable experience, catering to the unique needs and preferences of various customer groups.

5. **Efficient Resource Allocation:** Mall management can allocate resources such as staff and facilities more efficiently by focusing on the specific needs and demands of different customer segments, optimizing overall operational efficiency.

2.   PROJECT PLAN

2.1.   Problem Statement

Shopping malls attract a diverse range of customers with varying preferences, behaviors, and demographics. Understanding these customer segments is crucial for optimizing marketing efforts, tailoring services, and improving overall customer satisfaction. The goal is to implement a customer segmentation strategy using data-driven methods, specifically leveraging techniques like k-means clustering.

The objective is to segment mall customers into distinct groups based on their shopping patterns, preferences, and demographics. This segmentation will provide actionable insights to enhance marketing strategies, improve customer engagement, and optimize resource allocation within the mall.

*Key Challenges in Mall Customer Segmentation

1. **Diverse Customer Profiles:**
   Capturing and managing diverse demographics, including age, gender, income levels, and preferences, is a challenge.

2. **Shopping Behavior Variation:**
   Identifying and understanding varied shopping behaviors, such as frequency of visits, average spending, and preferred stores, is a challenge.

3. **Optimizing Retailer Placement:**
   Optimizing the placement of retailers within the mall to enhance the overall shopping experience and maximize sales presents a challenge.

4. **Personalized Marketing:**
   Implementing effective personalized marketing strategies requires a clear understanding of customer segments, which can be challenging.

5. **Resource Allocation:**
   Efficiently allocating resources, such as security, staff, and promotional activities, based on customer segments to enhance the overall mall experience is challenging.

2.2. Requirement Specification

1. **Data Collection:** Gather information about mall customers, including demographics, purchase history, and visit frequency.

2. **Data Preprocessing:** Load the data into the Jupyter Notebook and clean it by handling missing values and outliers.

3. **Customer Segmentation (K-means):** Utilize the K-means clustering algorithm from scikit-learn, determining the optimal number of clusters (k) using techniques like the elbow method. Apply K-means to group customers based on their characteristics.

4. **Visualization and Analysis:** Visualize customer segments using charts or plots. Explore and interpret the unique traits of each segment.

5. **Strategies and Documentation:** Propose marketing or operational strategies for each segment. Document the analysis in the Jupyter Notebook, providing explanations and recommendations.

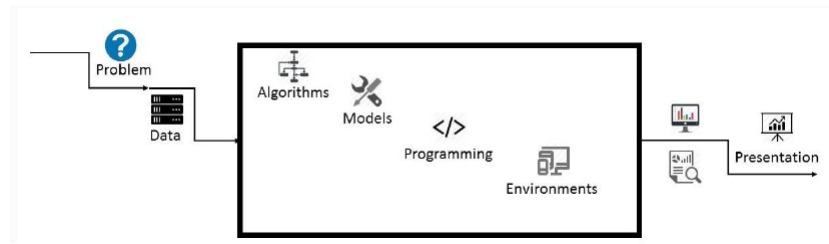# 3. PROPOSED SYSTEM AND METHODOLOGY

3.1. System Architecture

Fig. 1: System Architecture

Data: The size of the dataset is (200, 5) which is 200 rows and 5 columns. Also the dataset does not contain any NULL or NaN values.

Programming Language: Python 3.6

Environment (Libraries and Technologies): Numpy, Pandas, Matplotlib, Seaborn, Jupyter Notebook, Google Colab.

3.2. Methodology (Algorithms used)

Algorithms K-means algorithm is used in this project to analyze and form clusters of customers based on their income and spending score features.

3.3. Implementation

In this project, the prescribed sequence is:

- Creating an approach to solve the given problem statement

- Exploring the dataset and obtaining useful insight from the same

- Cleaning the dataset by handling nan values, remove duplicate records, etc.

- Data Visualization used to obtain important information from the data

- Data Preprocessing is performed to make the data ready to fit the model this includes

- feature scaling, splitting the dataset into features and labels, etc.

