GRT INSTITUTE OF

ENGINEERING AND

TECHNOLOGY, TIRUTTANI - 631209

Approved by AICTE, New Delhi Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PHASE 3

PROJECT TITLE

Customer Segmentation Using Data Science

COLLEGE CODE : 1103

Lokeshwar .K

3rd yr, 5th sem

Reg no. : 110321104027

lokeshwarnaidu2@gmail.com

PHASE 3 : CUSTOMER SEGMENTATION USING DATA SCIENCE

CUSTOMER SEGMENTATION IS A COMMON DATA SCIENCE AND MARKETING TECHNIQUE USED TO CATEGORIZE CUSTOMERS INTO DISTINCT GROUPS BASED ON SHARED CHARACTERISTICS. BY DIVIDING YOUR CUSTOMER BASE INTO SEGMENTS, YOU CAN BETTER TAILOR YOUR MARKETING STRATEGIES, PRODUCT OFFERINGS, AND CUSTOMER EXPERIENCES TO MEET THE SPECIFIC NEEDS AND PREFERENCES OF EACH GROUP. HERE'S A STEP-BY-STEP GUIDE ON HOW TO IMPLEMENT CUSTOMER SEGMENTATION USING DATA SCIENCE:

1. DATA COLLECTION:

- Gather relevant data on your customers. This data can include demographic information (age, gender, location), behavioral data (purchase history, browsing behavior, engagement with your products or services), and psychographic data (lifestyle, interests, values). You may collect this data through your CRM system, website analytics, surveys, or other sources.

2. DATA PREPROCESSING:

- Clean the data to ensure it's accurate and consistent. This involves handling missing values, removing duplicates, and dealing with outliers. You may also need to transform data into a suitable format for analysis.

3. FEATURE SELECTION/EXTRACTION:

- Decide on the features (variables) that are most relevant for segmentation. Some features may be more valuable in identifying meaningful segments. You can use techniques like Principal Component Analysis (PCA) or feature selection algorithms to help with this.

4. CHOOSE A SEGMENTATION TECHNIQUE:

- There are several methods to segment customers, and the choice depends on your data and business goals:

CLUSTERING: Use unsupervised learning techniques like K-Means, hierarchical clustering, or DBSCAN to group customers with similar attributes.

CLASSIFICATION: You can use supervised learning algorithms like decision trees, random forests, or support vector machines to classify customers into predefined segments if you have labeled data.

RFM ANALYSIS: Recency, Frequency, and Monetary analysis segments customers based on their transaction history.

CUSTOMER PROFILING: Manually create segments based on domain knowledge and create personas.

5. DATA MODELING:

- Apply the chosen segmentation technique to your data. This will result in the creation of customer segments.

6. EVALUATION:

- Assess the quality of your segments. Use metrics appropriate to the method you employed (e.g., silhouette score for K-Means clustering). You want segments to be distinct and internally homogeneous.
-

7. INTERPRETATION:

- Examine the characteristics of each segment. What sets them apart? What are their common behaviors and preferences? This will help you understand your customer base better.

8. IMPLEMENTATION:

- Once you have well-defined segments, you can use this information in various ways:

Targeted Marketing: Tailor your marketing messages and campaigns to each segment.

Product Development: Create or modify products and services to better meet the needs of specific segments.

Pricing Strategies: Adjust pricing based on the willingness to pay within each segment.

Customer Service: Customize customer service experiences based on segment preferences.

Recommendation Engines: Implement personalized recommendation systems for online platforms.

9. MONITORING AND ITERATION:

- Customer segments may evolve over time. Regularly monitor and update your segments to ensure they remain relevant. Use ongoing data to refine your strategies.

10. PRIVACY AND ETHICAL CONSIDERATIONS:

- Be mindful of data privacy and ethical concerns when handling customer data. Ensure that you are compliant with relevant regulations (e.g., GDPR) and respect customer privacy.

In addition to these steps, it's essential to keep the business goals in mind when implementing customer segmentation. Each business is unique, and the segmentation strategy should align with the specific objectives and customer dynamics of your organization.

