



**CUSTOMER SEGMENTATION**

**USING DATASCIENCE**

**APPLIED DATASCIENCE PHASE 5**

**TEAM MEMBERS:**

1. **U. SWATHI**
2. **A.VIDHYA**
3. **U.SNEHA SRI**
4. **K.PAVITHRA**

**TABLE OF CONTENTS:**

**1. INTRODUCTION**

**2. PROBLEM STATEMENT IN CUSTOMER SEGMENTATION**

**3. DESIGN THINKING**

**4. FEATURE ENGINEERING**

**5. APPLYING CLUSTERING ALGORITHMS**

**6. VISUALIZATION**

**7. INTERPRETATION**

**8. DATA PREPROCESSING**

**9. DATASET USED**

**10. CONCLUSION**

**INTRODUCTION:**

****

## 

## Customer segmentation involves implementing data science methods to divide the customer base into smaller groups based on certain characteristics. It assists marketing managers in better understanding their customers' preferences and presenting them with better-targeted advertisements Customer.

**PROLEM STATEMENT IN CUSTOMER SEGMENTATION:**



**PROBLEM:**

**Many businesses are faced with the challenge of effectively understanding and catering to the diverse needs and preferences of their customer base. To address this challenge, customer segmentation is a crucial strategy. The problem statement in customer segmentation can be defined as follows:**

**BACKGROUND:**

**Customer segmentation is the process of dividing a company's customer base into distinct groups based on various characteristics such as demographics, behaviors, preferences, and purchasing habits. The goal is to identify homogeneous groups of customers to better tailor marketing strategies, product offerings, and customer experiences, ultimately increasing customer satisfaction and revenue.**

**SOLUTIONS:**

**1) Review industry data and market analysis:  
  
Industry analysis: Assesses the general industry environment in which  
you compete. Target market analysis: Identifies and quantifies the  
customers that you will be targeting for sales. Competitive analysis:  
Identifies your competitors and analyzes their strengths and weaknesses.**

**2) Examine your current customer base:  
  
Your customer base is made up primarily of people who repeatedly buy  
your products or use your service**

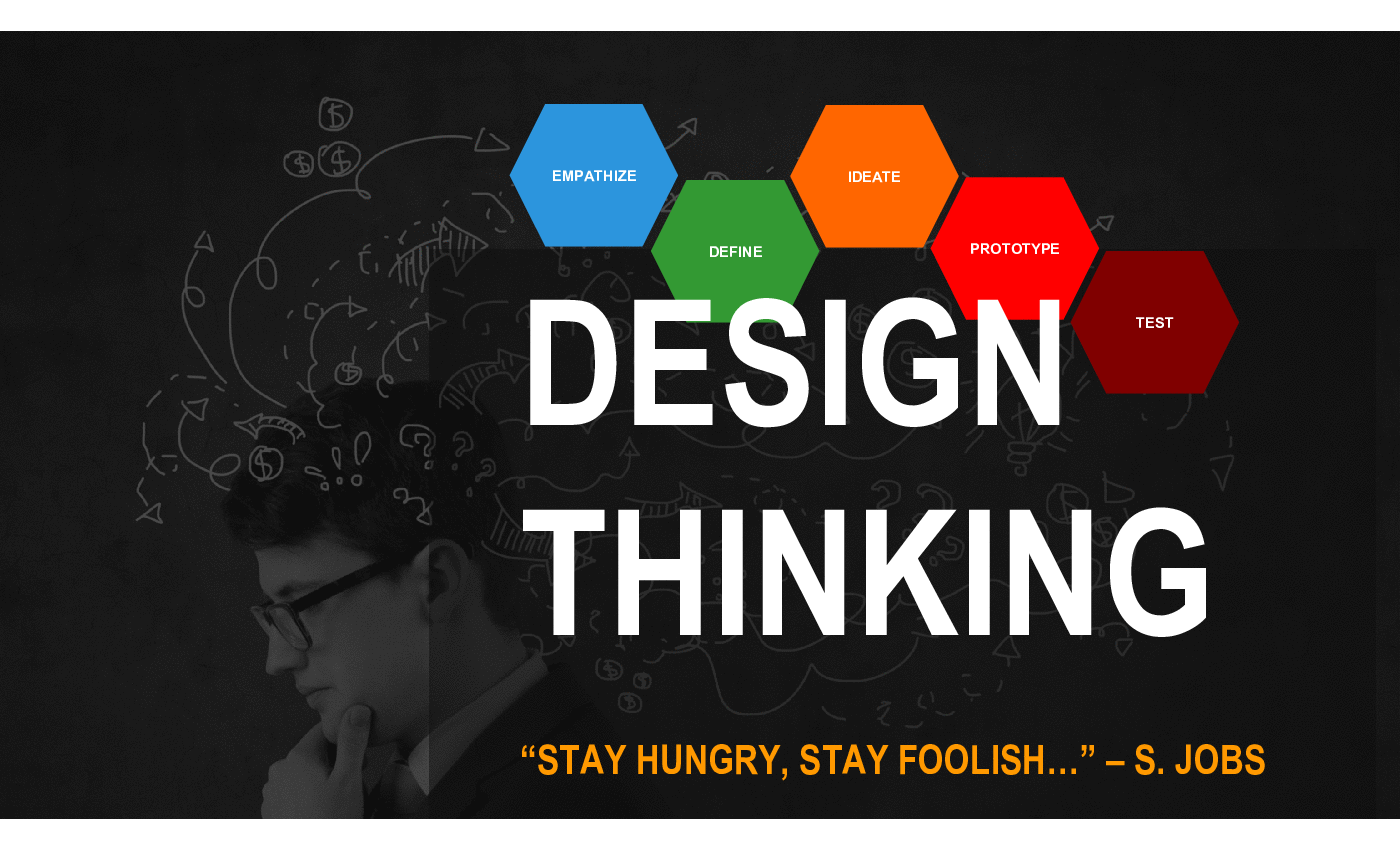
**3) Choose a segmentation model:  
  
Demographic segmentation divides customers based on factors like age,  
gender, and income.**

**4) Consider customer segmentation software:  
  
Customer segmentation software is a tool that groups your customers  
based on certain similarities in their demographics, behavior, or any  
other characteristics.  
  
5) Collect customer experience data:  
  
The process of collecting and analyzing customer data, with the goal of  
better understanding customer needs, viewpoints, and experiences with  
your products and services.**

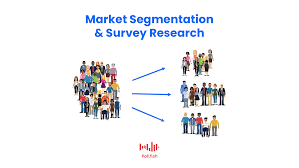
**6) Analyze customer experience data:  
  
Customer Experience Analysis in 3 Steps  
Step 1: Choose Your Key Customer Journey Points. ...  
Step 2: Gather Data on Those Customer Journey Points. ...  
Step 3: Analyze Your Data to Identify Patterns and Trends.**

**7) Refine your customer segment:  
  
Another important way to refine your customer segments over time is to analyze data from your customer interactions and transactions. You can use tools, such as analytics, CRM, or email marketing platforms, to track and measure how your customers behave, respond, and convert across different channels and touch points.**

**DESIGN THINKING IN CUSTOMER SEGMENTATION:**



**RESEARCHING THE PROBLEM IN-DEPTH IN CUSTOMER SEGMENTATION:**



# Data Quality and Integration:

**Problem: Many companies struggle with incomplete, inaccurate, or  
outdated customer data. Integrating data from various sources can also be  
complex.  
 Impact: Poor data quality can lead to incorrect segmentation, resulting in  
ineffective marketing efforts and missed opportunities.**

Privacy and Regulatory Compliance:

