# **CUSTOMER SEGMENTATION USING DATA SCIENCE**

**PHASE:3**

**DATA:**

Customer segmentation is a common application of data science, and it involves dividing your customer base into distinct groups based on various characteristics.

Here are the general steps to perform customer segmentation using a data science model:

**\* Data Collection:** Start by collecting relevant data about your customers. This data may include demographic information, purchase history, website activity, and more.

**\* Data Pre-processing:**

a. Data Cleaning: Remove any duplicate, missing, or irrelevant data.

b. Data Transformation: Convert categorical data into numerical format, if necessary, using techniques like one-hot encoding.

c. Feature Scaling: Normalize or standardize numerical features to ensure they have a similar scale.

d. Feature Engineering: Create new features if they can provide valuable insights.

**\* Exploratory Data Analysis (EDA):** Analyze the data to gain insights into customer behavior and characteristics. This step can involve data visualization and summary statistics.

**\* Customer Segmentation Algorithm Selection:** Choose an appropriate clustering algorithm. Common choices include K-Means, hierarchical clustering, and DBSCAN.

**\* Feature Selection:** Decide which features to include in the clustering analysis. Not all features may be relevant.

**\* Model Training:** Use the selected algorithm to train the model on the preprocessed data.

**\* Segmentation**: Apply the trained model to cluster your customers. Each cluster represents a segment with similar characteristics.

**\* Evaluation**: Evaluate the quality of your segmentation using metrics like silhouette score, Davies-Bouldin index, or domain-specific criteria.

**\* Interpretation**: Analyze the characteristics of each segment to understand their behaviors and needs.

**\* Actionable Insights:** Based on your analysis, develop strategies and marketing campaigns tailored to each segment’s preferences and needs.

**\* Monitoring and Refinement:** Continuously monitor and refine your segmentation as new data becomes available.

**You can use popular Python libraries like Numbly, pandas, sickie-learn, and Marplot/Seaborne for data preprocessing and analysis. Here’s a simplified example using Python:**

Import pandas as pd

# Load the dataset

Data = pd.read\_csv(‘customer\_data.csv’)

# Data pre-processing steps

# 1. Handle missing values

Data.dropna(emplace=True)

# 2. Feature selection/engineering

# For example, create a new feature ‘total purchase’ from ‘purchase history’ data.

# 3. Categorical encoding (if needed)

# Use one-hot encoding or label encoding for categorical variables.

# 4. Normalize/Scale numerical data

From sklearn.preprocessing import Standardisable

Scalar = Standardisable()

Data[‘numerical column’] = scaler.fit\_transform(data[‘numerical\_column’])

# Now, the data is ready for customer segmentation analysis.