BUSINESS CASES WITH DATA SCIENCE

MASTER DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS

Wonderful Wines of the World

Group G

Ehsan Meisami Fard, number: 20201050

Steffen Hillmann, number: 20200589

Maximilian Maukner, number: 20200645

February, 2021

TABLE OF CONTENTS

Github Repository: https://github.com/ehsanmeisami/BC1_GroupG

1. INTRODUCTION

Wonderful Wines of the World (abbreviated WWW) has been offering a unique range of wines for over 6 years, which can be ordered in the store, online via the website but also by phone. In order to continue to operate successfully in the market, customer analysis is of high importance to continue to achieve a high level of customer satisfaction with the services. Only in this way WWW can present itself to its customers as an attractive supplier of wine, achieve a high level of customer loyalty and thus gain market share in the wine industry. Up to now, an aggressive marketing strategy was chosen, in which customers received a catalog with the latest wine every 6 weeks. So far, there is no loyalty program, which could be a useful strategy as it positively influences the attitudes and buying behavior of customers and thus strengthens brand loyalty.
For this purpose, a customer segmentation is carried out in the following, which is based on the data of almost 10,000 customers of WWW. Therefore, the individual customers and their respective attributes are first analyzed and selected in detail. Afterwards, different clustering techniques are compared and the most adequate one selected to cluster the customers. Subsequently, the profiling of the individual clusters follows, which builds the basis for the possible marketing approaches, that are briefly explained in the end.

With the results obtained, it is possible to understand the behavior and characteristics of current customers, to gain insights for the acquisition of new customers and to improve the relationship with the current customers. The main goal of the analyses is to identify different customer segments in order to develop strategies.

2. BUSINESS UNDERSTANDING

2.1. Business Objectives

Wonderful Wines of the World has created a customer database over the last four years. So far it has mass-marketed its customers yet has the plan to target more specifically. Therefore, it has the objective to gain more intelligence about its database to better understand its consumers and hence to develop more targeted marketing programs. Furthermore, as a success criteria, at the end of the project WWW should be able to differentiate its customer base from each other based on multiple different perspectives.

2.2. Situation assessment

The dataset provided by the client represents a sample of 10,000 customers from its active database and consists of 30 features. It includes socio-demographic data, information about the customer's taste preferences, profitability measures such as RFM and LTV and lastly buying behavior. The customers selected for the campaigns were chosen randomly within the defined constraints: individual customers that have purchased something from the client in the past 18 months.

2.3. Determine Data Mining goals

Subsequently, we translate the business objectives into rather more technical data mining objectives. In this case, the objective is to generate a clear distinction among the current consumers from

socio-economics, profitability and taste preferences perspectives. Segmentation of the customers into thought-provoking and meaningful subgroups that assist the marketing campaigns to target WWW customers better.

2.4. Produce Project Plan

The project plan consists of multiple phases while each phase is defined through multiple deliverables. First of all, understanding the business and its organizational objectives is fundamental for subsequent data mining objectives. Here it is essential to understand the problem or desire of the business first prior to defining the technical aspect of the project. Lack of defining the business objectives could lead to working on a solution to an inaccurate question. Furthermore, as Wonderful Wines has provided us with the information regarding their customer base, we can skip data collection and move on to selecting feature selection, data cleaning, and feature engineering. Once the data preparation has been completed, selecting the appropriate model for the problem at hand is the next phase of the project. In this phase, we implement multiple techniques to generate interesting results. Subsequent to modeling, we start off with the evaluation and assessment of the generated output and clusters through cluster analysis. These clusters need to satisfy the business objectives and provide a profound solution for them.

3. DATA MINING PROCESS

3.1. Data Understanding

The investigated dataset consists of 10,000 observations and 30 features which was provided by the client and thereby collection of data is no further required. Each observation represents one customer of WWW. There are no duplicates among the rows and the columns of the dataset consist of two different data types, namely 9 integers (*int64*) and 21 floats (*float64*). The dataset provides socio-demographic data, information about the customer's taste preferences, and profitability measures (RFM).

3.2. Data Preparation

Data preparation consists of selection of specific features, cleaning the data and eventually implementing feature engineering to discover new information. As the given dataset is the only data source for this project, we will not merge any external data sources into the current dataset.

A closer look at the columns reveals that the columns *Custid* will not be relevant for the upcoming model as it exposes no further information about the customer and thus the column will be removed from the dataset. Prior to data exclusion, we separate the data into two distinct subgroups, namely metric and non metric features. The metric features are used as input for the clustering methods implemented. Yet upon closer examination of non metric features, we observe a clear discrepancy in the distribution of the non metric features. Thereby, we drop all of the non metric features except for *Kidhome* and *Teenhome*. As these two features display a similar form of information, we implement feature engineering to create the feature *Child* that summarizes both of these two columns into one[1].