**Problem: Increasing concerns about data privacy and stricter regulations   
(e.g., GDPR, CCPA) can limit the types of data that can be collected and  
used for segmentation.**

**Impact: This can hinder the depth and breadth of segmentation, making it  
challenging to create highly targeted segments.**

# Segmentation Overlap:

**Problem: Some customers may fit into multiple segments simultaneously,  
making it difficult to determine which segment should be targeted.  
 Impact: Overlap can lead to conflicting marketing strategies, potentially  
overwhelming customers with irrelevant messages.**

# Segmentation Drift:

**Problem: Customer preferences and behaviors can change over time,  
causing segments to become less accurate or outdated.  
 Impact: Outdated segments can result in misaligned marketing efforts and  
lost opportunities.**

# Resource Intensiveness:

**Problem: Developing and maintaining a segmentation strategy can be  
resource-intensive in terms of time, expertise, and technology.  
 Impact: Smaller businesses with limited resources may struggle to  
implement effective segmentation stregy.**

**IDEATING THE POSSIBLE SOLUTION:**

# 100 Problem Solution Essay Topics with Sample Essays - Owlcation

# Data Quality and Integration:

**Solution: Invest in data cleaning and validation processes to ensure data  
accuracy. Implement a data integration strategy to consolidate customer data  
from various sources into a single, comprehensive view.**

# Privacy and Regulatory Compliance:

**Solution: Stay compliant with data privacy regulations and obtain clear**• **consent for data usage. Collect only necessary data and anonymize or  
pseudonymize sensitive information when possible.**

# Segmentation Overlap:

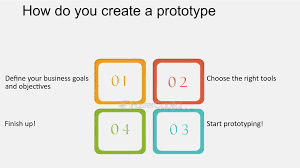
**Solution: Develop a hierarchy or prioritization system for segments to**• **handle overlapping customers. Create clear rules for assigning customers to segments based on their primary characteristics.**

# Segmentation Drift:

**Solution: Regularly update and reevaluate customer segments to account for changing behaviors and preferences. Implement automated tracking systems to monitor segment performance.**

Resource Intensiveness:  
  
 **Solution: Prioritize segmentation efforts based on potential impact. Consider outsourcing data analysis or segmentation tasks to experts if resources are limited.**

**CREATING A PROTOTYPE:**

  
  
**Creating a prototype for customer segmentation involves developing a simplified version of your segmentation strategy or model to test its  
feasibility and effectiveness before implementing it on a larger scale.**

TESTING AND TROUBLESHOOTING:  
 **Creating and troubleshooting customer segmentation is an iterative process.  
Regularly review and adapt your segmentation strategy to meet evolving customer needs and market dynamics. Keep stakeholders informed and engaged throughout the process to ensure alignment and support for your segmentation efforts.**

# Troubleshooting Common Issues:

**1. Inaccurate Segmentation  
2. Segment Overlap  
3. Segment Drift  
4. Data Quality Problems  
5. Lack of Action ability  
6. Resistance to Change  
7. Low ROI  
8. Ethical Concerns  
9. Ineffective Personalization**

**MAKING IMPROVEMENTS TO AND RELEASING THE FINAL PRODUCT:**

**Creating and improving a customer segmentation strategy is an ongoing process that involves continuous refinement and adaptation. To release a final product or version of your customer segmentation strategy, follow these steps:  
  
1. Review Existing Segmentation  
2. Define Objectives  
3. Collect Comprehensive Data  
4. Data Quality Assurance  
5. Update Segmentation Variables  
6. Choose Segmentation Methods  
7. Segment the Data  
8. Profile the Segments  
9. Develop Marketing Strategies  
10. Implementation Plan**

PHASES OF DEVELOPMENT:

# DATA COLLECTION:

**Surveys are a great way to get data related to demographics, purchasing patterns, preferences, and other distinct categories. You can also use interviews, existing customer data, focus groups, and other data collection methods.**

**Customer Satisfaction Tools**

**1. Net Promoter Score.  
2. Live Chat.  
3. Social Media Mentions.  
4. Marketing Emails.  
5. Short Message Service (SMS)  
6. Churn Rate.  
7. Follow-Up Surveys**DATA INSPECTION: **Data inspection in customer segmentation refers to the  
systematic examination and analysis of the data collected  
from customers or potential customers in order to  
understand its quality, characteristics, and suitability for  
segmentation purposes. This process involves inspecting  
the data to identify patterns, trends, anomalies, and other  
important insights that will inform the segmentation strategy.  
It is a critical step in the customer segmentation process as  
it helps ensure that the data used for segmentation is clean,  
relevant, and of high quality, which, in turn, leads to more  
accurate and effective customer segments.**

DATA CLEANING:  
 **Data cleaning is the act of sifting through your data to detect and  
correct inaccuracies or any corrupted elements. It is especially  
valuable for digital marketers who are combining performance  
data from multiple sources. It is part of the overall data  
management process.**DATA TRANSFORMATION: **Data transformation in customer segmentation is the set of  
procedures and operations applied to customer data to make it more  
usable, consistent, and relevant for the purpose of creating  
meaningful customer segments. This may include actions such as  
encoding categorical variables, scaling numerical features, handling  
missing values, creating new features, or reducing dimensionality.  
The transformed data provides a solid foundation for generating  
customer segments that can be used for targeted marketing, product  
customization, and other business strategies.**

EXPLORATORY DATA ANALYSIS:  
 **Exploratory data analysis (EDA) in customer segmentation is the  
process of examining and visualizing customer data to gain a  
preliminary understanding of its key characteristics, patterns, and  
relationships.  
EDA helps in uncovering valuable information about customer  
behavior, preferences, and demographics, which is crucial for  
creating effective customer segments and enhancing marketing and  
business strategies.**

DATA VISUALIZATION:  
 **In various tools and formats to create visual representations of your  
segments, such as charts, graphs, tables, dashboards, or  
info graphics. The goal is to make your segments easy to understand  
and compare and to highlight the key differences and opportunities  
among them.  
  
4 Simple Steps to Customer Journey Data Visualization  
1. Step 1: Define the Main Aspects of the Customer Experience.  
2. Step 2: Choose the Right Data to Visualize.  
3. Step 3: Choose the Right Data Visualization Tools.  
4. Step 4: Get Actionable Insights and Make Better Decisions.  
5. Conclusion.**DATA SPLITTING:   
 **Data splitting is when data is divided into two or more subsets.  
Typically, with a two-part split, one part is used to evaluate or test the  
data and the other to train the model. Data splitting is an important  
aspect of data science, particularly for creating models based on  
data.**

DATA PREPROCESSING: **Data Segmentation is the process of taking the data you hold and  
dividing it up and grouping similar data together based on the chosen  
parameters so that you can use it more efficiently within marketing  
and operations.**

**The steps to be taken are:**

**1. Read image.  
2. Resize image.  
3. Remove noise (Demise)  
4. Segmentation.  
5. Morphology (smoothing edges)**

**FEATURE ENGINEERING:**

**Feature engineering is a crucial step in customer segmentation. It involves selecting, transforming, or creating relevant features from your data to improve the quality of your customer segments. Effective feature engineering can lead to more meaningful and actionable segments.**

**APPLYING CLUSTERING ALGORITHMS:**

**Applying clustering algorithms in customer segmentation is a critical step in the process of identifying and grouping customers with similar characteristics. Clustering algorithms help automate the segmentation process and uncover hidden patterns within your customer data.**

**VISUALIZATION:**

**Visualization plays a crucial role in customer segmentation, as it helps you understand the structure of your customer segments and communicate your findings effectively. Here is some common visualization techniques used in customer segmentation:**

