CUSTOMER SEGEMENTATION

INTRODUCTION

The problem is to implement data science techniques to segment customers based on their behavior, preferences, and demographic attributes. The goal is to enable businesses to personalize marketing strategies and enhance customer satisfaction. This project involves data collection, data preprocessing, feature engineering, clustering algorithms, visualization, and interpretation of results.

In this phase the design to innovation and data flow of customer segementation is going to be done.

DATASET

The data is obtained from https://www.Kaggle.com/data

COLUMNS USED

From Mall_Customers.csv data the following columns are used

- CustomerID
- Gender
- Age
- Annual Income (k$)
- Spending Score (1-100)

LIBRARIES USED

The Python 3 environment comes with many helpful analytics libraries installed and several helpful packages to load.

The essential libraries used in this project are :

- Importing OS (for kaggle inputs)
- Numpy and Pandas libraries
- Matplotlib
- Seaborn

TRAIN AND TEST

Training the dataset by describe(), isnull().sum(), drop(), show(), and by using k-means algorithm we train the data

Testing the data by importing sklearn.cluster from k-means with ensuring the plot range and axis labels producing the k value, scattering the data by kmeans.cluster_centers and producing 3D plot.

REST OF THE EXPLANATIONS

Data Collection

The process involves gathering customer data, which includes information about their purchase history, demographics, and interaction patterns.

Data Preprocessing

The task involves preparing and cleaning data, handling missing values, and converting categorical features into numerical representations.

Feature Engineering

Data preparation and cleaning, handling missing values, and the transformation of categorical features into numerical representations are all part of the task.

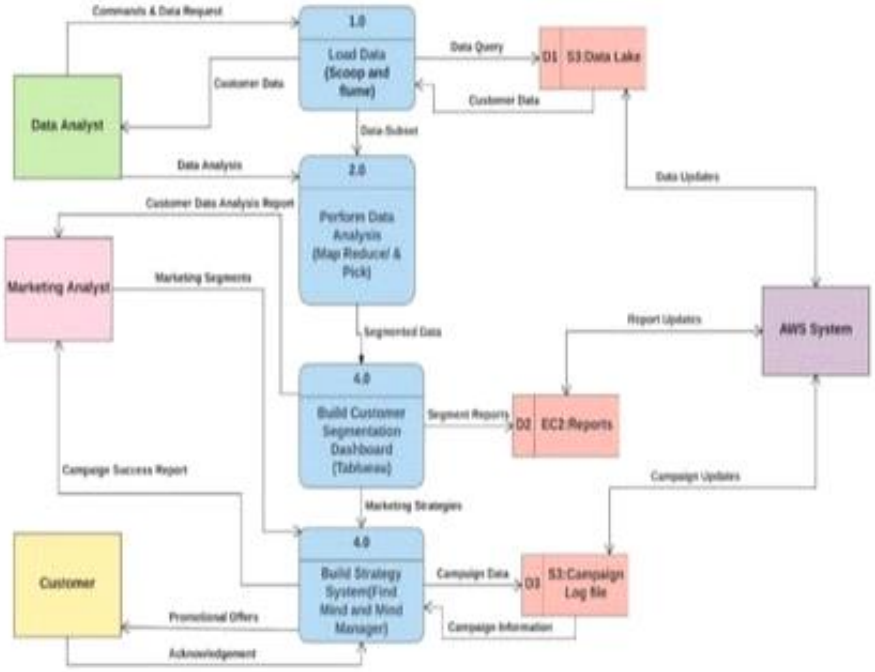
ALGORITHMS USED

Apply clustering algorithms like K-Means, DBSCAN, or hierarchical clustering to segment customers.

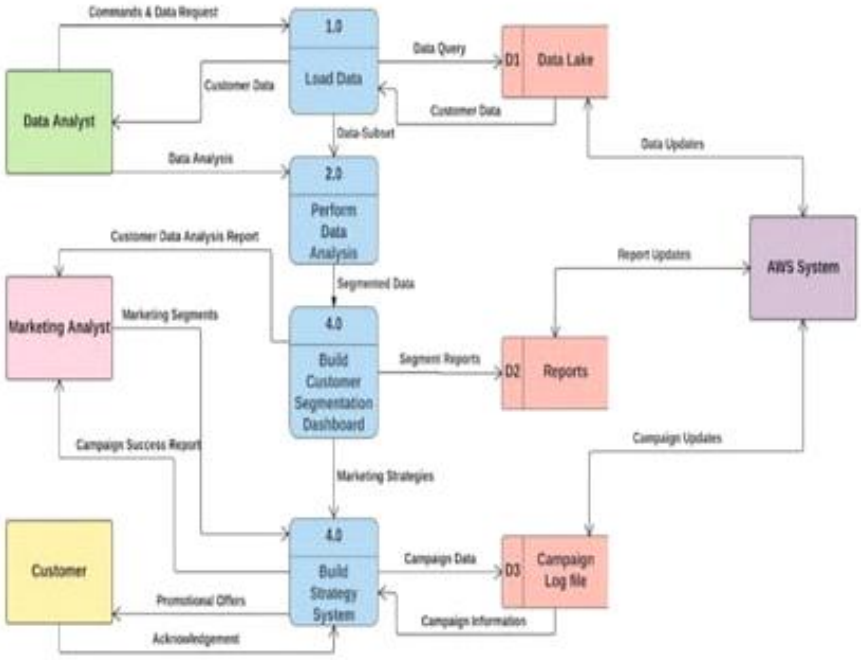
Visualization: Visualize the customer segments using techniques like scatter plots, bar charts, and heatmaps. Interpretation: Analyze and interpret the characteristics of each customer segment to derive actionable insights for marketing strategies.

