TEAM MEMBER

732521104030 : NAVEEN.S

Phase 4Submission Document

Project :**Customer Segmentation using Data Science**

**Problem Definition:**

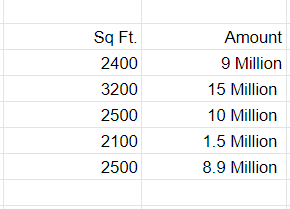
The problem is to implement data science techniques to segment customers based on their behavior, preferences, and demographic attributes. The goal is to enable businesses to personalize marketing strategies and enhance customer satisfaction. This project involves data collection, data preprocessing,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

Data s**et Link:**[**https://www.kaggle.com/datasets/akram24/mall-customers**](https://www.kaggle.com/datasets/akram24/mall-customers)

**Feature engineering:**



*f****eature engineering is a machine learning technique that leverages data to create new variables that aren’t in the training set. It can produce new features for both supervised and unsupervised learning, with the goal of simplifying and speeding up data transformations while also enhancing model accuracy. Feature engineering is required when working with machine learning models. Regardless of the data or architecture, a terrible feature will have a direct impact on your model.***



# Applying clustering algorithms :

* ***Choose the number of clusters k. The first step in k-means is to pick the number of clusters, k.***
* ***Select k random points from the data as centroids. ...***
* ***Assign all the points to the closest cluster centroid. ...***
* ***Recompute the centroids of newly formed clusters. ...***
* ***Repeat steps 3 and 4***.

**Visualization :**

* ***The physical or imagining information.creation of images, diagrams, or animations to communicate a message. Data and information visualization, the practice of creating visual representations of complex data***

**(1) outcome visualization:** [envisioning the desired future end-point](https://www.betterup.com/blog/career-planning), and

**(2) process visualization:**envisioning every step toward that desired outcome. With all the senses engaged

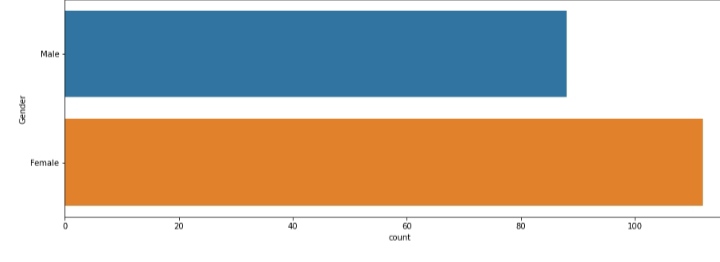
**Interpretation :**

* ***The basic definition of*interpretation*, according to the Webster dictionary is the “action of explaining the meaning of something; the way something is explained or understood.”***
* ***In terms of language, the definition of*interpretation*should be broader: rendering a spoken or signed message into another spoken or signed language, preserving the register and meaning of the source language content***.
* ***It is the spoken or signed language communication between users of different languages. A language interpreter or sign language interpreter must not only quickly and carefully interpret meaning, but also tone and intent of the original message into the target or interpreted language***

.

**Feature Engineering**: Create additional features that capture customer behavior and preferences, such as total spending, frequency of purchases, etc.

**Barchart:**



***Heatmaps*** : A heatmap is **a graphical representation of data that uses a system of color coding to represent different values**. Heatmaps are used in various forms of analytics but are most commonly used to show user behavior on specific web pages or webpage templates.

**Interpretation:**

Analyze and interpret the characteristics of each customer segment to derive actionable insights for marketing strategies.

Program:

import numpy as np linear algebra

import pandas as pd

import os

for dirname,. filenames in os.walk(' C:\Users\Students\Documents\New folder\Mall Customers '):

for

filename in filenames:

print(os.path.join(dirname, filename))

import numpy as np

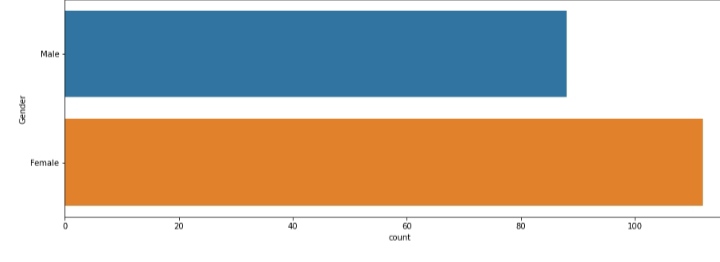
import pandas as pd import matplotlib.pyplot as plt

import seaborn as sns

df pd.read\_csv('/kaggle/input/mall-customers/Mall Customers.csv')

df.rename(columns=('Genre: 'Gender' }, inplace=True)

df.head()

**output:** 

plt.figure(1,figsize=(15,6))

n = 0

for x **in** ['Age','Annual Income (k$)','Spending Score (1-100)']:

n +=1

plt.subplot(1,3,n)

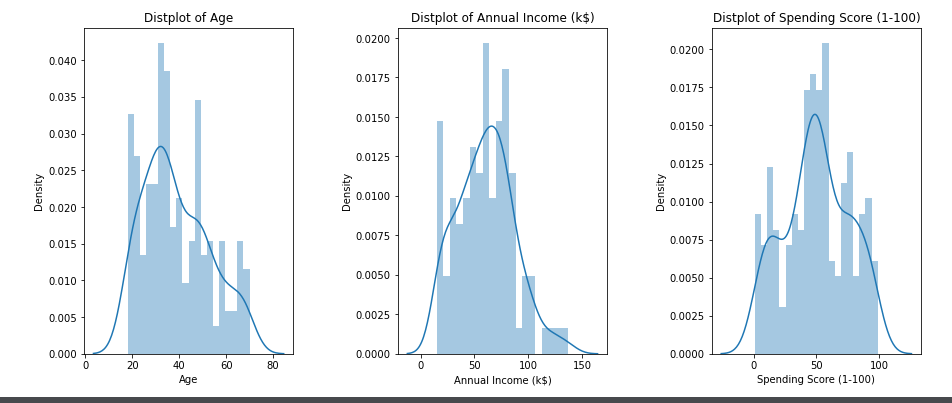
plt.subplots\_adjust(hspace=0.5,wspace=0.5)

sns.distplot(df[x],bins=20)

plt.title('Distplot of **{}**'.format(x))

plt.show()

**output :**

****

**plt.figure(1,figsize=(15,6))**

**n = 0**

**for** **cols in ['Age','Annual Income (k$)','Spending Score (1-100)']:**

**n +=1**

**plt.subplot(1,3,n)**

**sns.set(style="whitegrid")**

**plt.subplots\_adjust(hspace=0.5,wspace=0.5)**

**sns.violinplot(x = cols,y = 'Gender',data=df)**

**plt.ylabel('Gender' if n== 1 else '')**

**plt.title('Violin Plot')**

**plt.show()**

**OUTPUT:**