**INTERPRETATION:**

**Interpretation in customer segmentation involves making sense of the results and drawing meaningful insights from the generated customer segments. Effective interpretation is essential for understanding the characteristics and behaviors of different customer groups and for making data-driven decisions.**

**DATA PREPROCESSING:**

**Data Segmentation is the process of taking the data you hold and  
dividing it up and grouping similar data together based on the chosen  
parameters so that you can use it more efficiently within marketing  
and operations.  
The steps to be taken are:  
1. Read image.  
2. Resize image.  
3. Remove noise (Denoise)  
4. Segmentation.  
5. Morphology (smoothing edges)visual**

import pandas as pd

import numpy as np

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans

import matplotlib.pyplot as plt

data = pd.read\_csv("C:\\Users\\prath\\Downloads\\Mall\_Customers.csv")

# Display the first few rows to inspect the data

print(data.head())

X = data.iloc[:, [3, 4]].values

# Feature scaling (standardization)

scaler = StandardScaler()

X = scaler.fit\_transform(X)

wcss = []

for i in range(1, 11):

    kmeans = KMeans(n\_clusters=i, init='k-means++', random\_state=42)

    kmeans.fit(X)

    wcss.append(kmeans.inertia\_)

plt.figure(figsize=(8, 6))

plt.plot(range(1, 11), wcss, marker='o', linestyle='--')

plt.title('Elbow Method')

plt.xlabel('Number of Clusters')

plt.ylabel('WCSS (Within-Cluster Sum of Squares)')

plt.show()

import numpy as np

import pandas as pd

 import matplotlib.pyplot as plt

 from sklearn.preprocessing import StandardScaler

 from sklearn.cluster import KMeans, AgglomerativeClustering from sklearn.metrics import silhouette\_score from sklearn.decomposition import PCA data=pd.read\_csv('/content/archive (1).zip)

data = data.drop(columns=["CustomerID"])from sklearn.preprocessing import StandardScaler

scaler = StandardScaler() scaled\_data = scaler.fit\_transform(data)

scaled\_df = pd.DataFrame(scaled\_data, columns=data.columns)

wcss = [] for i in range(1, 11): kmeans = KMeans(n\_clusters=i, init='k-means++', max\_iter=300, n\_init=10, random\_state=0) kmeans.fit(scaled\_df)

wcss.append(kmeans.inertia\_)

plt.figure(figsize=(10, 6)) plt.plot(range(1, 11),

wcss, marker='o', linestyle='--') plt.title('Elbow Method') plt.xlabel('Number of clusters') plt.ylabel('WCSS') # Within-Cluster-Sum-of-Squares plt.show()

optimal\_clusters = 5 # as per the elbow method kmeans = KMeans(n\_clusters=optimal\_clusters, init='k-means++', max\_iter=300, n\_init=10, random\_state=0) cluster\_labels = kmeans.fit\_predict(scaled\_df)

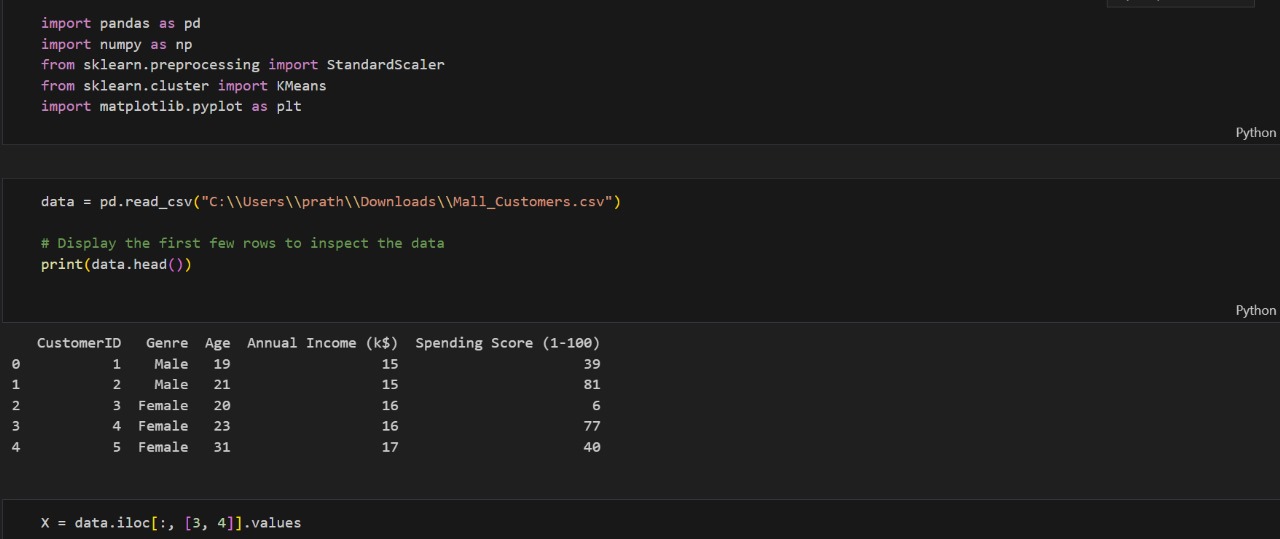
data['Cluster'] = cluster\_labels

from sklearn.decomposition import PCA

pca = PCA(n\_components=2) principal\_components = pca.fit\_transform(scaled\_df) pca\_df = pd.DataFrame(data=principal\_components, columns=['PC1', 'PC2']) pca\_df['Cluster'] = cluster\_labels

sns.pairplot(data=pca\_df, hue='Cluster', diag\_kind='kde') plt.show()

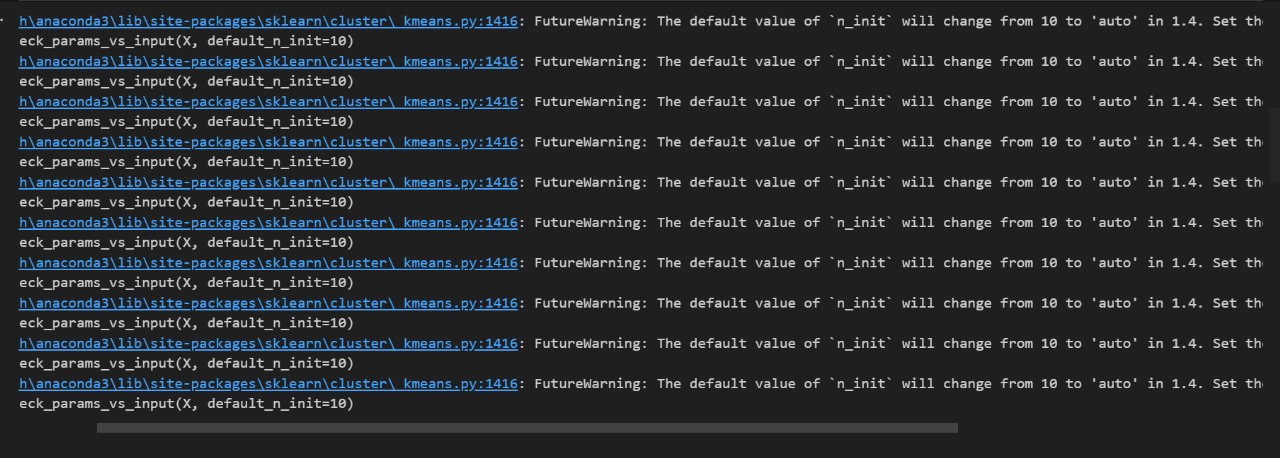
cluster\_centers = pd.DataFrame(scaler.inverse\_transform(kmeans.cluster\_centers\_), columns=data.columns[:-1]) print(cluster\_centers)

