BUSINESS CASES WITH DATA SCIENCE

MASTER'S DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS – MAJOR IN DATA SCIENCE

Business Case 1 – Wonderful Wines of the World

Group H

Hiromi Nakashima, m20201025

Manuel A. F. Carreiras, m20200500

Luis F. R. Agottani, m20200621

Venâncio Munhangane, m20200579

February, 2021

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

TABLE OF CONTENTS

FIGURES INDEX

TABLE INDEX

1. INTRODUCTION

In this project we were designated to analyze a wine company's dataset named Wonderful Wines of the World (WWW). WWW is a 7-year-old enterprise, which sells wine through three channels: catalogs, web site and physical stores (10 branches). The purchase can be done in the physical stores, telephone or online.

At this moment, WWW keep clients engaged by sending them a newsletter with the updates of wine world. Despite the fact that the database of WWW has only 4-year-old, the company recently organized a marketing activity which aggressively increased the data volume. One of the current pain points is a lack of cross selling strategies which supports the trade profit.

This project was developed with 10.000 samples of the current WWW's customers database that purchased in the last 18 months. The report was adjusted in four main parts based in CRISP-DM methodology (Pete Chapman, 1999).

The GitHub repository where all the present analysis is saved can be accessed through the following link: https://github.com/hnakashima96/Business-Case-.

2. BUSINESS UNDERSTANDING

At this stage we defined the essential business guidelines to grant a good result of the project. In order to develop the best solution to WWW the business understanding was based on the current reality of the company presented at the introduction.

2.1. BUSINESS OBJECTIVES

The goals of WWW are:

- Improve the familiarity of the database by creating a classification for each client and develop marketing strategies by profile;

- Be able to classify new customers;

- Understand the customer value.

2.2. BUSINESS SUCCESS CRITERIA

Based on the business objectives description, two main results were defined to guarantee the success of this project: identify the profile of the new customers since the first purchase, develop marketing strategies to reach all market segments and improve the trading profit.

2.3. DETERMINE DATA MINING GOALS

Based on the business goals we converted to Data Mining language as shown in the table below

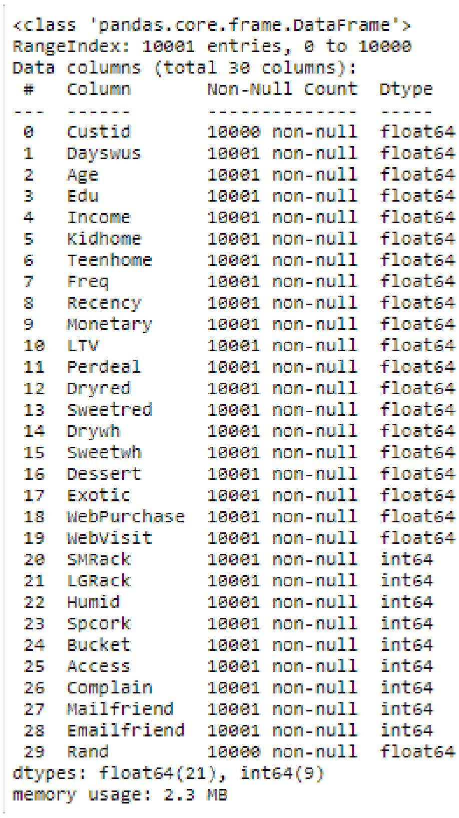
Table 1 - Data Mining Goals.

Business Goal	Data Mining Goal
Classify the currently clients by profile	Clustering the clients
Ranking the clients to understand the ROI	Apply the recency, frequency and monetary value (RFM)
Identify the new customer profile	Create a predictive model

3. PREDICTIVE ANALYTICS PROCESS

3.1. DATA UNDERSTANDING

The initial process of understanding the problem was explained in the introduction and data structure (Figure 1 and appendix I) 30 columns and 10001 entries were identified, all of them represented by numeric features. After a better understanding of the metadata and the features, we recognized that 10 features were binary then we converted them to Boolean type. Also, on this step some columns were drop due to its insignificance for the project (Data Exploration, (Notebook BC1, 2021)).