- Model Building

3.3.1. Block Diagram
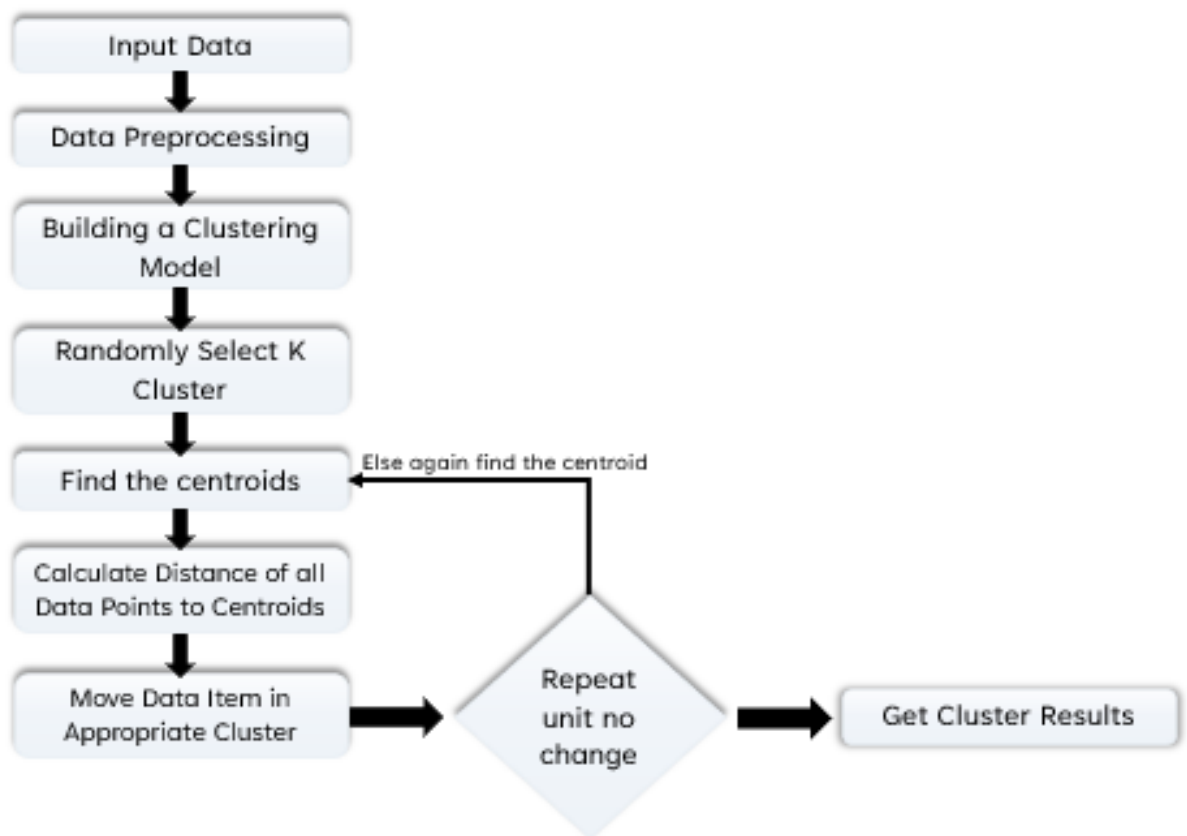
3.3.2. Flow Chart Diagram

Fig. 2: Flow Chart

4. RESULTS AND EXPLANATION

4.1. Implementation Approaches

Step 1: Define Objectives
Clearly articulate the objectives of customer segmentation, such as improving marketing effectiveness, enhancing customer experience, and optimizing business strategies.

Step 2: Gather Data
Collect relevant customer data from diverse sources, including demographics, purchase history, visit frequency, and any other information that might contribute to understanding customer behavior.

Step 3: Explore and Clean Data
Load the dataset into the Jupyter Notebook and perform exploratory data analysis (EDA). Explore the structure of the data, check for missing values, outliers, and inconsistencies. Clean the data to ensure accuracy in subsequent analyses.

Step 4: Select Features
Identify and select features that are meaningful for customer segmentation. Common features include age, spending habits, visit frequency, and any other relevant variables.

Step 5: Standardize Features
Standardize numerical features to ensure that they are on a consistent scale. This step is crucial for algorithms like k-means clustering that are sensitive to the scale of features.

