**Customer Segmentation: Know Your Customer**

Name: Anusha Savaram

Net ID as3444

Email ID:[as3444@rutgers.edu](mailto:as3444@rutgers.edu)

RU ID: 205003732

A good understanding of customers is extremely important for running a successful business. KYC (Know your customer) makes a lot of difference for many companies. Customer Analytics is a vital part of marketing Analytics.

‘Customer Segmentation- Know your customer’ is a customer personality Analysis- an attempt to understand the customers, their consumption, spending behavior. We are working on a B2C business model. This dataset is a collection of demographic and psychographic data. This dataset belongs to 2012-2014 data of the company with 2240 rows and 31 columns. An EDA (Explanatory Data Analysis is carried out to understand the data)

Chart, bubble chart

Description automatically generated

Fig1: Bubble Chart: Education and income per capita of customers

Chart, pie chart

Description automatically generated

Fig2: Pie Chart: A story of company’s platforms usage by customers

We use k-means clustering and study cluster wise behaviors in terms of the commodity the customer is interested in, age, educational background, marital status, number of children at home (if they are teen or kid), are they web, in store, catalog purchasers or are they loyal consumers of fruits, fish, sweet, gold, do they buy on a discount.

This analysis helps businesses in understanding their customers to customize their products, campaign to select targeted customers with selected campaigns. This will cut the Advertisement expenditure significantly and channelize the budget as deemed necessary to increase sales.

This is a part of STP (Segmentation, Target, and positioning) framework. This project helps this company in market segmentation. This is achieved by scaling down the data using standard scaler and dimensional reduction us Principle Component Scaler (PCA) and subjecting the subsequent data to a machine learning model called k-means clustering

Chart, line chart

Description automatically generated

Fig3: k-inertia to choose k value

This graph between inertia and k value is seeing a dip at k=3 (low k and inertia). So, we choose 0 to 3.i.e., 4 clusters

We used data from machine learning model to figure out customer’s income and spending capacity.

Chart, scatter chart

Description automatically generated

Fig4: Scatter plot on income vs spend trend

We use clustering to study the behavior of clusters. Cluster 0 seem to the elite class with high income and high spending capabilities. Cluster 1 is the group with low income and low spending.

Fig5: Bar graph on Company’s consumer Chart, bar chart

Description automatically generateddistribution among the customers.

Company’s customer base is tilted to cluster 0. Clearly this company has large number of elite class customers.

**Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated**

Fig6: Pair Bar chart: Customer Forum preference. Fig7: Pair Bar Chart: Product preference

Chart, bar chart

Description automatically generatedFig8: Bar Chart: Company’s website hits.

Cluster 2 followed by cluster 0 hits the customer’s website frequently.

This company is known for its wine. Since it is 2012-2014 data. Internet boom might not be so prevalent.

The below cluster wise network graphs reiterate the same.

Cluster 0 hits the website more frequently but its consumption is fairly low both instore and online. Cluster 0 is the elite class who frequent to company’s website and form the major chunk of customer base. They are the highest complainers implying lower customer satisfaction. They need to be appealed and incentivized to maintain a longer, healthy relationship.

Cluster 3 stands as the cluster with next best spending capacity. They are lowest complainers and hardly have kids. They are highest wine purchasers and purchase lot of it instore. Customer relationship with cluster 3 can be maintained positioning wine instore.

Cluster 1 and Cluster 3 are experimentative in nature, they have an appetite for wide range of products. Cluster 1 has been with the company for longer. They are a class of low income low spend. They are one of the highest buyers of wine with highest purchases in store. They can be checked on the new product lines in case the company plans to release.

Cluster 2 hits the company’s website the highest. They can be targeted and satisfied by ecomerce business. Their sale is fairly low butter better than cluster 0. They are one of oldest customers of the company.

The trends show that company is moving towards elite class customers. Elite class are coming to the company’s stores and websites. This is a recent development. They can be satisfied both on web and store. Their product consumption needs to increase for better sales.

A picture containing skiing, slope, line

Description automatically generatedChart

Description automatically generated

Fig9: Network Graph: Forum Preferences. Fig10: Network Graph: Product Preference

**Chart, line chart, histogram

Description automatically generated**

Fig10: Illustrates the Age Desnsity distribution amongst clusters. Cluster 0 has higher middle aged consumers who frequent to the company.

**Potential Areas:** Company’s ecommerce can be explored. Cluster 0 is the next big things. Company’s future strategies need to be aligned to their satisfaction.

**Future Scope of project**: Some more information on geographic and psychographic data (like data on Impulsive or planned buyers. Customer biases) would have given more insight on customer preferences and business as a whole.