3.1DATASET AND ITS DETAIL EXPLANATION AND IMPLEMENTATION OF CUSTOMER SEGMENTATION USING DATA SCIENCE

- Customer segmentation is a technique used to group customers based on their characteristics. It helps organizations understand their customers better and make strategic decisions regarding product growth and marketing1.
- Machine learning can be employed to automate customer segmentation, which can be a tedious task when done manually. There are different methodologies for customer segmentation, and they depend on four types of parameters: geographic, demographic, behavioral, psychological1.
- There are several datasets available online that can be used for customer segmentation. One such dataset is the e-commerce dataset that contains the annual income of approximately 300 customers and the amount they spend annually on an e-commerce website2.
- To implement customer segmentation using machine learning techniques, you can use ensemble techniques such as Support Vector Machine (SVM), Logistics Regression, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, AdaBoost Classifier, and Gradient Boosting Classifier3. You can also use clustering algorithms such as k-means and hierarchical clustering to derive the optimum number of clusters and understand the underlying customer segments based on the data provided

3.2 BEGIN BUILDING THE PROJECT BY LOAD THE DATASET

CUSTOMERIDS.

- Age.
- Gender.
- Annual Income.
- Spending Score.

- To begin building your customer segmentation project using the Mall Customers dataset from Kaggle, you'll need to load the dataset and start exploring it. You can use Python and popular libraries like Pandas for data manipulation and Matplotlib or Seaborn for data visualization. Here's a step-by-step guide:

1. IMPORT NECESSARY LIBRARIES:

- You'll need to import the necessary Python libraries to work with the dataset.

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns
```

2. LOAD THE DATASET:

- Download the Mall Customers dataset from Kaggle and save it as a CSV file in your working directory. You can then load it into a Pandas DataFrame.

```
# Load the dataset

df = pd.read_csv(' C:\Users\CSE\Downloads\Mall_Customers.csv')
```

3. EXPLORE THE DATASET:

- Now, let's start exploring the dataset to understand its structure and the type of information it contains.

```
# Display the first few rows of the dataset
print(df.head())


# Check the basic statistics of the dataset
print(df.describe())


# Check for missing values
print(df.isnull().sum())


# Check the data types of each column
print(df.dtypes)
```

This code will give you an overview of the dataset, including the first few rows, summary statistics, missing values, and data types of each column.

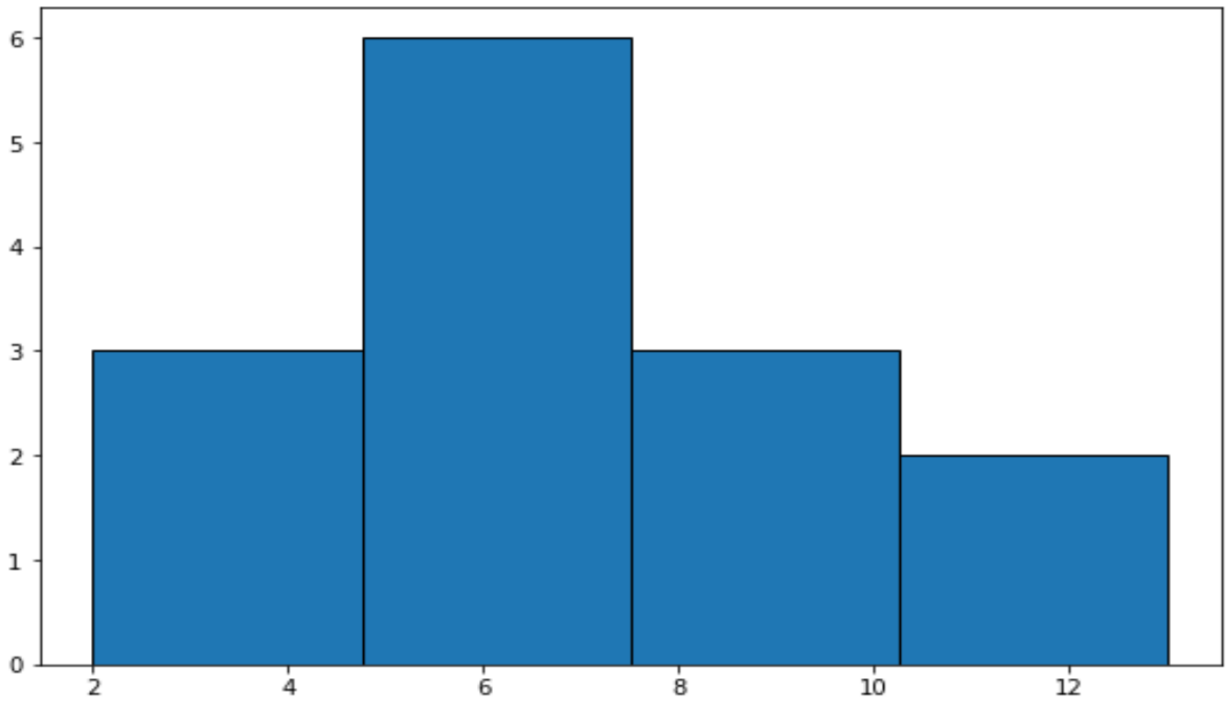
4. DATA VISUALIZATION:

- Data visualization is an important step in understanding the dataset and identifying potential trends and patterns. You can use libraries like Matplotlib and Seaborn for this purpose.