Step 6: Determine Optimal Number of Clusters
Use techniques such as the elbow method or silhouette analysis to determine the optimal number of clusters (k) for the k-means algorithm. This step helps in finding a balance between model complexity and performance.

Step 7: Apply K-means Clustering
Implement the k-means clustering algorithm using the chosen k value. Assign each customer to a cluster based on their feature values.

Step 8: Visualize Customer Segments
Create visualizations to represent the customer segments. Scatter plots, cluster plots, or other

visualizations can help in understanding how customers are grouped.

Step 9: Profile Customer Segments
Analyze each customer segment's characteristics and behaviors. Understand the unique traits that define each segment, such as average age, spending patterns, and visit frequency.

Step 10: Implement Customer Strategies
Develop marketing or operational strategies tailored to each customer segment. Consider personalized promotions, targeted communication, or loyalty programs to address the specific needs of each group.

Step 11: Evaluate and Validate
Evaluate the effectiveness of the segmentation approach. Use metrics such as silhouette score, within-cluster sum of squares, or domain-specific KPIs. Validate results against real-world observations or historical data to ensure practical relevance.

Step 12: Document and Report
Create a comprehensive Jupyter Notebook documenting the entire analysis. Include detailed explanations of each step, code comments, visualizations, and key findings. This documentation facilitates collaboration and future reference.

Step 13: Conclusion and Recommendations
Summarize key findings, draw conclusions, and make actionable recommendations based on the identified customer segments.

4.2. Pseudo Code

Step 1 : Importing essential libraries
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

Step 2: Load and Explore the Dataset
df = pd.read_csv("Mall_Customers.csv")

Step 3: Data Preprocessing
df_cleaned = preprocess_data(df)

Step 4: Finding optimal number of clusters using Elbow Method

Elbow method is a tool used for analysing the clusters formed from our dataset and helps to interpret the appropriate number of optimal clusters in dataset.

wcss = []

for i in range(1,11):

kmeans_model = KMeans(n_clusters=i, init='k-

means++', random_state=42)

kmeans_model.fit(X)

wcss.append(kmeans_model.inertia_)

Step 5 :Model Building Training the KMeans model with n_clusters=5

kmeans_model = KMeans(n_clusters=5, init='k-means++', random_state=42)

y_kmeans = kmeans_model.fit_predict(X)