Due to the different scales in our metric dataset that we use for the clustering problem, we use the MinMaxScaler to transform and normalize the input features.

[1] Jupyter Notebook: Part 2.1 - Non Metric Visualisation

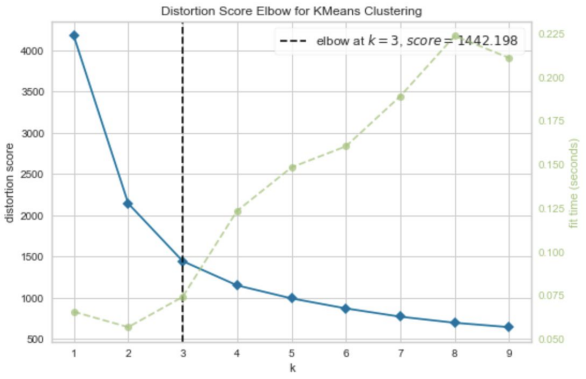
Furthermore, we use the Principal Component Analysis to implement dimensionality reduction. The principal components represent each a proportion of the explained variance of the set of observations, with its first principal component explaining most of the variance. In this case, we have decided to use the first three principal components which represent a cumulative explained variance of 80.3% of the observations.

3.3. Modeling

Once the data is suitably prepared, the model creation process can be initialized. To achieve the best possible outcome, selected clustering techniques (partitional, hierarchical, and density-based) are used and compared with each other.

3.3.1 K-Means

K-Means is a partitional clustering technique that groups unlabeled data points into "k" number of clusters, while keeping the intra-clusters distances as small as possible. K-Means' advantage is that it is quite fast and can handle large datasets. However, the main drawback is that the algorithm requires to define the number of clusters prior to model execution.



One popular technique to select the optimal number of clusters is the Elbow-Method, which fits the model with a range of values for K. Here the Elbow curve shows clearly that the optimal number of components is 3 with an R2 of 65%.

3.3.2 Hierarchical Clustering

Hierarchical Clustering, on the other side, initially considers each data point as an individual cluster. At each iteration, the similar clusters merge with other clusters until one cluster or K clusters are formed. Advantages of this clustering technique are that the number of clusters doesn't need to be specified a priori and it's robustness in dealing with noise. However, hierarchical clustering techniques generally suffer from time complexity and space complexity.

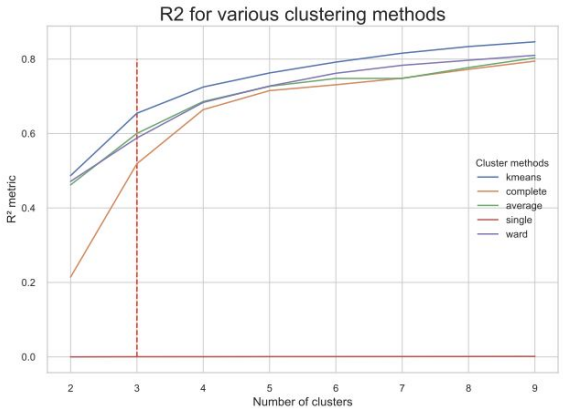
3.3.3 Density-Based Clustering

DBSCAN groups data points that are close to each other based on a distance measurement (epsilon) and a minimum number of points (neighborhood). Its advantages over partitional clustering techniques is that it is capable of detecting clusters of arbitrary shape and size containing even noise and outliers. However, DBSCAN struggles when applying it on large dimension datasets. This effect can be seen after running the model with the predefined number of neighbors on the dataset: It almost places all data points in one cluster with a resulting R-squared of 0.3%.

3.4. Evaluation And Model Selection

As previously mentioned DBSCAN is not the algorithm to go with here[2]. Further, to compare the different hierarchical clustering techniques and K-Means the R2 metric is used.

Ultimately, the higher R-squared score is the decisive criterion why the K-Means algorithm is used for clustering in the following. Moreover, hierarchical clustering should rather serve as an indicator for the number of clusters in K-Means than as the final clusterer.



4. RESULTS EVALUATION

Following up on the previous results, the customers of the client are grouped into 3 different clusters using K-Means. In the further course, the features of the respective clusters are examined in more detail. For the sake of clarity, the customers are divided into clusters corresponding to the following frequencies: Cluster 1: 2970, Cluster 2: 3494, Cluster 3: 3536.

