# CUSTOMER SEGMENTATION USING DATA SCIENCE

Name: **Yerukala Manjunath**

[au723921244060]

Project Name: **CUSTOMER SEGMENTATION USING DATA SCIENCE**

**Phase-3: Development-1**

**Topic: In this part you will begin building your project by loading and pre processing the Dataset**

INTRODUCTION:

Certainly, to begin a customer segmentation project, you'll need to collect and preprocess the customer data. Here's a step-by-step guide on how to do this:

1. Define the Purpose of Segmentation:

Start by understanding the goal of your customer segmentation. Why are you segmenting customers? Is it for targeted marketing, product recommendations, or improving customer service? Knowing the purpose will guide your data collection and preprocessing.

2. Data Collection:

Collect relevant data on your customers. This data can come from various sources, such as:

Customer Records: This includes information like name, contact details, purchase history, etc.

Demographic Data: Age, gender, location, income, occupation, etc.

Behavioural Data: Online behaviour, transaction history, website visits, etc.

Psychographic Data: Interests, lifestyle, values, and opinions.

Feedback and Surveys: Customer feedback and responses to surveys.

Social Media Data: If applicable, data from social media platforms.

3. Data Cleaning:

Before you start analysis, it's crucial to clean the data. This involves:

- Handling missing values.

- Removing duplicates.

- Correcting inconsistencies and errors.

- Standardizing data formats (e.g., date formats, currency symbols).

4. Data Integration:

If you collected data from various sources, you might need to integrate them into a single dataset. Ensure that the data is in a format that's easy to work with.

5. Feature Engineering:

Create new features that might be valuable for segmentation. For example, you could calculate customer lifetime value, recency, and frequency metrics for RFM (Recency, Frequency, Monetary) analysis.

6. Data Transformation:

Depending on your analysis method, you might need to transform the data. This can include normalizing or standardizing numerical features or encoding categorical variables.

7. Data Exploration:

Before diving into segmentation, it's essential to understand your data. Visualize key statistics, create histograms, scatter plots, and other exploratory data analysis (EDA) techniques to gain insights.

8. Segmentation Analysis:

Perform the actual segmentation using appropriate techniques. Some common methods include:

- RFM Analysis: Segments customers based on their recency, frequency, and monetary value.

- K-Means Clustering: Unsupervised clustering based on similarities in customer behaviour or attributes.

- Hierarchical Clustering: Agglomerative or divisive clustering methods.

- Decision Trees or Random Forests: Supervised methods to segment customers based on predictors.

- Principal Component Analysis (PCA): For dimensionality reduction and feature selection.

9. Evaluate and Validate Segments:

It's important to assess the quality and effectiveness of your customer segments. Use relevant metrics to measure how well they meet your project's goals.

10. Documentation:

Document all your data preprocessing steps, analysis techniques, and results. This documentation will be invaluable for future reference and collaboration.

Remember to respect data privacy and adhere to any relevant data protection laws when collecting and processing customer data. Additionally, regularly update your segmentation as customer behaviour and preferences change over time.

**Importance of Loading and pre-processing datasets**

Loading and pre-processing datasets is a fundamental and critical step in data analysis, machine learning, and data-driven decision-making. Here are several key reasons for the importance of these tasks:

**Data Quality Assurance:** Loading and pre-processing data allow you to inspect and ensure the quality of the dataset. You can identify and address issues like missing values, outliers, and inconsistencies, which are essential for reliable and accurate analysis.

**Data Understanding:** Loading the data provides an initial understanding of its structure, size, and format. It helps you become familiar with the dataset, which is necessary for planning your analysis, choosing the right techniques, and making informed decisions.

**Data Cleaning:** Pre-processing involves data cleaning, which includes handling missing data, removing duplicates, and addressing outliers. Clean data is vital for obtaining meaningful and reliable insights, as well as for building robust machine learning models.

**Feature Engineering:** Pre-processing often involves feature engineering, where you create, transform, or select features to enhance the predictive power of your models or to gain better insights in data analysis.

**Submitted by…..**

Yerukula Manjunath