DESIGN AND DATAFLOW

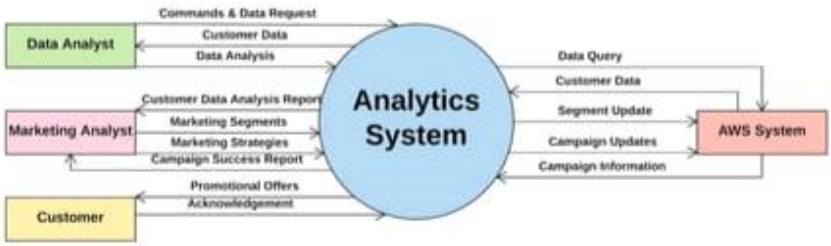
1.Physical data flow diagram:



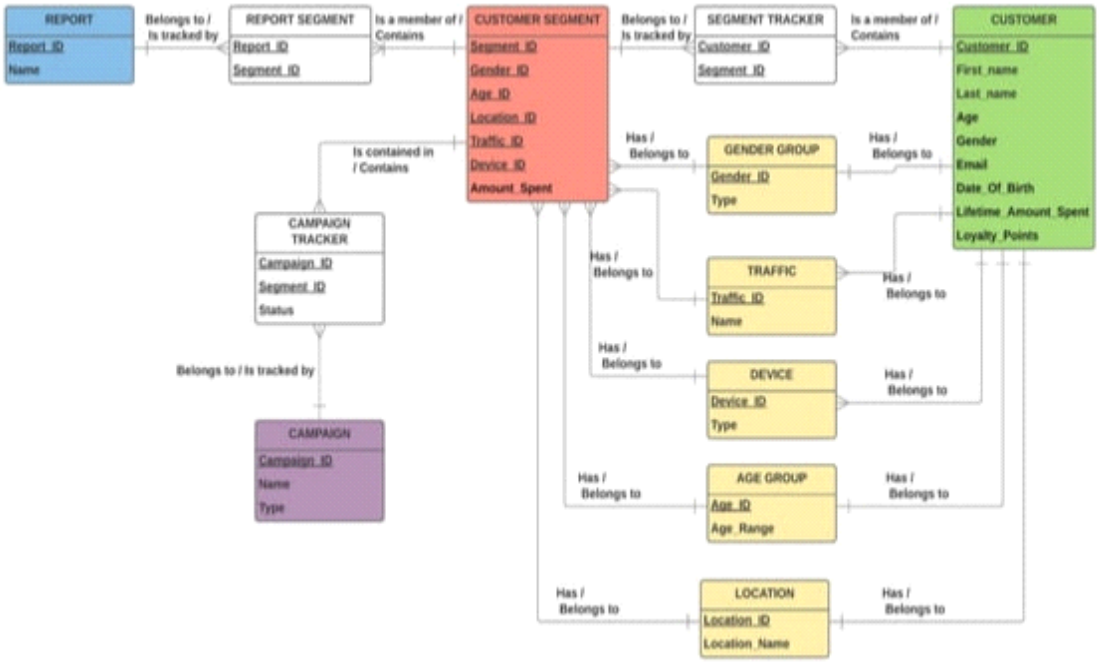
2.Logical data flow diagram:



3.Data flow diagram:



4.Logical entity relational diagram:


REPORT
Report_ID
Name
Belongs to / Is tracked by
REPORT SEGMENT
Report_ID
Segment_ID
Is a member of / Contains
CUSTOMER SEGMENT
Segment_ID
Gender_ID
Age_ID
Location_ID
Traffic_ID
Device_ID
Amount_Spent
Belongs to / Is tracked by
SEGMENT TRACKER
Customer_ID
Segment_ID
Is a member of / Contains
CUSTOMER
Customer_ID
First_name
Last_name
Age
Gender
Email
Date_Of_Birth
Lifetime_Amount_Spent
Loyalty_Points
Is contained in / Contains
CAMPAIGN TRACKER
Campaign_ID
Segment_ID
Status
Belongs to / Is tracked by
CAMPAIGN
Campaign_ID
Name
Type
Has / Belongs to
GENDER GROUP
Gender_ID
Type
Has / Belongs to
Has / Belongs to
TRAFFIC
Traffic_ID
Name
Has / Belongs to
Has / Belongs to
Has / Belongs to
DEVICE
Device_ID
Type
Has / Belongs to
Has / Belongs to
Has / Belongs to
AGE GROUP
Age_ID
Age_Range
Has / Belongs to
Has / Belongs to
Has / Belongs to
LOCATION
Location_ID
Location_Name
Has / Belongs to