

# Customer Segmentation

## Team Members:

- Krishna Koushik Parimi
- Rupa Dinesh Lnu
- Ravi Tarun Prasad Nimmalapudi

## Objective:

Customer segmentation is the process of dividing customers into groups based on shared characteristics. The objective of this customer segmentation project is to use machine learning techniques to categorize online retail customers into distinct and meaningful groups based on their purchasing behavior and characteristics. The goal is to enhance the understanding of the customer base, identify trends and behaviors, and develop personalized marketing efforts for each segment. This segmentation will, in turn, lead to improved customer satisfaction and increased profitability for the online retail business.

## Problem Statement:

The challenge of this project is to effectively analyze the transactional trends of different products in a particular time frame that are bought by people residing in different locations. So, coming to **why segment customers?**

From a business perspective, mainly to get a **competitive market edge** and moreover putting people into categories is only one aspect of consumer segmentation. Customers are better understood when they are segmented, and you can use this knowledge to develop content that addresses the demands and difficulties of each section.

## Data Sources:

The dataset, titled "Online Retail," is sourced from the UCI Machine Learning Repository, providing a valuable collection of transnational data capturing transactions from December 1, 2010, to September 9, 2011, for a UK-based non-store online retail business specializing in unique all-occasion gifts. The data offers a glimpse into the actual operations and customer interactions of this retail company, shedding light on purchase behaviors, transaction details, and customer demographics.

## **Algorithm / Solution Technologies:**

We thought of using appropriate clustering machine learning algorithms for this customer segmentation problem, such as K-Means, Hierarchical Clustering and other unsupervised learning methods etc. (if needed). Exploratory data analysis will be performed to ensure that the data has been appropriately pre-processed.

## **Risks:**

The data we are dealing with has approximately more than 5 lakh records. So, while dealing with this huge load of data, the data needs to be cleaned and preprocessed before analyzing it as it might have a lot of anomalies which might lead to inaccurate results. Incomplete, inaccurate, or inconsistent data can lead to biased analysis and incorrect conclusions. Ensuring data quality through proper validation, cleaning, and preprocessing is essential.

## **Challenges:**

As mentioned above, we would be dealing with a huge volume of data. Handling and analyzing large volumes of data efficiently can be challenging and can be computationally intensive. Processing, cleaning, aggregating, and extracting patterns from such a large dataset can be time-consuming. Efficiently representing and summarizing data while maintaining visual clarity is a common challenge in real-time.

## **Citations:**

<https://www.kaggle.com/datasets/vichoudhary7/customer-segmentation-tutorial-in-python>

Ref - We came across a similar problem of customer segmentation, but this data only contained 200 observations or so, which we felt wouldn't be enough for a thorough analysis.