Moreover, we will examine different perspectives against which we will measure the success of our customer segmentation:

Moreover, we examine each cluster based on the different perspectives as we defined before. The resulting cluster analysis on distinct perspectives should allow us to measure the success of our customer segmentation.

4.1 Socio-Demographic Differences

This perspective includes the variables *Age*, *Income*, *Education* and *Childs*. Here, a clear discrepancy can be seen for the two first-mentioned variables. Customers of Cluster 1 are young (the average age is 29), possess a low income and 86% of them have at least one child at home. In contrast, customers of Cluster 2 are on average 67 years old and have an significantly above-average income compared to the individuals of the other 2 clusters.

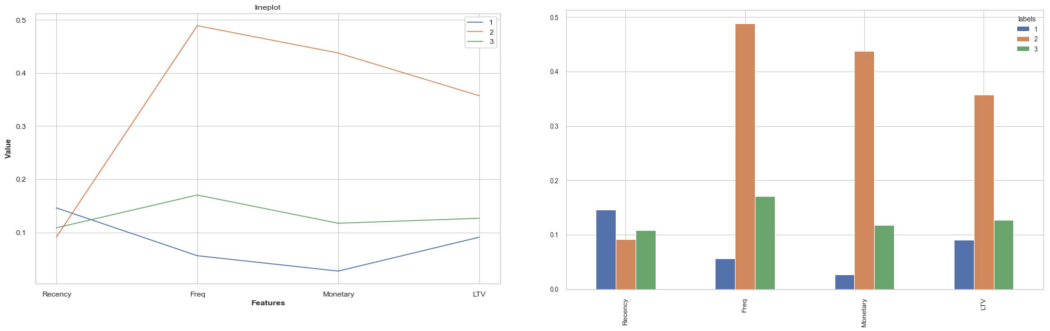
4.2. Differences in Customer Profitability (RFM)

As mentioned in the socio-demographic perspective, the clusters differ in age as well as in income. In order to gain further insights into profitability, the clusters are distinguished in terms of *Recency*,

[2] Jupyter Notebook: Part 2.6 - TSNE, to see density of data points, to visualize that DBSCAN isn't appropriate for this BC

Frequency, *Monetary* and *LTV*. It becomes clear that Cluster 1 has been less active recently compared to the other two Clusters. Recency is measured in days since the last purchase. The first Cluster possesses a recency score of 80 while Cluster 2, on the other hand, has a value of 50 and Cluster 3 a value of 60. In addition to Recency, Frequency is also a decisive factor which is the amount of purchases in the past 18 months. It can be seen that Cluster 1 has the lowest value with 4, whereas Cluster 2 has 50 and Cluster 3 has a value of 60. Furthermore, from the analysis it is understood that Cluster 1 generates the lowest sales per purchase with just 88. Next to Cluster 3 which achieves a moderate turnover per purchase with 362, it becomes clear that Cluster 2, on the other hand, generates the highest turnover with an amount of 1138.

Profitability measures compared by clusters



4.3. Differences in Tastes Preferences

It is also interesting to take a closer look at the customers' wine tastes within the clusters. Dryred wine is the most popular among each cluster, especially for Cluster 3 (in this group it is purchased in over 70% of the transactions). Furthermore, individuals belonging to cluster 3 seem to dislike sweet and dessert wine (they are significantly under-average in terms of buying this type of wine). The younger customers of Cluster 1 are enjoying every wine. Besides Dryred they also have a faible for exotic wine with 30%.

4.4. Differences in Buying Behavior

Last but not least, this perspective provides an even deeper insight into the behavior of the customers of the client. Clusters 1 and 3 like to shop online and seek out for bargains. Almost 60% of the individuals belonging to Cluster 1 and 40% of Cluster 3 customers made purchases bought on discount, whereas Cluster 2 neither buy on discount (only 5%) and rarely shop online, 20%.

An overview of the above-mentioned findings are visualized in the table below.

	Cluster 1	**Cluster 2**	**Cluster 3**
Socio-Demographic	Young, low income, 1 child	Elderly and high income	Middle age, moderate income and 1 child
Profitability	Lowest LTV and lowest RFM scores	Most valuable customers in RFM terms	Moderate LTV and RFM scores
Taste	Dryred is the favourite among all clusters		

	Faible for Exotic wines	Wine enthusiasts	Dislike sweet and dessert wines
Buying Behavior	Buy online and on discount	Not usually buy online nor on discount	Similar to cluster 1

Overall, when evaluating the above-mentioned perspectives, one can see a clear distinction between the clients' customers and thus a clear segmentation of the customers into meaningful and significant subgroups that will allow the Wonderful Wines of the World to create more targeted marketing campaigns per segment/cluster. The results illustrate that we have been successful in discriminating subgroups of customers of WWW and hence we move on into the next section of the report which is to formalize strategies to approach each Cluster individually.

