



ORGANIC & NATURAL FOOD PRODUCTS CUSTOMER SEGMENTATION

Vasantharaj Tejas

ABOUT PROJECT

- i. Client is one of the largest United States super market chain by revenue and the second-largest general retailer in the country.
- ii. Client want to employ customer data, predictive analytics and data-driven strategy to drive sales growth and customer loyalty.
- iii. They lack sophisticated, tools and technology to turn customer data into knowledge, hence, enquired for data science solutions.

Business Objective:

- a) Client want to quickly becoming a top destination for customers who are looking for affordable and accessible organic and natural foods.
- b) Hence, wants to understand their customers better, i.e. segments/groups for better targeting, in simple words, best groups based on common characteristics for better targeting.
- c) So, business objective is to provide effective groups or best customer segmentations from their historical data.

A QUICK DATA UNDERSTANDING

- i. Client has provided the organic and natural products customer transactional data for the period of one year that consists of some 20-30 columns.
- ii. However, we identified key seven variables for next level of analysis. Below is a snapshot of them.
- iii. First quick preparation was with respect to Order date, by converting it to date format.
- iv. Further, due to some double entries understood exist few negative values for quantity ordered and same has been removed.

Variables	Number of Non Missing Values	Unique Values
Customer ID	32275	1754
Order No	32275	22497
Product Code	32275	2935
Quantity Ordered	32275	1381
Unit Price	32275	3740
Order Date	32275	339
Classification Code	32275	82

DATA UNDERSTANDING AND APPROACH

- i. Since, our business objective is to find appropriate customer segments for the given data, a closer look provides that the product's unit price, ordered quantity and date for each customer helps us in achieving the required.
- ii. Order number, product code and classification code are mere identifications ones for present.
- iii. As, from the above information of price and ordered quantity of each product, getting the average amount spent by each customer and classifying them into certain brackets can provide different groups based on their average value but it is crude way of arriving at segments.
- iv. Hence, decided to use Unit Price, Ordered Quantity and Ordered Date variables to find out how recency customer has purchased product, how often they purchase and how much customer spends along with their average spent.
- v. Thus, derived RFM (Recency, Frequency and Monetary) variables.
- vi. And, later employed advanced machine learning technique namely SOM based clustering for arriving customer segments.

DATA PREPARATIONS

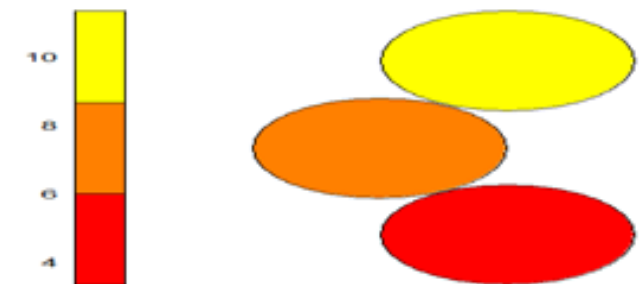
- i. Generation 'Recency' variable - Convert date format to numerical ranks and sort them to ascending order to get least recency value which signifies the recent visit of the customer.
- ii. Generate Frequency variable- multiple visits on a day are considered as 1 frequency, so consider distinct customer visits.
- iii. Create new variable 'Amount', i.e. total order per transaction by multiplying Unit Price and Quantity Ordered.
- iv. Generate Monetary variable - Sum of all transactional amount for each customer.
- v. Remove all duplicates and got the final data for the model with Customer ID , Recency, Frequency and Monetary.

DATA MODELING & RESULTS

- ✓ SOM Clustering Technique (also known as Kohonen Technique) - maps multidimensional data into lower dimensional subspaces where geometric (distance) relationships between points indicate their similarity.
- ✓ Therefore, we have subjected our three generated RFM variables for obtaining distinct groups based on distance similarity.
- ✓ Understood that for given data three cluster are most distinct in nature for better targeting, same has been finalized and below (right) graphical figure represents its uniqueness of three clusters.
- ✓ Then, validate whether clusters making sense with respect to amount spent, visits etc.
- ✓ Glimpse of averages of the same are provided in below table:

Cluster	Count	Avg. Spent (\$)	Avg. Recency	Avg. Frequency
Dealers	20	507,874	1.8 Days	185 visits
Potentials	856	17,106	43.9 Days	13.6 visits
By-passers	873	1,125	286 Days	1.64 visits

Output from SOM clustering (3) clusters being unique among themselves



BUSINESS INSIGHTS

- i. As it is observed, for client the business was new and being a good discount provider, dealers turned out to be a good source for the product.
- ii. Apart, from dealers we found next best segment which has around 850 potential customer who can be of immediate target for the client to promote the product.
- iii. Further, client can ignore last group who are just by-passers, i.e. just visited as a pass-through by or might be far away customers. Targeting them may not make sense as their probability of response will be to low.
- iv. Hence, identified potential segment provides good segment for targeting.
- v. Later, understood in client feedback in their first targeted campaign they received 27% response and later increased to max of 43%.