**Research on Customer Personality Analysis**

Sharath Chandra Tummanapally

Department of Data Science, Bellevue University

DSC550 T301 2223 – Data Mining

Prof. Brett Werner

March 2, 2022

**Introduction**

**Is it possible for a company to develop its business by researching the personality of its ideal customers?**

Customer Personality Analysis is a detailed analysis of a company’s ideal customers. It helps a business to better understand its customers and makes it easier for them to modify products according to the specific needs, behaviors and concerns of different types of customers. It also helps a business to modify its product based on its target customers from different types of customer segments.

**Importance to solve this problem**

By designing the correct website interaction techniques, we can ensure that we direct onlookers’ attention to the right spots. This will benefit the company. We also saw that customers who has high income tend towards the purchasing more on company website.

My main objective is to create a solution or build a model to address this problem of understanding the customer’s attributes like their purchases, income or location and provide a helpful insight to a company that implement this solution.

**Pitch to stakeholders to gain buy-in**

In any retail company like Amazon.com, the customers tend to spend and purchase more when there is a high activity on website of the company. Customers who are just browsing through company’s services, they've expressed an interest but have yet to decide. We can make the company's website interesting and appealing and keep the design fresh.

**Dataset source**:

The dataset I retrieved is from an online source Kaggle, where it provides the attributes of the Customers, Companies products details, Promotions offers by the company and the mode of purchases by the customers. The dataset contains each individual customer and with respective provide data, number of purchases etc. We will get through the data in detailed as we dig deep into data and clean it for our analysis.

**Summary of Milestones 1- 3**

**Exploratory Data Analysis:**

Visualization 1**:**

**Chart, bar chart

Description automatically generated**

This graph helps on analyzing which product is generating more revenue in company. By looking at the graph, it shows that Wine product is generating more revenue to the company in the last 2 years.

Visualization 2:

Chart

Description automatically generated

This graph helps to understand the customer satisfaction. The above Histogram explains about the number of customers complained in the last 2 years. We can understand that there are very few customers who complained on the products.

Visualization 3:

Chart, pie chart

Description automatically generated

This graph gives detailed view of customer purchases through which segment. The above Pie chart shows that customers are more inclined towards going to store and purchasing the products. So, we can focus on the updating the store with customer friendly environment.

**Data preparation:**

As a next step, I moved to the Data Preparation. For this, first I identified the unwanted columns and deleted them from data frame as they are not required for our analysis. This refined the dataframe a little. To help us more in analysis, I added new features - ‘Age’, 'NumPurchases' and’ MntTotal' as columns to the dataframe and added their appropriate values. We have added a new column 'Age' to determine the age of the customers and understand the demographics of the customer population and their purchases. I have converted the data frame columns to appropriate datatypes. And finally, I verified the missing data and updated null values to zeros. Thus, we have a refined dataframe.

**Model building and evaluation:**

When we look at the dataset and our objective of the prediction, we have multiple independent variables impacting a variable - ’MntTotal’. We can say that our Total amount spent on all products (MntTotal) is dependent on Customer’s attribute of Income, Age and Total Number of Purchases made by the customer. Hence we are choosing a Multi Linear Regression model. Multiple Linear Regression (MLR) is basically indicating that we will be having many features and our output feature. in other words, we have one dependent feature and multiple independent features. Accuracy of this model is around 75%. This better/good accuracy score of our model suggests that our regressive model has fitted well to the existing data. This also suggests that our data is suitable for multi linear regression.

**Conclusion**

As a curious kid from the childhood and with a knack of problem solving around the household, building a solution to a problem gives a great joy. Although it different in terms of process and steps involved, Model Building is similar. Here, we collect data, comprehend and pay attention to what matters in the data, and then build a model to acquire a better understanding and make predictions. Going back to our model I built for this, after analyzing it, I would recommend implementation because the model does perform well, and it has the accuracy around 75%. This also signifies the importance of understanding the customer’s attributes like their purchases, income or location which can further provide a helpful insight to a company that implement this solution. We can improve the model by using it in more datasets that are on the customer purchases across different demographics and refine the model.