4.3. Analysis (graphs/chart)
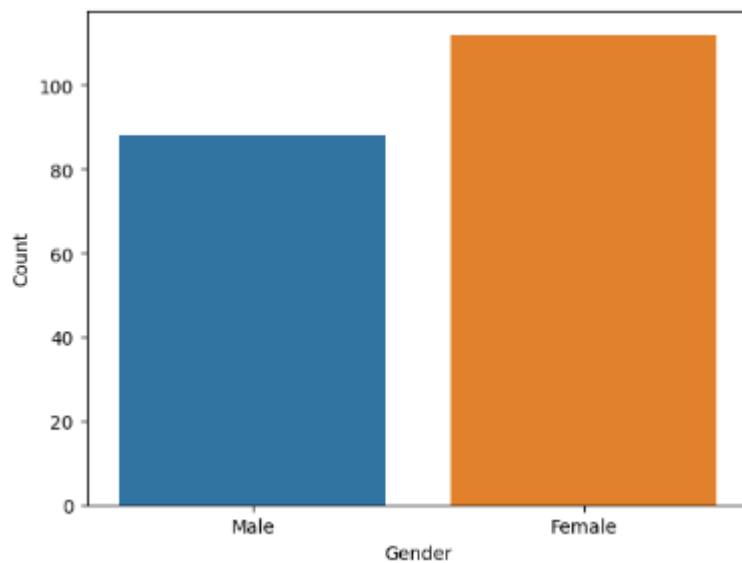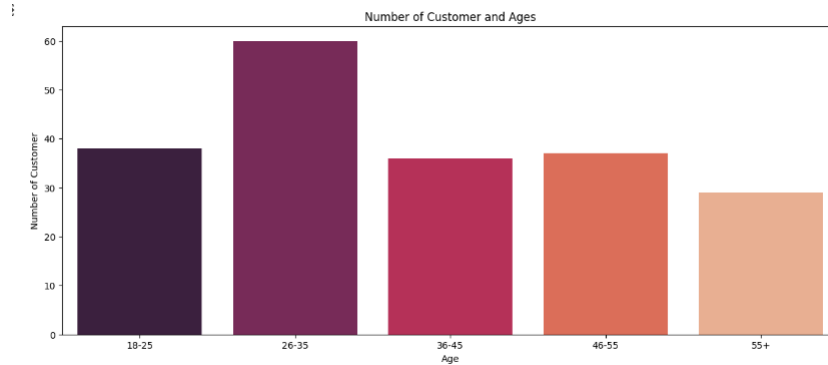

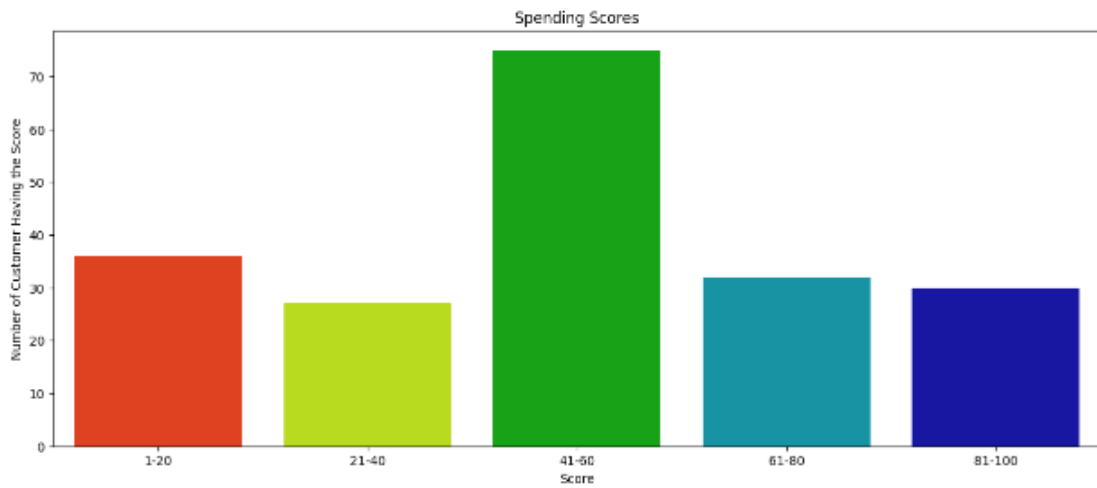Fig. 3: Gender Count Plot

Fig. 4: Age Graph
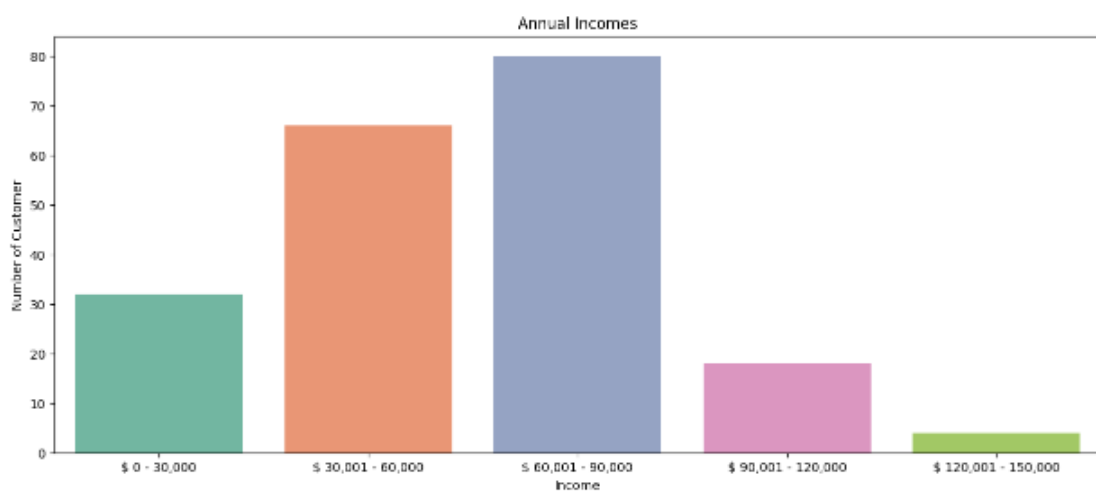

Fig. 5: Spending Score
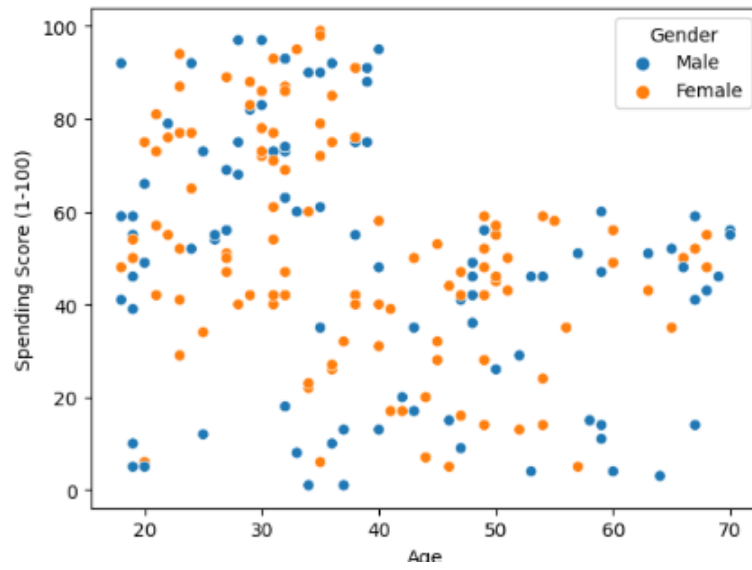

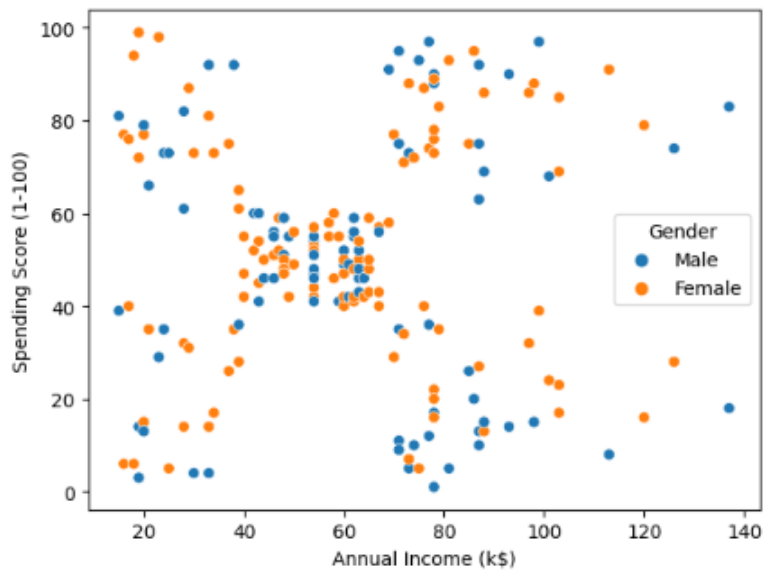Fig. 6: Annual Income

Fig. 7: Age Vs Spending Score

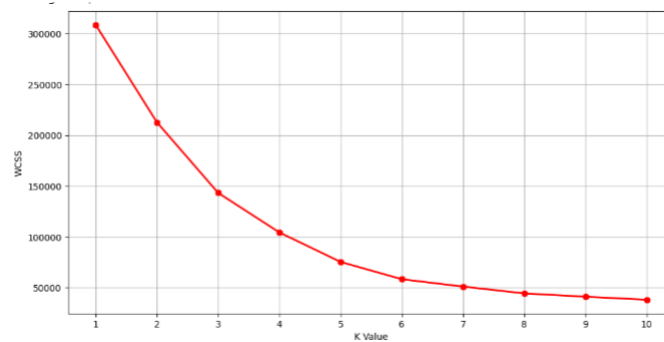Fig. 8: Annual Income Vs Spending Score

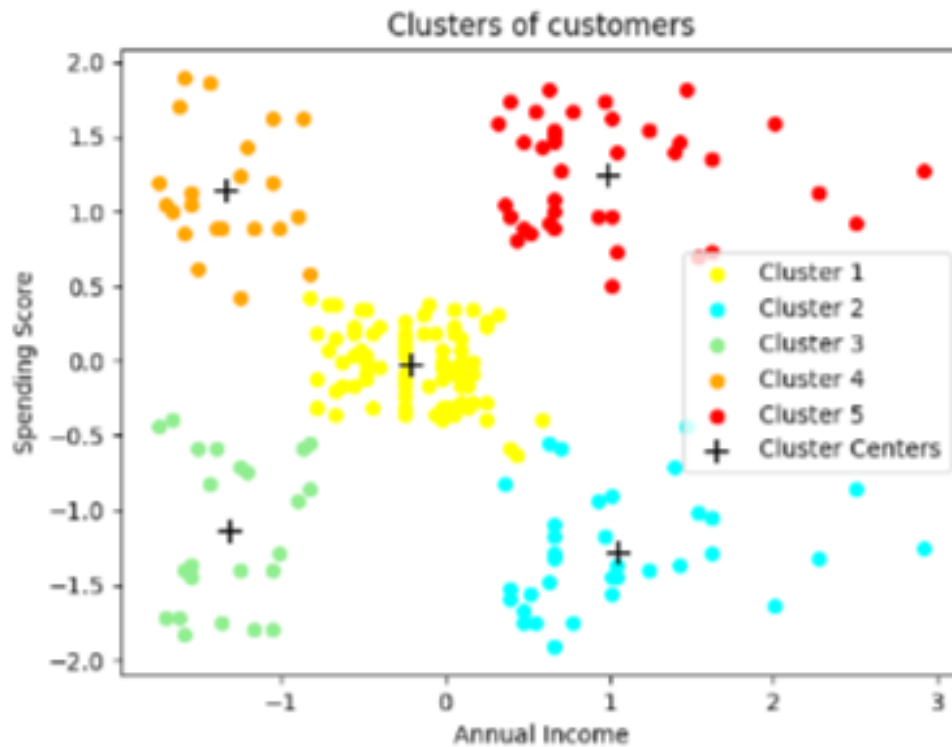Fig. 9: Elbow Method for determining Number of K clusters

Fig. 11: Cluster Analysis

Clustering - Observation

a. High Income, High Spending Score (Cluster 5) -
Target these customers by sending new product alerts which would lead to increase in the revenue collected by the mall as they are loyal customers.
b. High Income, Low Spending Score (Cluster 3) -
Target these customers by asking the feedback and advertising the product in a better way to convert them into Cluster 5 customers.