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10001 entries, 0 to 10000
Data columns (total 30 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Custid       10000 non-null  float64
 1   Dayswus      10001 non-null  float64
 2   Age          10001 non-null  float64
 3   Edu          10001 non-null  float64
 4   Income       10001 non-null  float64
 5   Kidhome      10001 non-null  float64
 6   Teenhome     10001 non-null  float64
 7   Freq         10001 non-null  float64
 8   Recency      10001 non-null  float64
 9   Monetary     10001 non-null  float64
 10  LTV          10001 non-null  float64
 11  Perdeal      10001 non-null  float64
 12  Dryred       10001 non-null  float64
 13  Sweetred     10001 non-null  float64
 14  Drywh        10001 non-null  float64
 15  Sweetwh      10001 non-null  float64
 16  Dessert      10001 non-null  float64
 17  Exotic       10001 non-null  float64
 18  WebPurchase  10001 non-null  float64
 19  WebVisit     10001 non-null  float64
 20  SMRack       10001 non-null  int64
 21  LGRack       10001 non-null  int64
 22  Humid        10001 non-null  int64
 23  Spcork       10001 non-null  int64
 24  Bucket       10001 non-null  int64
 25  Access       10001 non-null  int64
 26  Complain     10001 non-null  int64
 27  Mailfriend   10001 non-null  int64
 28  Emailfriend  10001 non-null  int64
 29  Rand         10000 non-null  float64
dtypes: float64(21), int64(9)
memory usage: 2.3 MB
```

Figure 1 – Variable Information

3.2. DATA PREPARATION

The diagram below presents the data preparation steps followed to reach the final model in this project (Figure 2).

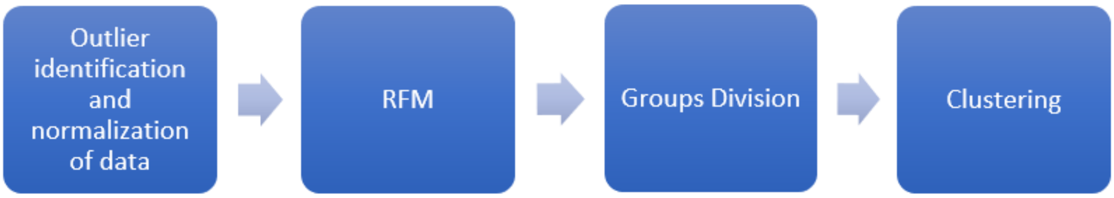


Figure 2 - Data preparation process

Firstly, we identified the outliers from data. Subsequently a RFM analysis was developed to classify the value of each customer, by this classification we reached to 4 clusters which characterize the client quality. The table 2 show the result of this step:

Table 2 - Characterization of the RFM classification.

RFM quality classification	**Characterization**
Loyal	Clients who frequently shop with high consumption and have more recently purchased in WWW.
Recovery	Clients who used to shop frequently and spend high values at WWW.
New Customers	Clients who recently have been at WWW but have not spent much in shops.
Volatile Customers	Clients who only shopped in sales and marketing actions.

