**A case study on Clustering**

A supermarket called K-mart wants to analyse its customers. It wants to understand its customers, their spends, their marketing campaigns and make it easier for them to take business decisions and modify products and strategies accordingly.

This process is called customer personality analysis. It helps businesses make decisions like instead of spending money to market a new product to every customer in the company’s database, a company can analyze which customer segment is most likely to buy the product and then market the product only on that particular segment.

When analysing customers, its also valuable to cluster the customers into meaningful clusters so that we can make our decisions accordingly and make customer cluster specific decisions. We want to be in a position where we can try different strategies for different customer clusters. You are the official Data scientist of the K-mart and are given the task to do the customer personality analysis and also do the clustering of the customers.

*People*

**ID**: Customer's unique identifier

**Year\_Birth:** Customer's birth year

**Education:** Customer's education level

**Marital\_Status:** Customer's marital status

**Income:** Customer's yearly household income

**Kidhome:** Number of children in customer's household

**Teenhome:** Number of teenagers in customer's household

**Dt\_Customer:** Date of customer's enrollment with the company

**Recency:** Number of days since customer's last purchase

**Complain:** 1 if the customer complained in the last 2 years, 0 otherwise

*Products*

**MntWines:** Amount spent on wine in last 2 years

**MntFruits:** Amount spent on fruits in last 2 years

**MntMeatProducts:** Amount spent on meat in last 2 years

**MntFishProducts:** Amount spent on fish in last 2 years

**MntSweetProducts:** Amount spent on sweets in last 2 years

**MntGoldProds:** Amount spent on gold in last 2 years

*Promotion*

**NumDealsPurchases:** Number of purchases made with a discount

**AcceptedCmp1:** 1 if customer accepted the offer in the 1st campaign, 0 otherwise

**AcceptedCmp2:** 1 if customer accepted the offer in the 2nd campaign, 0 otherwise

**AcceptedCmp3:** 1 if customer accepted the offer in the 3rd campaign, 0 otherwise

**AcceptedCmp4:** 1 if customer accepted the offer in the 4th campaign, 0 otherwise

**AcceptedCmp5:** 1 if customer accepted the offer in the 5th campaign, 0 otherwise

**Response:** 1 if customer accepted the offer in the last campaign, 0 otherwise

*Place*

**NumWebPurchases:** Number of purchases made through the company’s website

**NumCatalogPurchases:** Number of purchases made using a catalogue

**NumStorePurchases:** Number of purchases made directly in stores

**NumWebVisitsMonth:** Number of visits to company’s website in the last month

Steps involved in the case study:

1. Exploratory Data analysis
2. Handling nulls and datatypes
3. Analysing people
4. Learning Box plot and distplot
5. Analysing products
6. Analysing promotions
7. Data pre-processing
8. Data scaling
9. Z score transform
10. Dimensionality reduction
11. K means clustering
12. Elbow method
13. Implementing k means in python
14. Visualising results