**Project 8: Customer Segmentation using Data Science**

T. Mohith Shakthi (411621104033)

Mohanrangan (411621104032)

Shaik Sharukh Ahmed (411621104048)

Krishnaraj (411621104027)

K. Salman Khan (411621104047)

**Phase 1: Project Definition and Design Thinking**

**Introduction:**

In the rapidly evolving landscape of business and technology, the importance of understanding and catering to customer needs cannot be overstated. As we are college students passionate about the intersection of data science and marketing, we embark on a project aimed at leveraging data-driven insights to enhance marketing strategies. The focus is on customer segmentation, a pivotal aspect that enables businesses to tailor their approaches, fostering a more personalised and satisfying customer experience.

**Problem Definition:**

In a dynamic landscape of business, understanding customers is vital for effective marketing strategies. Our project aims to employ data science techniques for customer segmentation, focusing on behaviour, preferences, and demographic attributes. Traditional one-size-fits-all approaches often fail to resonate with individual customers' unique characteristics, prompting the need for a data-driven strategy. The objective of our project is to overcome the limitations of generic marketing strategies by dissecting the diverse customer base into distinct groups. Our goal is to identify patterns within customer data and formulate personalised marketing strategies to enhance overall customer satisfaction.

**Design Thinking:**

The design thinking process outlines a structured approach to solving the problem of customer segmentation for personalized marketing. Each step is iterative, allowing for continuous learning and improvement. By understanding the needs of customer, defining data requirements, ideating creative solutions, and prototyping and testing various components, using this process we aims to deliver a robust and effective customer segmentation solution.

**Data Collection:**

We have collected relevant data on customer behaviour, preferences, and demographic attributes from various sources, such as customer databases, online interactions, surveys, and third-party data providers.

This may include:

**Behavioural Data:** Purchase history, website interactions, and responses to marketing campaigns.

**Preference Data:** Surveys, feedback, reviews.

**Demographic Data**: Age, gender, location, income.

We have Ensured data privacy and compliance with regulations like GDPR. We anonymize and aggregate data to protect customer privacy.

**Data Pre-processing:**

We clean and pre-process the data collected to ensure its quality and usability. Steps may include:

**Handling missing data:** Impute or remove missing values.

**Removing duplicates:** Ensure each customer is represented once.

**Data transformation:** Normalise numerical features, encode categorical variables.

**Feature Engineering:**

We can create new features or transform existing ones to extract more meaningful information. For example, you might create features like:

**RFM (recency, Frequency, and Monetary) features:** Calculate these metrics based on customer transaction history.

**Demographic features**: Create categories or bins for age, income, or other demographic attributes.

**Preference features:** Incorporate sentiment analysis scores from customer reviews or feedback.

**Clustering Algorithms:**

We must conduct EDA (Exploratory Data Analysis) to understand the data distribution, identify outliers, and visualise relationships between variables. Use summary statistics, histograms, scatter plots, and correlation matrices.

We are applying clustering algorithms to group or segment customers based on their attributes. Common clustering methods include:

**K-Means Clustering:** Divide customers into K clusters based on similarities.

**Hierarchical Clustering:** Build a hierarchical tree of clusters.

**DBSCAN:** Identify dense regions of data points as clusters.

**Model Evaluation:**

Evaluate the performance of the clustering models using appropriate metrics. Adjust the number of clusters (K) if needed.

**Visualisation:**

We are visualising the clusters to provide a clear representation of customer segments. Examples are:

**Scatter Plots:** Visualise clusters using scatter plots to understand the distribution of customers.

**Heat-maps:** Use heat-maps to showcase the relationships between different features.

**Dimensionality Reduction:** Apply techniques like PCA or t-SNE for a concise representation of clusters.

**Interpretation of Results:**

We analyse the characteristics of each cluster to understand what distinguishes one segment from another. We identify common behaviours, preferences, and demographic profiles within each cluster. We identify actionable insights for personalised marketing strategies.

**Conclusion:**

In this design thinking process we have outlined a structured approach for solving the problem of customer segmentation for personalised marketing. Each step is iterative, allowing for continuous learning and improvement. By understanding the needs of customers, defining data requirements, ideating creative solutions, and prototyping and testing various components, in this process we aims to deliver a robust and effective customer segmentation solution.