**Course code : CMPE-257 Sec 99, Group name : Insurgents\_Sales Prediction**

**Chaitrali Deshmukh(012501489)**

**Prathyusha Kasibhatta(013774163)**

**Sandhya Gadgoli(013798759)**

**Saylee Vyawahare(013801294)**

The dataset contains product information for different supermarkets and their sales records. The business objective is to cluster the products into groups according to their visibility and sales so that supermarket can predict their future sales for specific products.

We followed below steps:

1. Data preparation

2. Applying clustering algorithms

3. Analyzing the results based on clusters

**Data preparation:**

1. Read and parse the dataset csv file using the pandas Data frame.

2. Removed unnecessary columns and invalid/blank rows.

**Applying clustering algorithms:**

We applied elbow method and decided to group the data into five clusters. We applied following clustering algorithms using scikit.

1. K-Means

2. GMM

3. Agglomerative

4. DB Scan

5. Ward Hierarchical clustering

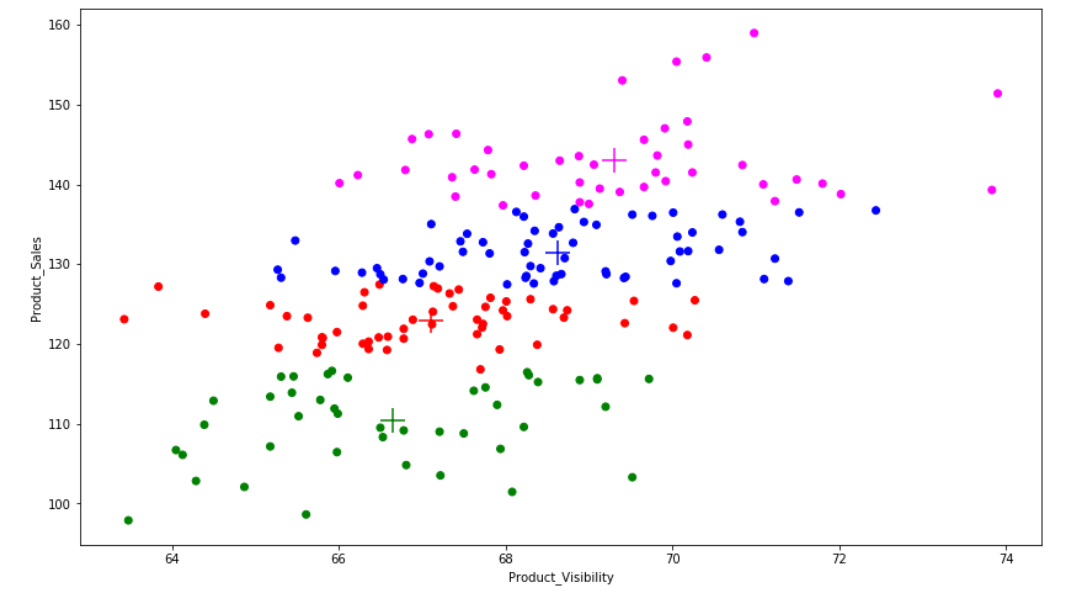
6. Mean shift

7. K-means with Tensor Flow

Algorithm Implementation:

So, here we are implementing k-means which initially randomly selects the data points and tries to converge into different clusters by grouping the similarities between different data points in the given dataset.

**Data Visualization**



So here we can see that k-means algorithm has grouped our data into 4 different clusters.

Conclusion of the data from the graph is as below:

Green cluster: Products having low visibility also has a low sale.

Red cluster: Products having moderate visibility has low-to-moderate sale.

Blue cluster: Products having high visibility has more than average amount of sale.

Pink cluster: Products having high visibility also has a high sale.

This conclusion gives a clarity that product visibility can play a vital role and can have huge amount of impact on the overall sales of the company.