1)

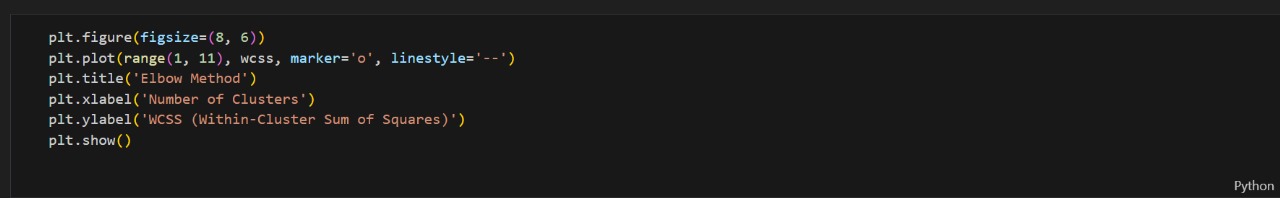
2)



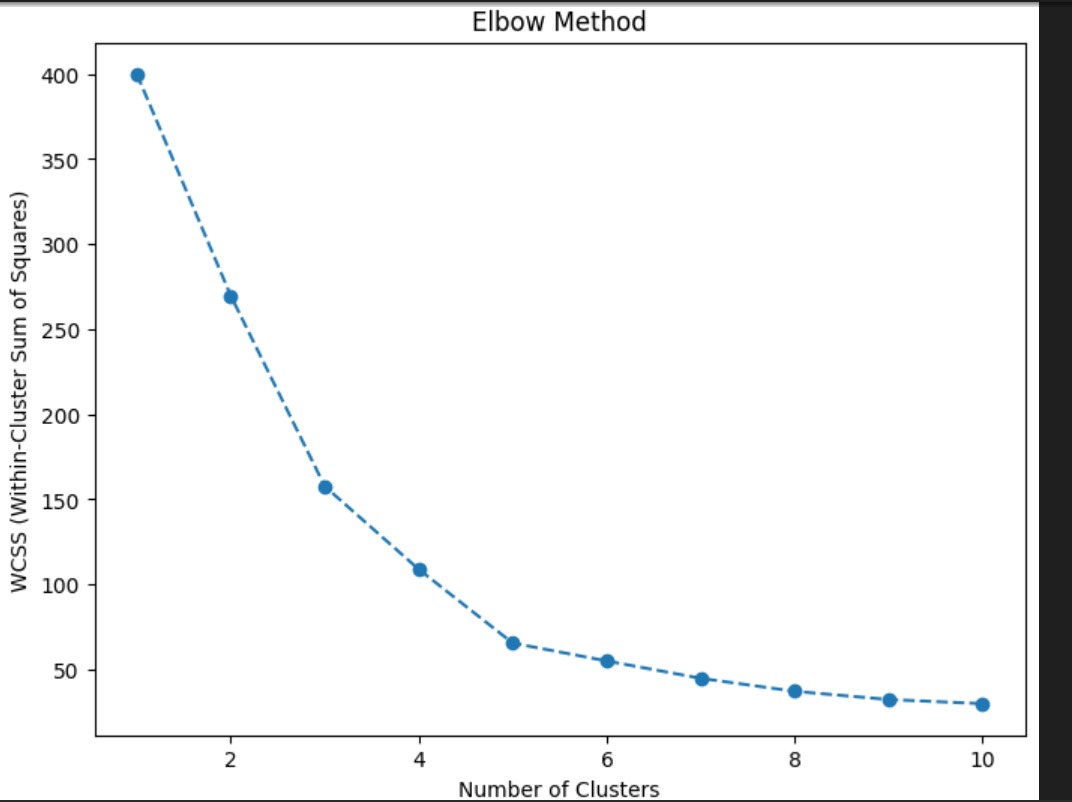
3)



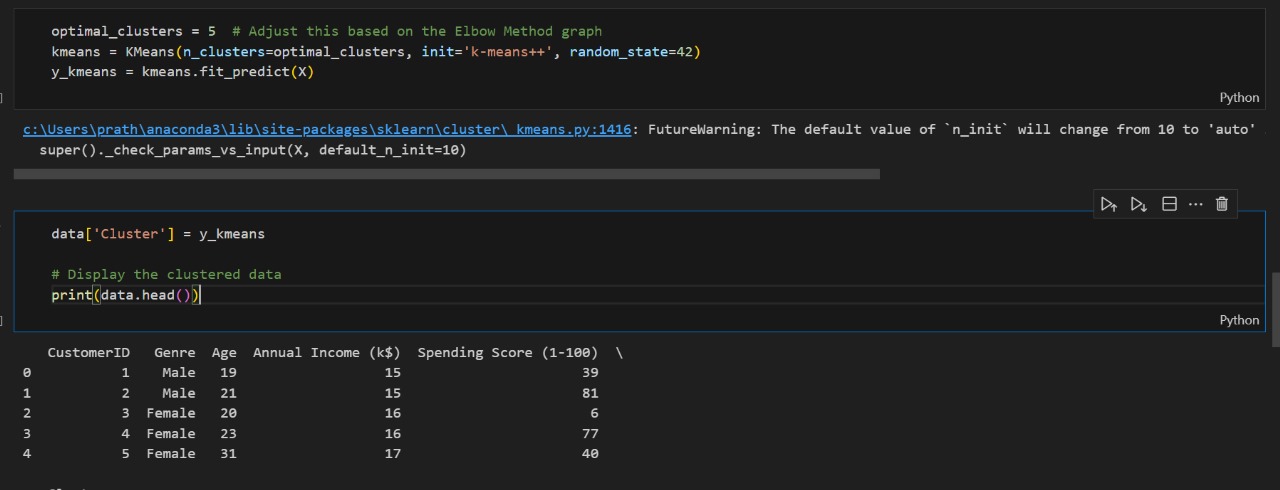
4)



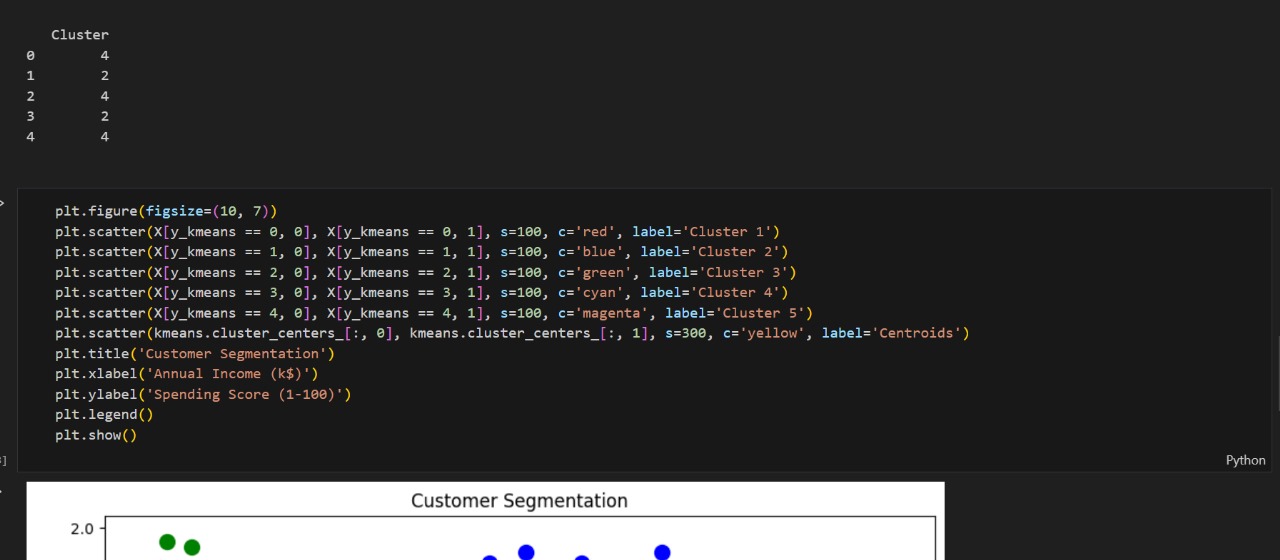
5) OUTPUT:



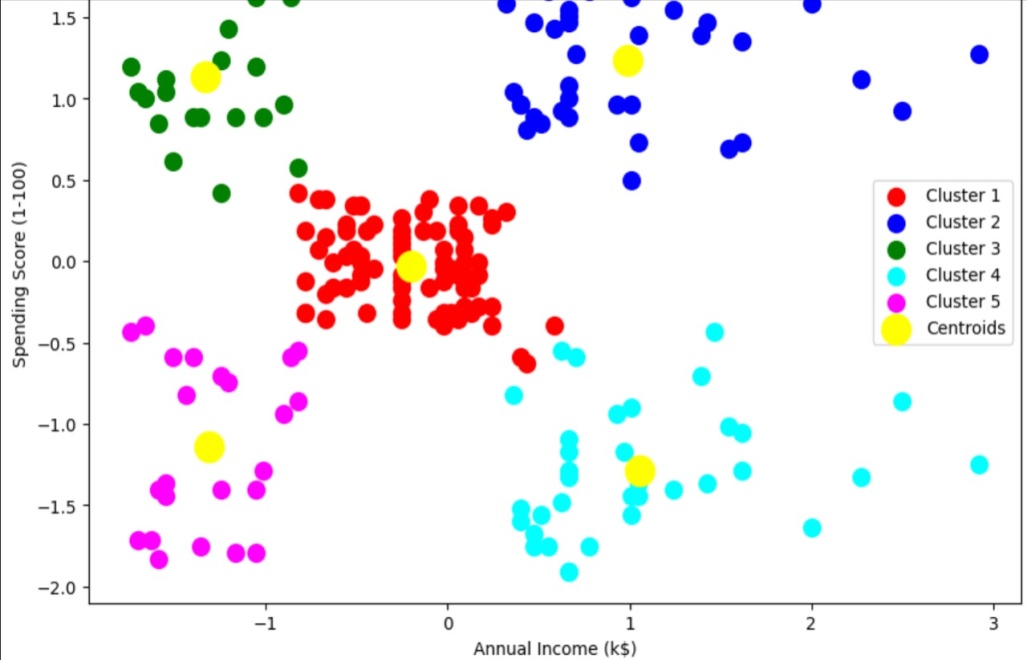
6)



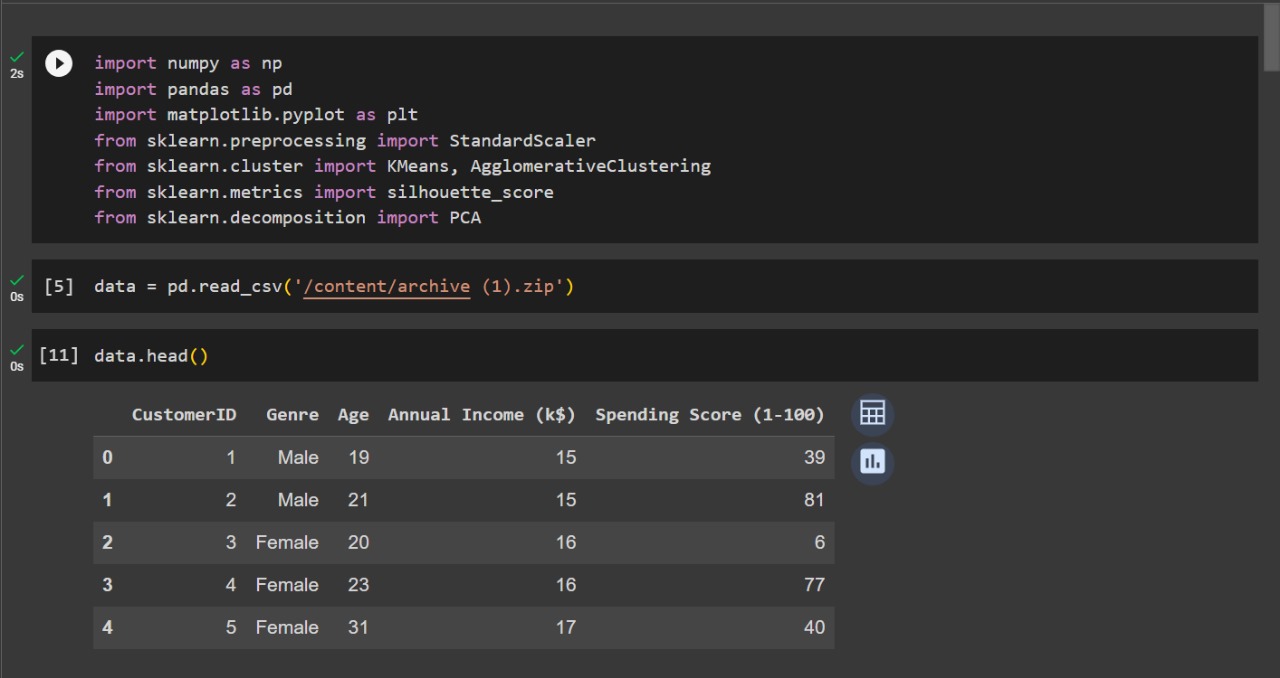
7)



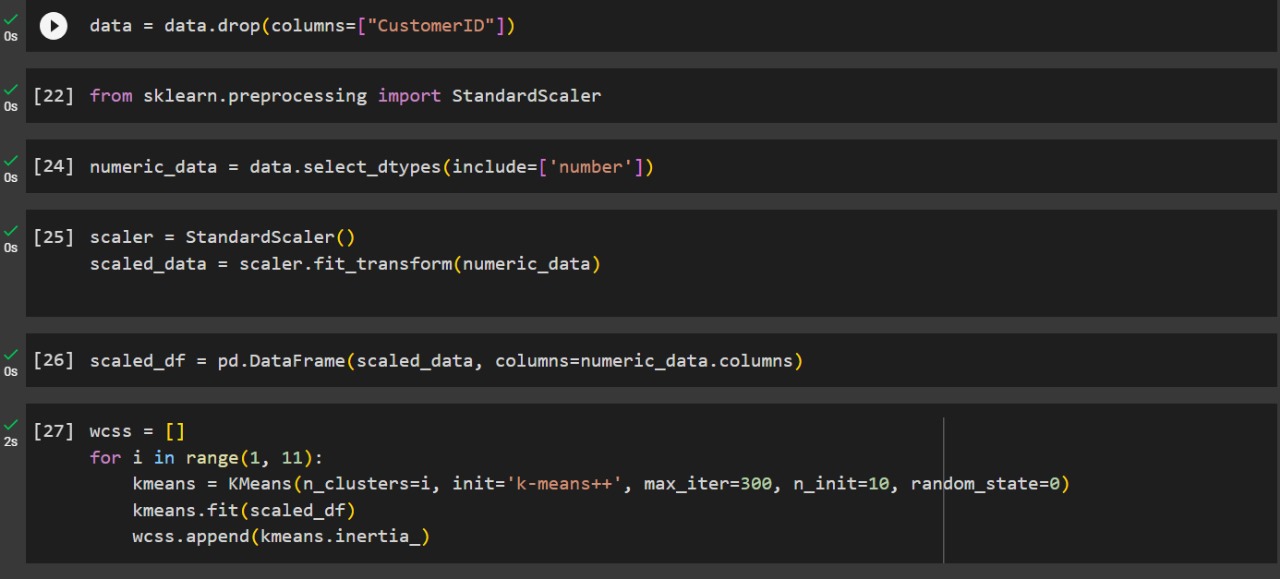
8) OUTPUT:



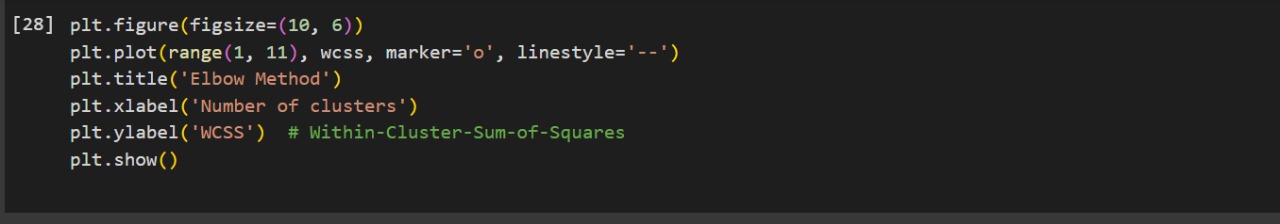
9)



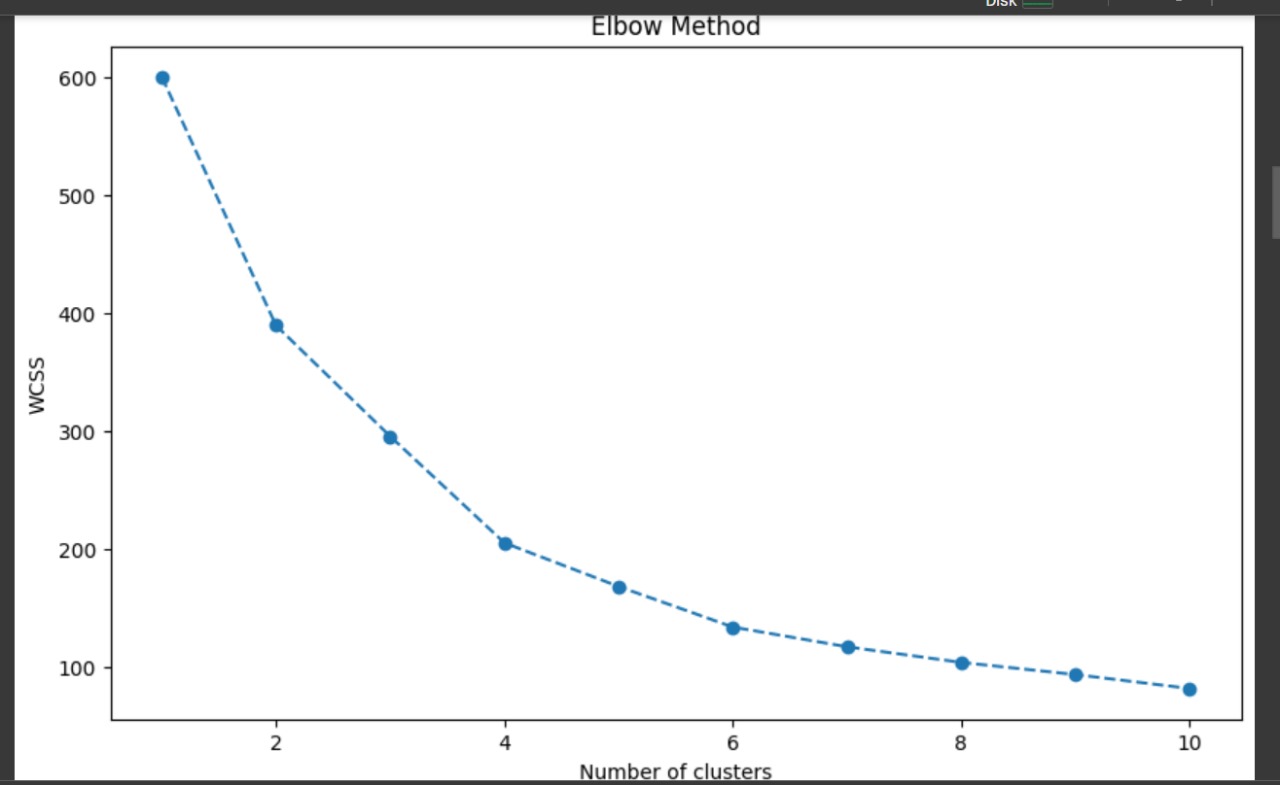
10)



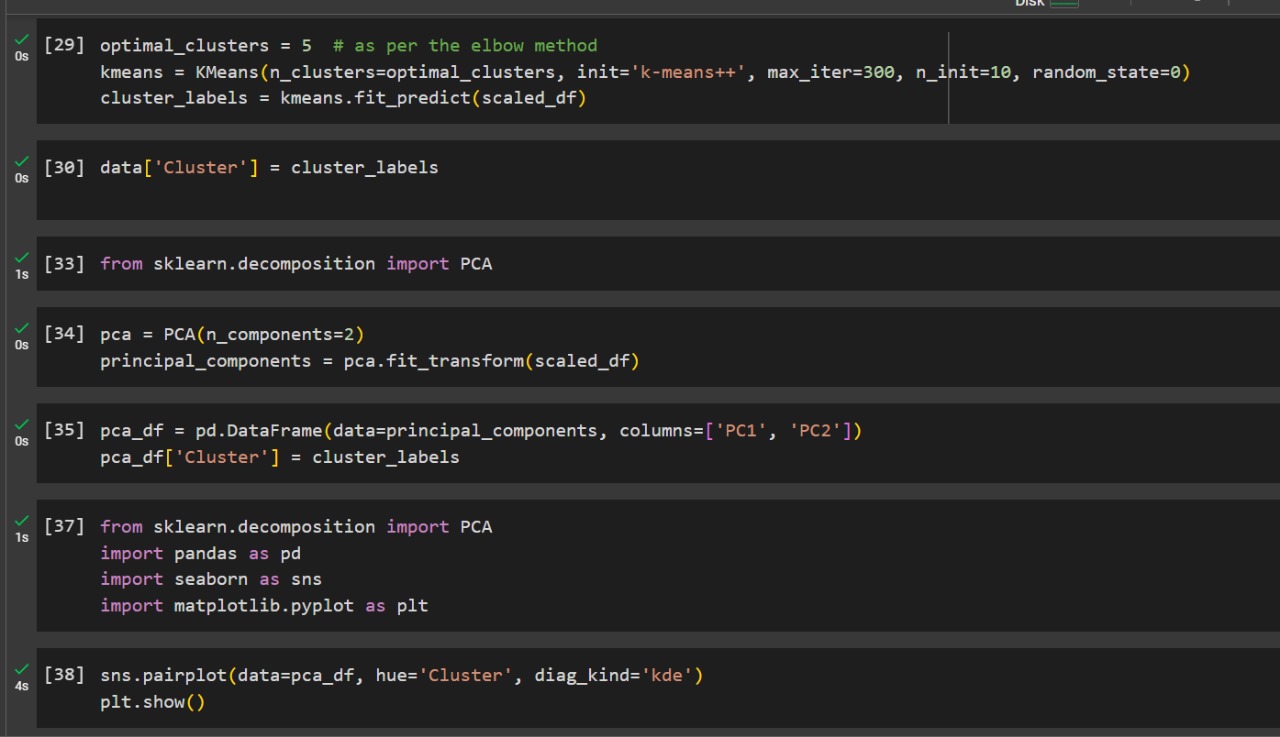
11)



