CUSTOMER SEGMENTATION

# Abstract:

Customer segmentation is a process of separating customers into groups of individuals that fall under same category for marketing, such as age, gender, interests and spending habits. Customer segmentation helps to customize its services with the customers for more precision.

Customer segmentation reminds us the fact that all individual customer is different from the other and helps to divide customers fall under same or similar category. Companies also hope to gain a deeper understanding of their customers' preferences and needs with the idea of discovering what each segment finds most valuable to more accurately tailor marketing materials toward that segment.

The algorithm used for customer segmentation is Clustering which falls under the unsupervised learning. Unsupervised learning is a type of Machine Learning that looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision. In contrast to supervised learning that usually makes use of.

# Introduction:

Clustering is the task of dividing the population or data points into several groups, so that the data points in the same groups are more similar to other data points in the same group than those of other groups. In simple words, the goal is to segregate groups with similar traits and assign them to clusters.

Cluster analysis is a class of techniques that are used to classify objects or cases into relative groups called clusters. Cluster analysis is also called classification analysis or numerical taxonomy. In cluster analysis there is no prior information about the group or cluster membership for any of the objects.

# Problem:

Companies want to increase their customers or users and should satisfy their needs without compromising. Customer segmentation helps to categorize the customer or user based in their interest or needs into categories so, that customers fall under same category would enjoy the company services.

Customer segmentation is the subdivision of a market into discrete customer groups that share similar characteristics. Customer segmentation can be a powerful means to identify unsatisfied customer needs.

# K-means clustering:

K-means clustering is a type of unsupervised learning, which is used when you have unlabelled data. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable k. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points to one of K groups based on the feature that are provided. Data points are clustered based on feature similarity. The results of the means clustering algorithm are:

1.The centroids of the K cluster, which can be used to label new data

2.Labels for the training data (each data point is assigned to a single cluster)

Rather than defining groups before looking at the data, clustering allows you to find and analyse the groups that have formed organically. Each centroid of a cluster is a collection of feature values which define the resulting groups. Examining the centroid feature weights can be used to qualitatively interpret what kind of groups each cluster represents.

## The approach K-means follows to solve the problem is called Expectation Maximization:

1. Specify number of clusters K.
2. Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
3. Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn’t changing.
4. Compute the sum of the squared distance between data points and all centroids.
5. Assign each data point to the closest cluster (centroid).
6. Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.

# Final Goal:

The goal of clustering is to maximize the similarity of observation within the cluster and maximize the dissimilarity between the clusters.

# Data Set:

Companies employing customer segmentation operate under the fact that every customer is different and that their marketing efforts would be better served if they target specific, smaller groups with messages that those consumers would find relevant and lead them to buy something. Companies also hope to gain a deeper understanding of their customers' preferences and needs with the idea of discovering what each segment finds most valuable to more accurately tailor marketing materials toward that segment.

1. Customer ID: It is the unique ID given to a customer

2. Gender: Gender of the customer

3. Age: The age of the customer

4. Annual Income (k$): It is the annual income of the customer

5. Spending Score: It is the score (out of 100) given to a customer by the mall authorities, based on the money spent and the behaviour of the customer.