```
# Example: Visualize the distribution of Age
plt.figure(figsize=(8, 6))
sns.histplot(df['Age'], bins=20, kde=True)
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```

You can create various types of plots and visualizations to gain insights into the dataset.

OUTPUT:



- Depending on your analysis goals and the segmentation method you plan to use, you may need to preprocess the data. This could include handling outliers, scaling features, and encoding categorical variables.

6. CUSTOMER SEGMENTATION:

- Once you've explored and preprocessed the data, you can move on to customer segmentation using one of the techniques mentioned earlier (e.g., clustering using K-Means). You may also want to select the features (columns) that are relevant for your analysis.

7. EVALUATE AND INTERPRET:

- After segmentation, evaluate the quality of the segments and interpret the results to gain insights into customer groups. You can use visualization and statistical methods to do this.

From there, you can implement personalized marketing strategies, product recommendations, or any other actions based on the segments you've identified.

8. MONITOR AND ITERATE:

- Keep monitoring your customer segments over time and adjust your strategies as needed to maintain their relevance and effectiveness.

Remember to handle sensitive customer data responsibly and in accordance with data privacy regulations.This is a basic outline to get you started with your project using the Mall Customers dataset. Your specific analysis and segmentation approach may vary depending on your business goals and the insights you want to gain from the data.

3.3 PREPROCESS DATASET

1. IMPORT LIBRARIES:

- First, import the necessary libraries, including Pandas for data manipulation.

```
import pandas as pd
```

2. LOAD THE DATASET:

- Load the dataset from the CSV file. Make sure to download the dataset from Kaggle and place it in your working directory.

```
# Load the dataset

df = pd.read_csv('Mall_Customers.csv')
```

3. EXPLORE THE DATASET:

- Explore the dataset to understand its structure, check for missing values, and review data types.

```
# Display the first few rows of the dataset
print(df.head())


# Check for missing values
print(df.isnull().sum())


# Check the data types of each column
print(df.dtypes)
```

4. HANDLE MISSING VALUES:

- In this dataset, it's possible that there are no missing values. However, if there were any missing values, you'd need to decide how to handle them. Options include dropping rows with missing values, filling them with a default value, or using more advanced imputation techniques.

5. ENCODE CATEGORICAL DATA (IF ANY):

- The Mall Customers dataset doesn't contain categorical variables that need encoding. However, if your dataset had categorical data (e.g., "Gender"), you'd need to encode it, typically using one-hot encoding or label encoding.

6. FEATURE SELECTION:

- Depending on your analysis goals, you may want to select a subset of features for segmentation. For example, you might choose to focus on "Annual Income" and "Spending Score" for clustering customers.

```
# Select specific columns for analysis
selected_columns = ['Annual Income (k$)', 'Spending Score (1-100)']
df = df[selected_columns]
```

7. STANDARDIZE/NORMALIZE DATA (IF NEEDED):

- If you're using clustering algorithms that rely on distances (e.g., K-Means), it's often a good practice to standardize or normalize the data to bring features to the same scale. This can be done using techniques like Min-Max scaling or Z-score standardization.

```
from sklearn.preprocessing import StandardScaler


scaler = StandardScaler()

df_scaled = scaler.fit_transform(df)
```

