

**BUSINESS CASES WITH DATA SCIENCE**

**MASTER DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS – MAJOR IN BUSINESS ANALYTICS**

**Business Case 1 – Wonderful Wines of the World**

Group H

Hiromi Nakashima, m20201025

Manuel A. F. Carreiras, m20200500

Luis F. R. Agottani, m202000621

<Fourth student name>, number: <student number>

February, 2021

INDEX

[1. INTRODUCTION 1](#_Toc65369142)

[2. BUSINESS UNDERSTANDING 1](#_Toc65369143)

[2.1. Business Objectives 1](#_Toc65369144)

[2.2. Business Success criteria 1](#_Toc65369145)

[2.3. Determine Data Mining goals 1](#_Toc65369146)

[3. PREDICTIVE ANALYTICS PROCESS 2](#_Toc65369147)

[3.1. Data understanding 2](#_Toc65369148)

[3.2. Data preparation 3](#_Toc65369149)

[3.3. Modeling 4](#_Toc65369150)

[3.4. Evaluation 4](#_Toc65369151)

[4. CONCLUSIONS 4](#_Toc65369152)

[4.1. Marketing Approach 4](#_Toc65369153)

[5. REFERENCES 5](#_Toc65369154)

[6. APPENDIX (OPTIONAL) 6](#_Toc65369155)

# 

# INTRODUCTION

This project we were designated to analyze a wine company 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.

Into this moment, WWW keep your clients engaged by sending them a newsletter with the updates of wine world. Even that the database of WWW has only 4-year-old, the company recently organized a marketing activity which aggressively increased the database. One of the currently pain points is a lack of cross selling strategies which support the recurrent profit.

This project was developed with a 10.000 sample of the currently WWW’s database of customers that purchased in the last 18 months. And the report was adjusted in four main parts based in CRISP-DM methodology (Pete Chapman, 1999).

# BUSINESS UNDERSTANDING

At this step were defined the essential business guide lines to garante a good result of the project. In order to develop the best solution to WWW the business understanding were based in the currently reality of the company presented on the introduction.

## Business Objectives

The goals of WWW are:

* Improve the familiarity of the database by create a classification for each client to develop marketing strategies by profile;
* Be able to classify the new customers;
* Improve the Return on Investment by understanding the client value.

## Business Success criteria

Based on the business objectives describe, were defined two main results to guarantee the success of this project: identify the profile of the new customers since the first purchase, develop marketing strategies to reach all classification profiles and recurrent profit to the company.

## Determine Data Mining goals

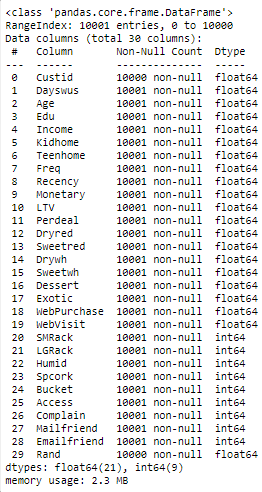
Based on the business goals we translated to Data Mining language as shown in the table below (Table X).

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

# PREDICTIVE ANALYTICS PROCESS

## Data understanding

The initial process of understanding the problem were explained in the introduction and the data structure (Figure X/METADATA PRINT). Were identified 30 columns and 10001 entries, where all of them presented to be numeric features. Although after a better understand 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 the insignificance for the project (Data Exploration, Notebook Reference)



## Data preparation

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

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

|  |  |
| --- | --- |
| **RFM quality classification** | **Characterization** |
| Loyal | Clients who frequently shops with high consume and has the most recently purchase in WWW. |
| Recovery | Clients who used to frequently shops and spend high values at WWW. |
| New Customers | Clients who recently been at WWW but have not spent much in shops. |
| Volatile Customers | Opportunities clients who probably shopped in sales and marketing actions. |

In order to avoid miss understanding of the data, we decided to spread the data in two groups: Taste (taste\_group, notebook) and Customer Characterization (cust\_group, notebook). Forward, those groups were clustered, where both reached to an optimum number of 4 clusters each and concatenated to a final one. The figure X presents the final cluster result.

VISUALIZAÇÃO DOS CLUSTERES

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”, are the clients who seems begin their wine journey. They are recently clients and usually visit the website and a great conversion rate, although promotions are not decision make to a purchase. They prefer Dryred wines and presents a relevant interest on accessories.

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 compose the elderly 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.

l

## 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 chose to reach this result were the Decision Tree due to (Michael J. A. Berry, 2004):

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

## Evaluation

Results described in technical terms (e.g., reached an Accuracy of 95%).

# CONCLUSIONS

## 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 be target on the 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 site and recover clients classify as Recovery, the WWW should apply an online advertisement and regular newsletter sending.

# Referências

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

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

# APPENDIX (OPTIONAL)