CUSTOMER PERSONALITY

ANALYSIS

**Contributors:**

Parnal Ghundare Patil

Supriya Bidanta

**TABLE OF CONTENT**

1. Abstract
2. Introduction
3. Methods
4. Results
5. Discussion
6. References

**ABSTRACT**

The objective is to understand the relationship between the economic division of society and customer satisfaction for a service. 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 segment. Here, analysis is done based on the income and the response towards a particular campaign across the two major categories of customers. The results are more inclined to a section of demographic where offers were accepted at the earliest compared to others, and the satisfaction level was quite overwhelming.

Key Words: Consumer behavior, Exploratory data analysis, K-means, Clustering, DBSCAN, EDA, Feature engineering, Scaling, PCA, Gaussian Mixture, Silhouette Score

**INTRODUCTION**

The world of business has always been driven by the customers and its necessity. The concept of customer satisfaction has always been the spine of marketing literature since the onset of the word itself. Which leads us to the question: What are customer expectations? Generally, customer expectations are a set of ideas about a product, service or a brand that a customer holds in their mind. For example, customers that buy an Apple iPhone over another phone brand have a set of expectations about that product. For example, they expect to see that the Apple phone has a sliding lock function, a ‘slate-style’ that has few to no physical buttons, with ‘Face ID’ facial recognition as standard. Customers want these expectations met for them to feel satisfied with the customer service and with their purchase. They’ll also have expectations that aren’t so easy to see. Our segmentation project intends to improve marketing tactics of a virtual company by grouping customers into segments with specific characteristics.

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 helps a business to modify its product based on its target customers from different types of customer segments. For example, 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 segment.

The dataset was downloaded from [**Kaggle**](https://www.kaggle.com/imakash3011/customer-personality-analysis) and consisted of 2240 records and 29 columns. There were 26 numeric columns and 3 categorical columns within the dataset. Columns were related to the customer’s income, expenses, amount spent on different products, marital status and educational status, number of responses from campaigns, etc.

**METHODS**

The different method we chose to segment the customers are describe below:

1. **K-Means**

* Initially started by cleaning the data by replacing the nan values with the mean of the columns
* Then comes EDA and based on the results from EDA, some of the columns which aren’t adding any value to the data are removed
* Feature engineering is done based on the observations from EDA
* The data is scaled using *StandardScaler* so that no column will have significantly higher weightage
* Different PCA components were checked for their variance ratios
* Then, a combination of PCA components and clusters were analyzed. For example, we checked these PCA and cluster combinations: PCA-3-K-4, PCA-3-K-3, PCA-5-K-3, PCA-7-K-3
* Based on that, different cluster combinations were analyzed

1. **DBSCAN**

* Followed the same process as above for KMeans till point 5
* Testing done for different PCA components and eps value, and minimum samples similar to K Means.

1. **Gaussian Mixture Model**

* Reduced dimensionality was used for the analysis.
* Calculate Silhouette Scores, based on GMM and select the best cluster for the given data set.

**RESULT**

We eventually went ahead with **K-Means** since it was giving a comparable split of the cluster segments in terms of the counts within the 3 clusters and the split on the different parameters too seemed logical.

Based on this analysis, we found classified cluster 1 as low-income cluster, cluster 2 as moderate income and cluster 3 as high income cluster

|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute** | **Cluster 1**  **Low Income** | **Cluster 2**  **Moderate Income** | **Cluster 3**  **High Income** |
| **Cluster Size** | Highest ~1100 | Low ~600 | Lowest ~500 |
| **Age** | Lower | Highest | Slightly lower |
| **Expenses** | Lowest | Moderate | Highest |
| **Income** | Lowest | Moderate | Highest |
| **Family Size** | Close to 2.5 | ~ Cluster 0 | Lowest |
| **%Amount Spent on Wines** | Low | Highest | Low |
| **%Amount Spent on Fruits** | Highest | Lowest | Moderate |
| **%Amount Spent on Meat** | Lower | Lower | Highest |
| **%Amount Spent on Fish** | High | Lowest | High |
| **%Amount Spent on Gold** | Highest | Lower | Lowest |
| **Education** | Higher graduate percentage | Higher graduate percentage | Does not have undergraduates, higher graduate percentage |
| **Marital Status** | More number of married people | More number of married people | More number of married people |
| **Complaints** | High | Medium | Low |
| **Web Visits** | High | Medium | Low |
| **Responses** | Low | Low | High |

**Cluster 1**

1. Low income, spends less, average age on the lower side, spends more on fruits and gold.
2. High graduate percentage, more married people
3. High complaints and web visits, low responses to campaigns

**Cluster 2**

1. Moderate income, moderate spending, average age on the higher side, spends more on wines.
2. High graduate percentage, more married people
3. Medium complaints and medium web visits, low responses to campaigns

**Cluster 3**

1. High income, spends more, average age on the lower side, spends more on meat and fish.
2. High graduate percentage and does not have any undergraduates, more married people
3. Low complaints and low web visits, high responses to campaigns

**DISCUSSION**

For the given data set, after analyzing the EDA results, we chose K-means, Gaussian Mixture Model and DBSCAN for the segmentation of customer into the different clusters based on their purchase history and purchase selection. After doing the segmentation analysis, we eventually chose KMeans since it was giving a better characteristic split of the three segments.

There is no proper segmentation amongst different categories but considering the economical aspect of the society we can assume that the purchase is more among the moderate income (based on cluster 2). Based on the analysis done for this dataset, we can assume that the data is highly imbalanced and it requires more categorical data. For instance, rather than generalizing the data to wine/meat/fish, there should be one hierarchical column set, alcohol->beer/whiskey/rum/vodka/wine -> brands etc., and similarly for fruits/sweets etc. More hierarchical data will help us predict more about the purchase trends and the type of product favored by segment of consumer. If this prediction is considered, then we can use it for targeting specific customer based on the campaigns/discounts/offers response and retain them for longer period. In market ecology, now and then we can see new products with varied number of features. Therefore, it is difficult to retain customer. Hence, the customer personality analysis is the paramount key to thrive in business.

**REFERENCES**

* <https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/>
* <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>
* <https://www.kaggle.com/ekami66/detailed-exploratory-data-analysis-with-python>
* <https://matplotlib.org/>
* <https://www.semanticscholar.org/paper/The-role-of-customer-personality-in-satisfaction%2C-Smith/1cbb40a481f0db28d33b80d457b81db19cf127df#paper-header>
* <https://docs.w3cub.com/scikit_learn/modules/generated/sklearn.mixture.gaussianmixture>
* <https://www.kaggle.com/vipulgandhi/gaussian-mixture-models-clustering-explained>
* <https://towardsdatascience.com/gaussian-mixture-model-clusterization-how-to-select-the-number-of-components-clusters-553bef45f6e4>