**CUSTOMER SEGMENTATION USING DATA SCIENCE**

**AGENDA:**

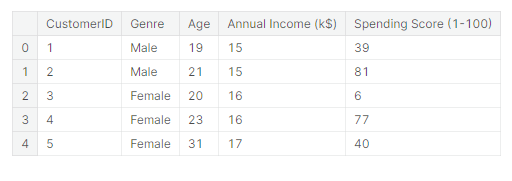
* Problem definition
* Design thinking
* Data collection
* Data preprocessing
* Feature engineering
* Clustering algorithms [K-Means]
* Visualizations
* Interpretation
* **PROBLEM DEFINITION:**

 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.

* **DESIGN THINKING:**

1. DATA COLLECTION:

Gather relevant customer data, which can include demographic information, purchase history, website behavior, and so on…



1. DATA PREPROCESSING:

Clean, format, and prepare the data for analysis. This may involve handling missing values, outliers, and transforming data into a usable format.

1. FEATURE ENGINEERING:

 Identify the most important features (variables) for segmentation. You may need to create new features that capture customer behavior or characteristics better.

1. CLUSTERING ALGORITHM:

K-Mean

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. In this topic, we will learn what is K-means clustering algorithm, how the algorithm works, along with the Python implementation of k-means clustering.

**PROGRAM:**

In[1]:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

dataset = pd.read\_csv('/kaggle/input/mall-customers/Mall\_Customers.csv')

dataset.shape

Out[1]:

(200, 5)

In[2]:

dataset.info()

Out[2]:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 200 entries, 0 to 199

Data columns (total 5 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 CustomerID 200 non-null int64

1 Genre 200 non-null object

2 Age 200 non-null int64

3 Annual Income (k$) 200 non-null int64

4 Spending Score (1-100) 200 non-null int64

dtypes: int64(4), object(1)

memory usage: 7.9+ KB

1. VISUALIZATION:

Another way to communicate your customer segments is to visualize them and their relationships using charts, graphs, or maps. Visualization can

help you highlight the similarities and differences between your clusters, as well as their patterns and trends over time or across dimensions

**PROGRAM:**

In[3]:

*#Visualizing the clusters of KMeans*

label = final\_model.fit\_predict(train\_data)

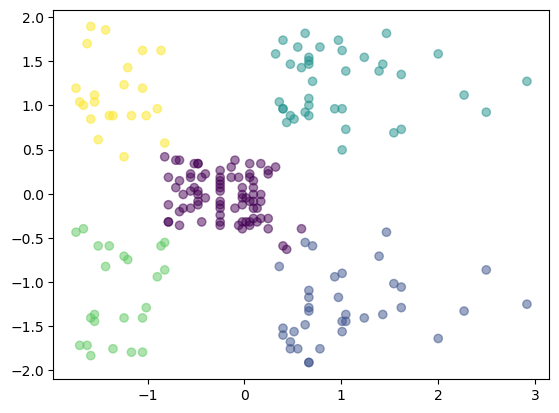
xs = train\_data[:,0]

ys = train\_data[:,1]

plt.scatter(xs,ys,c=label,alpha=0.5)

Out[3]:

<matplotlib.collections.PathCollection at 0x755c71ea6450>

****

1. INTERPOLATION:

*Interpretation* is the act of explaining, reframing, or otherwise

showing your own understanding of something. A person who translates one

language into another is called an interpreter because they are explaining what a person is saying to someone who doesn't understand. Interpretation requires you

to first understand the piece of music, text, language, or idea, and then give your explanation of it. A computer may produce masses of data, but it will require

your interpretation of the data for people to understand it.