5. DEPLOYMENT AND MARKETING STRATEGIES

For future segmentation of upcoming customers, WWW solve this issue by using these clusters to classify new customers with the k-nearest neighbors algorithm where an object is classified by a plurality vote of its neighbors, with the object being assigned to a cluster with most common among its k nearest neighbors. Important to note, that this a recommendation for WWW and this step has not been implemented yet. Moreover, we advise WWW to re-evaluated the clusters in predefined intervals as the customers always evolve and thereby change themselves.

The following marketing strategies are structured according to the 'Marketing Mix' which is a foundation model for businesses and includes products, price, place and promotion (also known as the 4 Ps).

5.1. Cluster One:

The customers within this cluster are young, low income parents that like to purchase wine online based on the current discount options given on their favourite brand which is the Dry red wine. Despite their low LTV compared to the other clusters, they are easy to target with discounts on Dry red wine products and represent 29.7% of the total current customer base of Wonderful Wines of the World. Due to their low recency scores we might trigger them to buy again by introducing a discount on their favourite wine such as Dryred and Exotic. Furthermore, due to their significant online shopping habits, we can assume that they are heavy internet users and thereby WWW could use online marketing tools such as newsletters and online advertisement to inform them about the new discounts on their preferred product.

5.2. Cluster Two:

Compared to other clusters, cluster 2 consists of high income and rather elderly customers. Most of their children have moved out of their home, meaning they are most likely living with their partners and enjoying retirement. As mentioned earlier. This cluster is by far the most valuable customers among the whole customer base of Wonderful Wines of the World. Additionally, the lack of purchases on discounted goods signals the fact that we are dealing with rather sophisticated customers that enjoy high quality of wine and are willing to spend a lot of money. Given these facts, a marketing strategy to target these types of customers would be to include and emphasize the

quality of the wine offered by WWW. Based on the purchase behaviour of this cluster, one could assume that this cluster deeply cares about the wine and are wine enthusiasts. Thereby, WWW could host an exclusive wine tasting event for this type of customers based on their geographic location. This would not only lead to higher immediate sales but also raises the brand image of WWW among customers that are highly profitable.

5.3. Cluster Three:

In terms of profitability, our third cluster comes second. The major difference between this cluster and other is the taste preferences. This information with the fact that they also prefer purchasing discounted products, gives us the opportunity to reactivate their spending habits and offer them good deals on their preferred product. Due to their relatively high years of education, one could assume that the customers within this cluster are also knowledgeable customers and could be triggered with more exposure to their original and background of the wine they are purchasing. This could mean printed brochures and handbooks for specific wines of their taste to read about as goodie along their wine purchase.

5.4. Recommendation for future data mining projects:

By increasing the dimension of the dataset of the current customers, one could understand, specially for price sensitive clusters, if they buy the wine they frequently purchase due to the discount given on them, or they buy for the taste of the product itself. In such a manner, we require more data on buying behaviour on discounted products. On the other hand, the second cluster performs poorly on discounted products. Could this mean that the more sophisticated customers do not like discounted products and this could actually lead to diminished sales? These are the type of questions that could be answered by increasing the amount of data generated regarding the purchase and sales details. For instance, WWW should integrate a discount on products. Meaning how often the price of a product has been discounted and the corresponding sales performance.

6. CONCLUSION

Overall, we have been able to help WWW to achieve their initial business objective which was to understand their customer base better and be able to find insights that could help them target their customers individually and more directly. Subsequent to multiple data mining techniques that have been implemented, we divided the customer base into three distinctive subgroups. These subgroups have been analysed based on multiple perspectives such as socio-demographics, profitability, buying, tasting preferences and lastly buying behaviour.

Now, for the clusters defined above, as well as the associated brief marketing approaches and ideas, precise strategies for the respective clusters should be specifically worked out and implemented by the marketing department. It is crucial that all steps and successes are documented and recorded in detail. After numerous months, a first assessment can eventually be made. Subsequently, as customers and circumstances usually change considerably over time, it is recommended that, based on this new data, a customer segmentation like this is carried out again. It is very likely that major changes will take place, resulting in completely new clusters and therefore completely new marketing strategies. The development and behavior of customers and the corresponding clusters is not static and linear, but rather dynamic and often unpredictable, which requires constant observations and adjustments.