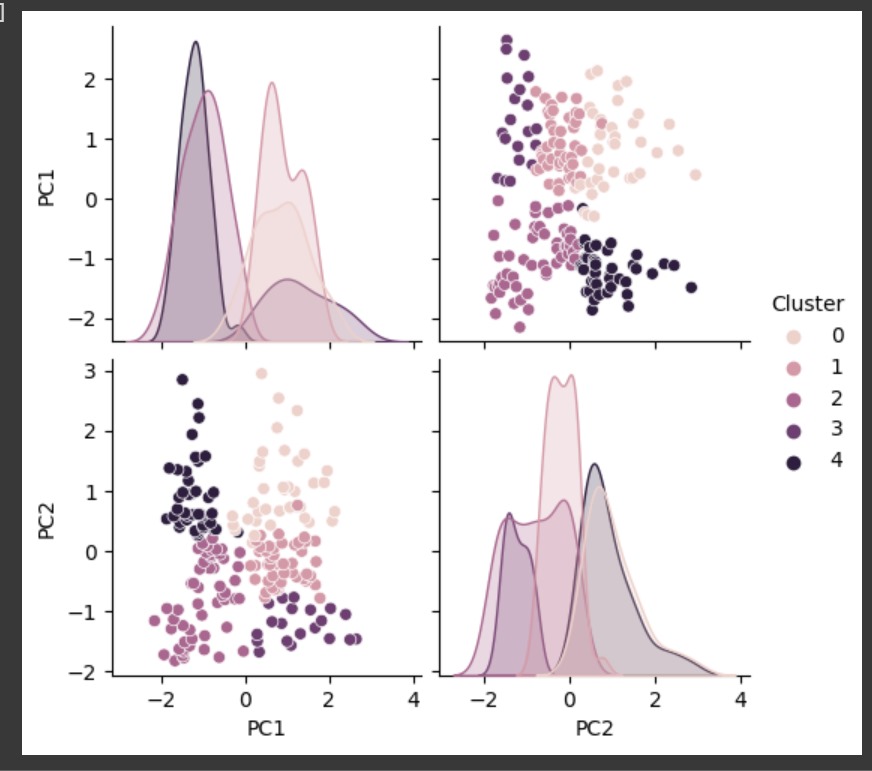
12) OUTPUT:



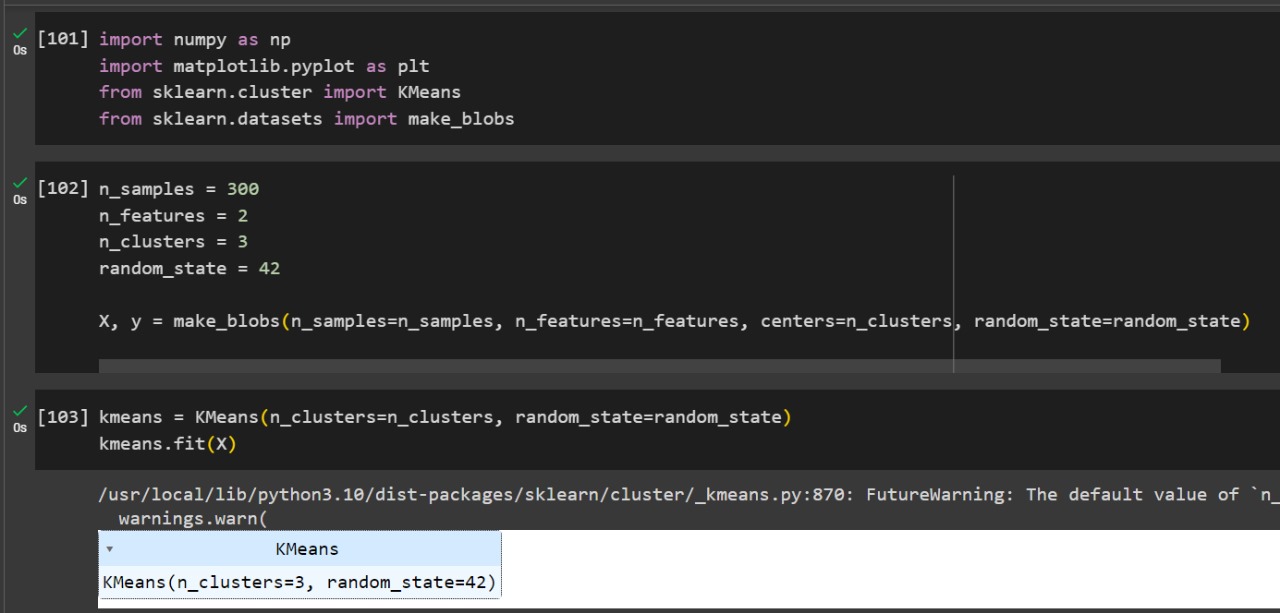
13)



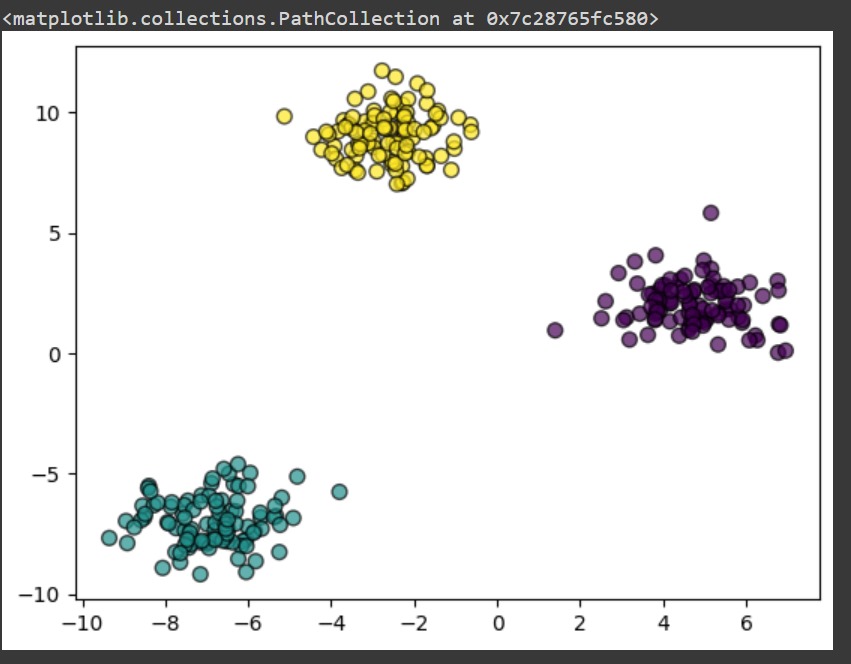
14) OUTPUT:



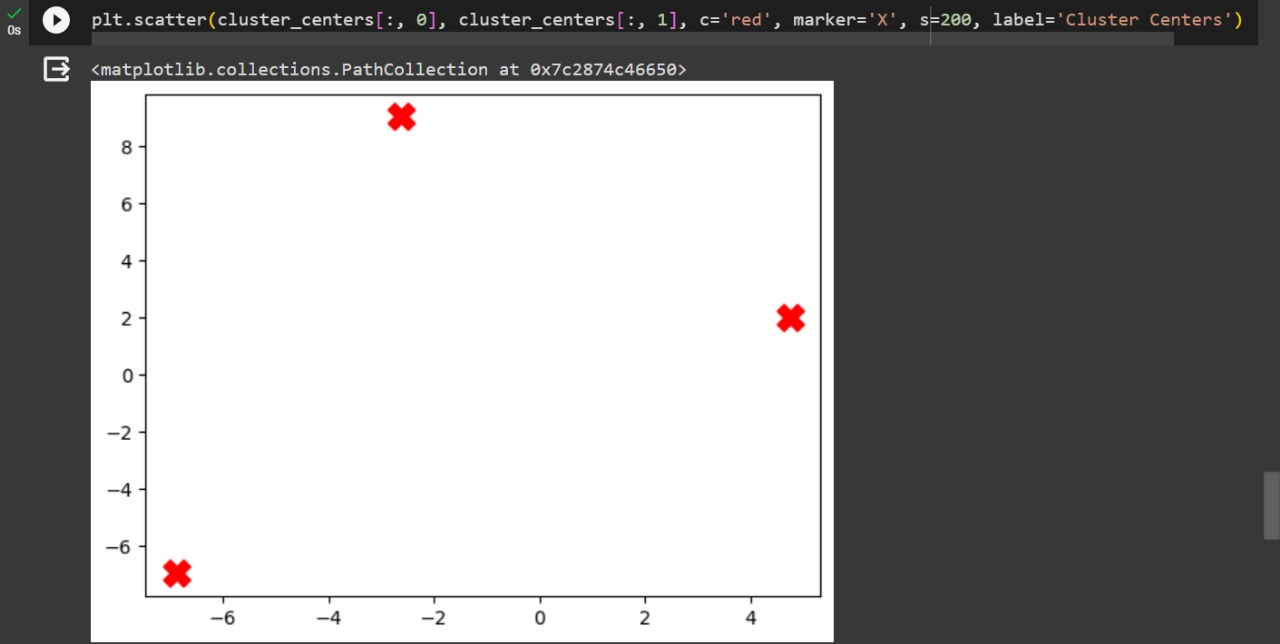
15)



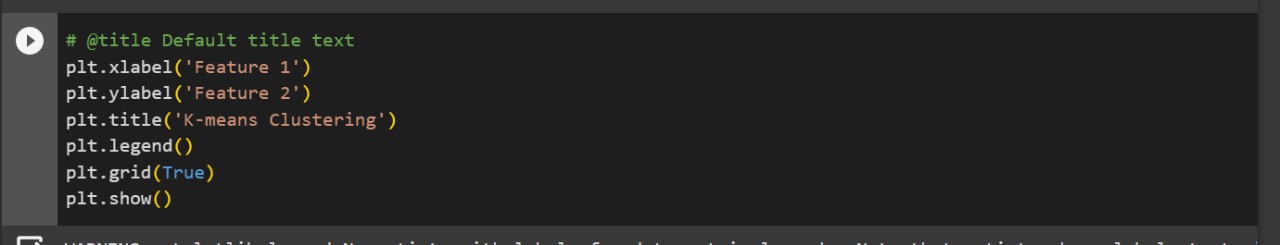
16) OUTPUT:



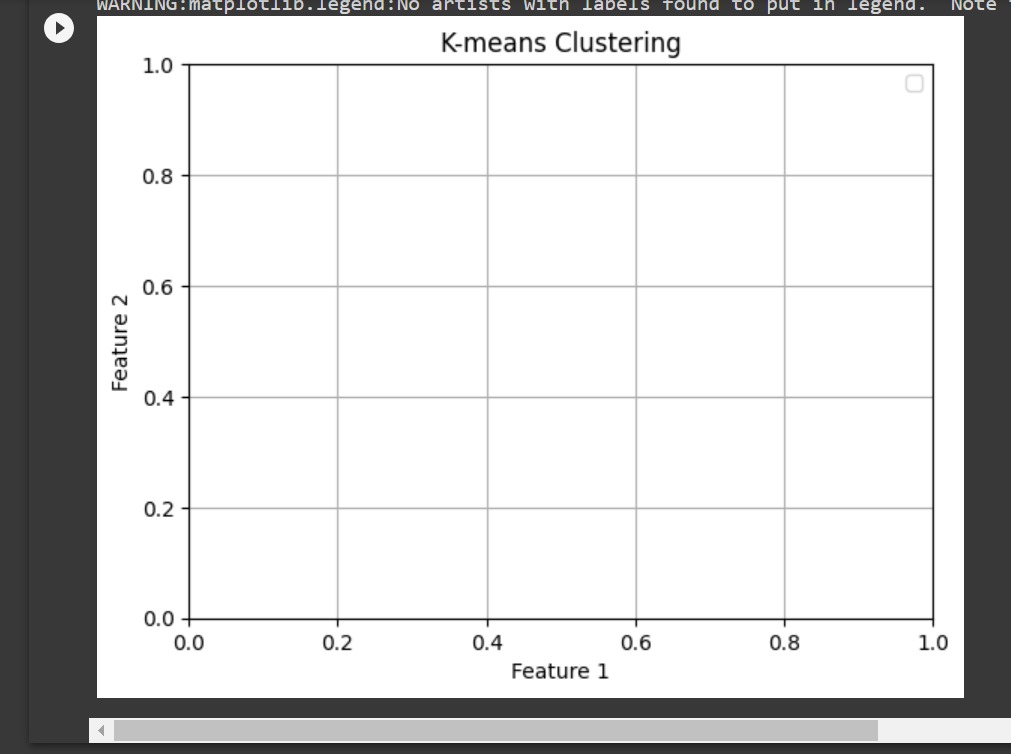
17) OUTPUT:



18)



19)OUTPUT:



**Description:**

- **We start by importing the necessary libraries, including NumPy, Pandas, Matplotlib, and scikit-**

**Learn&#39;s KMeans and StandardScaler.**

**- We load the customer data from a CSV file name &#39; customer\_data.csv&#39;.**

**- We define the features of interest, standardize the data, and perform K-Means clustering to create Customer segments.**

**- A scatter plot is used to visualize the clusters and centroids.**

**- We interpret the results by reverting the scaling to obtain cluster statistics and display them using Pandas.**

**DATASET USED:**

**1. Dataset Name: Mall Customers**

**2. Data source: It is taken from the survey made on the mall**

**CONCLUSION:**

**In this project we explained the below contents:**

**1. Summary of Objectives:**

**Objective 1: Understand Customer Diversity**

**Objective 2: Enhance Targeted Marketing**

**Objective 3: Improve Customer Experience**

**Objective 4: Increase Customer Retention**

**Objective 5: Optimize Resource Allocation**

**Objective 6: Boost Sales and Revenue**

**Objective 7: Measure Effectiveness**

**2. DATA AND METHODOLOGY:**

**Data and methodology are critical components of a customer segmentation project. The data provides the foundation for the analysis, while the methodology outlines the approach used to create meaningful customer segments.**

**3.** **Key Findings:**

**Key findings in customer segmentation represent the insights and discoveries derived from the analysis of customer data. These findings are crucial for businesses to make informed decisions and tailor their strategies for different customer groups. Here are some common key findings in customer segmentation:**

**Distinct Customer Segments**

**Segment Characteristics**

**Purchase Behavior**

**Customer Lifetime Value**

**Customer Preferences**

**4.** **Future Recommendations:**

Future recommendations in customer segmentation refer to suggestions for further enhancing and optimizing the segmentation strategy, ensuring that it remains relevant and effective in the dynamic business environment. Here are some key future recommendations for a customer segmentation project:

**Dynamic Segmentation**

**Predictive Modeling**

**Personalization at Scale**

**Omni-Channel Marketing**

**Multi-Touch point Attribution**

*THANK YOU*

**SUBMITTED BY:**

**U. SNEHA SRI**

**CSE-3RD YEAR**