

PROJECT - CUSTOMER SEGMENTATION

TEAM MEMBER

210621104012–DHARUNRAJ .N

Phase-1 Document Submission

Problem Definition

The main goal of this project is to use advanced techniques in data science to group customers based on how they behave, what they like, and information about them like their age or location. This grouping strategy is really important because it helps businesses create special marketing plans for each group of customers. This can make customers happier and more likely to keep coming back to the business. The project has several important steps like getting the data, cleaning it up, making new useful information from the data, using different ways to group the customers, showing the data in pictures, and understanding what the groups mean.

Design Thinking

1. Data Collection

Objective: Collect comprehensive customer data encompassing a spectrum of attributes, ranging from purchase history and demographic particulars to interaction behaviour and engagement metrics.

Rationale: A comprehensive dataset is the foundation upon which the subsequent analytical processes will be built. Ensuring the integrity and privacy of this data is paramount throughout the entire collection process.

Dataset Link : <https://www.kaggle.com/datasets/akram24/mall-customers>

1	Customer	Gender	Age	Annual Inc	Spending Score (1-100)		
2	1	Male	19	15	39		
3	2	Male	21	15	81		
4	3	Female	20	16	6		
5	4	Female	23	16	77		
6	5	Female	31	17	40		
7	6	Female	22	17	76		
8	7	Female	35	18	6		
9	8	Female	23	18	94		
10	9	Male	64	19	3		
11	10	Female	30	19	72		
12	11	Male	67	19	14		
13	12	Female	35	19	99		
14	13	Female	58	20	15		
15	14	Female	24	20	77		
16	15	Male	37	20	13		
17	16	Male	22	20	79		
18	17	Female	35	21	35		
19	18	Male	20	21	66		
20	19	Male	52	23	29		
21	20	Female	35	23	98		
22	21	Male	35	24	35		
23	22	Male	25	24	73		
24	23	Female	46	25	5		
25	24	Male	31	25	73		
26	25	Female	54	28	14		
27	26	Male	29	28	82		

Mall_Customers

Ready

2. Data Preprocessing

Objective: Systematically cleanse and preprocess the acquired data to eliminate inconsistencies, outliers, and any potential sources of noise or bias.

Tasks: First, carefully check the dataset to find and fix any problems with the data quality.

Then, use strong methods to deal with missing information, like filling in the gaps with educated guesses or removing the data if it doesn't make sense.

Lastly, change categories (like types of products or colors) into numbers so the computer can understand them better. This can be done in different ways, like making a list of 0s and 1s or giving each category a number.

Rationale: A meticulously cleaned and standardized dataset serves as a reliable foundation for subsequent analytical processes, ensuring the accuracy and reliability of the results.

"Customer ID" is not a significant feature So we'll remove it

```
data.drop(["CustomerID"], axis=1, inplace=True)
```

```
data.isna().sum()
```

```
Gender          0
Age             0
Annual Income (k$)  0
Spending Score (1-100)  0
dtype: int64
```

It seems like we don't have any missing values

3. Feature Engineering

Objective: Make new pieces of information that show different parts of how customers behave and what they like, like how much they spend or how often they buy things.

Tasks: Use what we know about this field to find customer things that matter and make new pieces of information that will help us group them better.

Use math and statistics to find important pieces of information from the raw data.

Rationale: Making these new pieces of information helps us have a more detailed and useful dataset. This will make it easier for the computer to group customers in a smart way.

4. Clustering Algorithms

Objective: Use smart computer methods to group customers based on their characteristics, like what they buy or where they live.

Tasks: Try different ways of grouping customers to see which one works best for this dataset.

Make small changes to how we group customers to get the best results.

Rationale: Using strong grouping methods is really important to make sure the groups we create show real and useful information about customers

We will consider 5 as optimal clusters

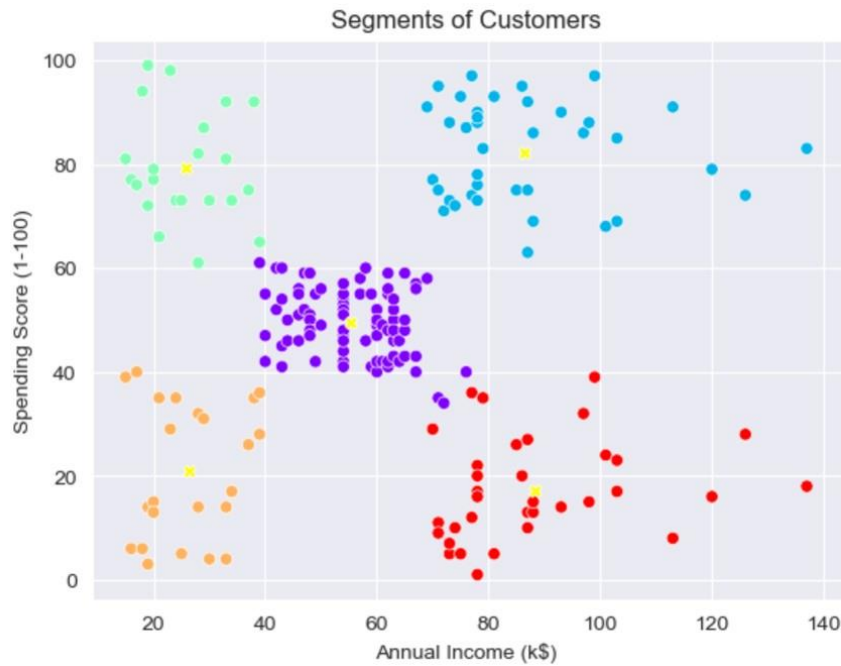
```
kmeans = KMeans(5)
clusters = kmeans.fit_predict(data_new)
data_new["Clusters"] = clusters
data_new.head()
```

	Annual Income (k\$)	Spending Score (1-100)	Clusters
0	15	39	3
1	15	81	2
2	16	6	3
3	16	77	2
4	17	40	3

5. Visualization

Objective: Employ a range of visualization techniques, including scatter plots, bar charts, and heatmaps, to vividly represent the identified customer segments.

Rationale: Visual representations provide an intuitive means of comprehending complex data, enabling stakeholders to grasp the nuances of customer segmentation at a glance.



6. Interpretation

Objective: Conduct a comprehensive analysis and interpretation of the characteristics exhibited by each customer segment.

Tasks: Identify distinguishing traits and behaviours that define each segment.

Derive actionable insights that can be leveraged in tailoring marketing strategies to cater to each segment.

Rationale: The insights gained through interpretation serve as the bedrock for formulating targeted marketing strategies that resonate with each unique customer group.