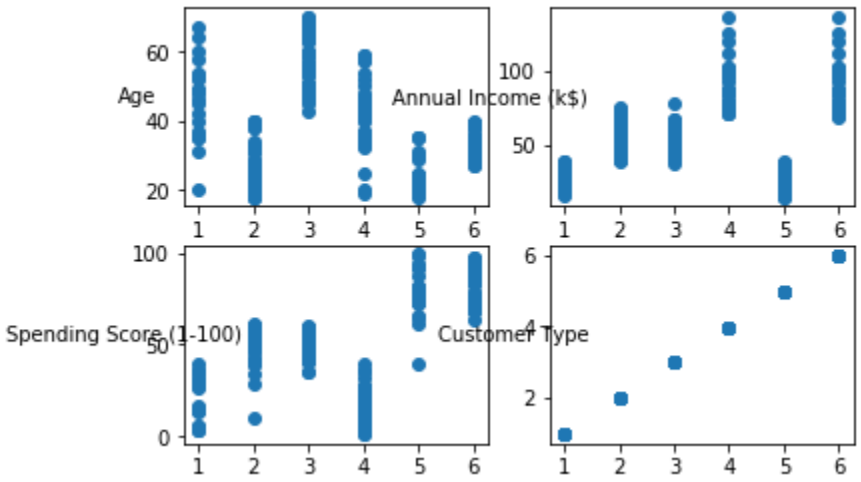
8. SAVE THE PREPROCESSED DATA (OPTIONAL):

- If you want to save the preprocessed data for future use, you can save it to a new CSV file.

```
df_preprocessed = pd.DataFrame(df_scaled, columns=selected_columns)

df_preprocessed.to_csv('mall_customers_preprocessed.csv', index=False)
```

OUTPUT:



The steps mentioned above provide a general outline for preprocessing the Mall Customers dataset. Keep in mind that the specific preprocessing steps can vary based on your analysis goals and the nature of the dataset. Additionally, you can add further preprocessing steps, such as handling outliers or creating new features, depending on your project requirements.

3.4 PERFORMING DIFFERENT ANALYSIS NEEDED

1.DESCRIPTIVE STATISTICS:

- Calculate summary statistics such as mean, median, standard deviation, and percentiles for features like 'Age,' 'Annual Income,' and 'Spending Score.'

2.DATA VISUALIZATION:

- Create various types of plots and visualizations to explore the data. For example:

Histograms to visualize the distribution of features.

Scatter plots to explore the relationship between 'Annual Income' and 'Spending Score.'

Box plots to identify outliers in the data.

Pair plots or correlation matrices to understand the relationships between different features.

3.CUSTOMER SEGMENTATION:

- Implement customer segmentation using clustering algorithms like K-Means, Hierarchical Clustering, or DBSCAN to group customers into different segments based on 'Annual Income' and 'Spending Score.'

Visualize the segments to understand their characteristics.

4.EXPLORATORY DATA ANALYSIS (EDA):

- Conduct EDA to uncover patterns or trends in the data. For example, you can explore the distribution of customers by age and gender or identify correlations between variables.

5.CUSTOMER PROFILES:

- Create customer profiles or personas based on common attributes or behaviors. This can help in tailoring marketing strategies.

6.RFM ANALYSIS:

- Perform RFM (Recency, Frequency, Monetary) analysis to segment customers based on their shopping behavior.

Identify high-value customers or those who might need re-engagement.

7.MARKET BASKET ANALYSIS:

- Analyze product associations by examining which products are frequently purchased together. This can help with product placement and recommendations.

8.CUSTOMER CHURN ANALYSIS:

- Analyze customer churn by tracking changes in customer behavior over time. Identify factors that lead to customers leaving and take steps to retain them.

9.PREDICTIVE MODELING:

- Build predictive models to forecast customer behavior, such as predicting future spending based on historical data or predicting customer churn.

10.CUSTOMER SATISFACTION ANALYSIS:

- Collect and analyze customer feedback and satisfaction data to identify areas for improvement.

A/B Testing:

If applicable, conduct A/B tests on different marketing strategies, product offerings, or store layouts to evaluate their impact on customer behavior.

11.CUSTOMER LIFETIME VALUE (CLV) ANALYSIS:

- Calculate the CLV of each customer to understand their long-term value to the business. This can inform marketing and retention strategies.

12.GEOSPATIAL ANALYSIS (IF LOCATION DATA IS AVAILABLE):

- If the dataset contains location information, you can perform geospatial analysis to understand customer distribution and behavior by region.

13.TIME SERIES ANALYSIS (IF APPLICABLE):

- Analyze time-dependent data, such as customer visits or spending, over time to identify trends and seasonality.

14.CUSTOMER RETENTION ANALYSIS:

- Analyze customer retention rates and understand why some customers continue to visit the mall while others don't.

Remember that the choice of analysis depends on your specific business goals and questions you want to answer. Also, consider combining multiple types of analysis to gain a more comprehensive understanding of your customer data and make informed business decisions.