In order to avoid misunderstanding of the data, we decided to spread the data in two groups: Taste (taste_group variable, (Notebook BC1, 2021)) and Customer Characterization (cust_group variable, (Notebook BC1, 2021)) Forward, those groups were clustered, where both reached to an optimum number of 4 clusters each and concatenated to a final one. The figure 3 presents the final cluster result.

The cluster 0 were nominated as “Unusual Drinker”, those clients that are attracted by promotions and usually do online shops. They are composed by the younger people from de database and the lower income. The Sweetred, Sweetwh, Dessert, Exotic are the wine’s types that that group mostly buy.

Following, the cluster 1 nominated as “New Drinker”, there are clients who seem to begin their wine journey. They are recent clients and usually visit the website with a great conversion rate, although promotions are not reasoning to a purchase. They prefer Dryred wines and present a relevant interest on accessories.

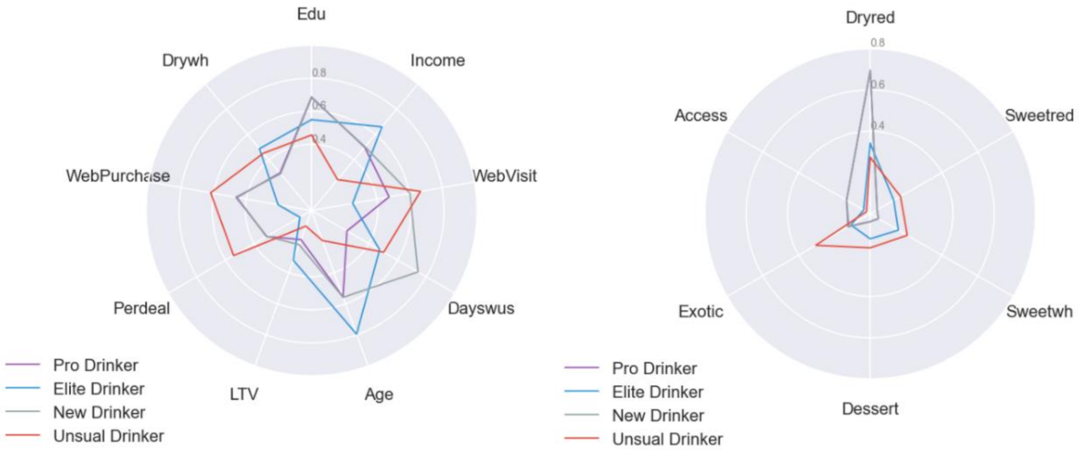


Figure 3 - Left plot presents the clustering of Customer Characterization. Right plot presents the clustering of the Taste plot.

The cluster 3, nominated as “Elite Drinkers”, has a high similarity to the cluster 1. We nominated as Elite due to the difference with the cluster 1, which are senior clients.

The cluster 2 were nominated as “Pro Drinkers”. They are composed by the eldest clients with the highest income. They hardly buy on internet neither visit the site. Also, this cluster do not present an interest on discounts. They are attracted by Dryred wines but seems to be interest on others type options.

3.3. Modeling

As defined in data mining goals in Business Understanding step, our main objective is to predict the classification of new clients based on the classification presented on the data preparation.

The model choosen to reach this result was the Decision Tree due to (Michael J. A. Berry, 2004):

- Return a transparent classification analysis.
- Possibility to identify the impact of each variable on predictive model.

In order to facilitate customer understanding, an application was developed in which the predictive model is implemented in a user-friendly interface.

4. CONCLUSIONS

4.1. MARKETING APPROACH

Based on the analysis shown throughout this report, we recommend to WWW a marketing approach that would guarantee the continuous cash flow and increase of customer loyalty.