c. Average Income, Average Spending Score (Cluster 2) -
Can target these set of customers by providing them with Low cost EMI's etc.
d. Low Income, High Spending Score (Cluster 1) -
May or may not target these group of customers based on the policy of the mall.
e. Low Income, Low Spending Score (Cluster 4) -
Don't target these customers since they have less income and need to save money.

5. CONCLUSION

Using on the k-means clustering we have managed to form different clusters based on different features. Mall management can target the clusters with average spending score to increase their profit and should also maintain good relationship with premium customers with high spending score.They should also work on coming up with new innovative ideas to upgrade the customers with low spending score.

The significance of customer segmentation in attracting the customers towards the products which inturn aids the increase in the business scale in the market.
Segmenting the customer group into the different groups according to the similarities they possess, on one hand, helps the marketers to provide customized ads, products, and offers.
where on other hand it supports the customers by avoiding them from the confusion of the products to buy.
Comparing the clusters obtained by deploying the three different clustering algorithms on the customers data using the metrics that measure the distinctness and uniqueness of the clusters.
It couldn't be said that the Kmeans is the most effective clustering algorithm every time.
It depends on the various factors such as the size of the data,attributes of the data,etc., This Project can further be enhanced by including different clustering algorithms that may depict more proficiency and by considering the large datasets which in turn increases the efficiency.

References

DATASET :

- https://www.kaggle.com/datasets/shwetabh123/mall-customers

WEBSITES

- https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning
- https://www.geeksforgeeks.org/k-means-clustering-introduction/