Firstly, create a loyalty card for clients where each purchase convert into points and the sum of those become accessories and discounts. This strategy would target New Clients and Volatile Customers.

At the same time, enrich the wine experience creating a “VIP opportunities” for the Loyal clients through wine tasting and special wine combinations.

In order to increase the traffic on the website and recover clients classified as Recovery, the WWW should apply an online advertisement and a regular newsletter.

5. REFERENCES

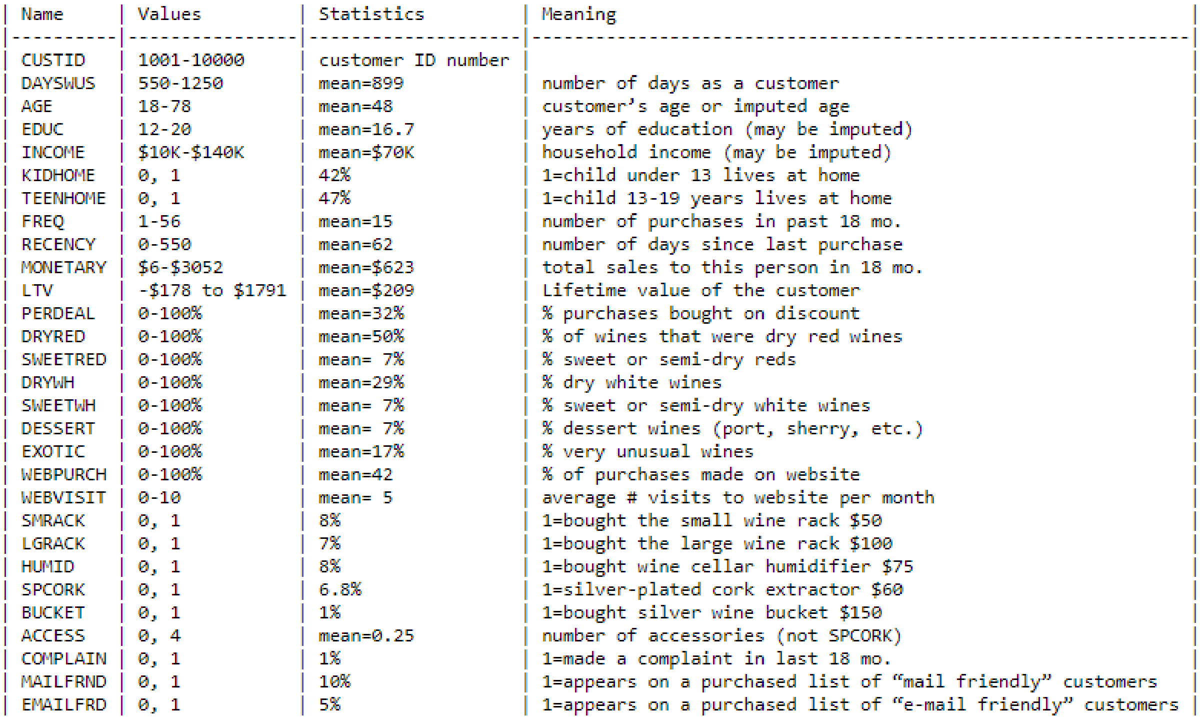
Michael J. A. Berry, G. S. (2004). *Data Mining Techniques - For Marketing, Sales and Customer Relationship Management.* Wiley Publishing.

Notebook BC1. (February de 2021). Business Case 1 - Wonderful Wines of the World. Lisbon, Portugal.

Pete Chapman, J. C. (1999). CRISP-DM 1.0. *Step-by-step data mining guide.*

6. APPENDIX (OPTIONAL)

APPENDIX I – METADATA



Name	Values	Statistics	Meaning
CUSTID	1001-10000	customer ID number	
DAYSWUS	550-1250	mean=899	number of days as a customer
AGE	18-78	mean=48	customer's age or imputed age
EDUC	12-20	mean=16.7	years of education (may be imputed)
INCOME	$10K-$140K	mean=$70K	household income (may be imputed)
KIDHOME	0, 1	42%	1=child under 13 lives at home
TEENHOME	0, 1	47%	1=child 13-19 years lives at home
FREQ	1-56	mean=15	number of purchases in past 18 mo.
RECENCY	0-550	mean=62	number of days since last purchase
MONETARY	$6-$3052	mean=$623	total sales to this person in 18 mo.
LTV	-$178 to $1791	mean=$209	Lifetime value of the customer
PERDEAL	0-100%	mean=32%	% purchases bought on discount
DRYRED	0-100%	mean=50%	% of wines that were dry red wines
SWEETRED	0-100%	mean= 7%	% sweet or semi-dry reds
DRYWH	0-100%	mean=29%	% dry white wines
SWEETWH	0-100%	mean= 7%	% sweet or semi-dry white wines
DESSERT	0-100%	mean= 7%	% dessert wines (port, sherry, etc.)
EXOTIC	0-100%	mean=17%	% very unusual wines
WEBPURCH	0-100%	mean=42	% of purchases made on website
WEBVISIT	0-10	mean= 5	average # visits to website per month
SMRACK	0, 1	8%	1=bought the small wine rack $50
LGRACK	0, 1	7%	1=bought the large wine rack $100
HUMID	0, 1	8%	1=bought wine cellar humidifier $75
SPCORK	0, 1	6.8%	1=silver-plated cork extractor $60
BUCKET	0, 1	1%	1=bought silver wine bucket $150
ACCESS	0, 4	mean=0.25	number of accessories (not SPCORK)
COMPLAIN	0, 1	1%	1=made a complaint in last 18 mo.
MAILFRND	0, 1	10%	1=appears on a purchased list of "mail friendly" customers
EMAILFRD	0, 1	5%	1=appears on a purchased list of "e-mail friendly" customers