

**BUSINESS CASES WITH DATA SCIENCE**

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

**Business Case #1 - Wine store**



Data4Business Consulting

Débora Santos, number: 20200748

Diana Furtado, number: 20200590

Pedro Medeiros, number: 20200742

Rebeca Pinheiro, number: 20201096

February, 2021

INDEX

[1. INTRODUCTION 1](#_Toc65173054)

[2. BUSINESS UNDERSTANDING 2](#_Toc65173055)

[2.1. Background 2](#_Toc65173056)

[2.2. Business Objectives 2](#_Toc65173057)

[2.3. Business Success criteria 3](#_Toc65173058)

[2.4. Situation assessment 3](#_Toc65173059)

[2.4.1. Terminology 3](#_Toc65173060)

[2.4.2. Risks and contingencies 3](#_Toc65173061)

[2.5. Determine Data Mining goals 3](#_Toc65173062)

[2.6. Project Plan 3](#_Toc65173063)

[3. PREDICTIVE ANALYTICS PROCESS 4](#_Toc65173064)

[3.1. Data understanding 4](#_Toc65173065)

[3.2. Data preparation 4](#_Toc65173066)

[3.3. Modeling 7](#_Toc65173067)

[3.4. Evaluation 7](#_Toc65173068)

[4. RESULTS EVALUATION 8](#_Toc65173069)

[5. DEPLOYMENT AND MAINTENANCE PLANS 9](#_Toc65173070)

[6. CONCLUSIONS 10](#_Toc65173071)

[6.1. Considerations for model improvement 10](#_Toc65173072)

[7. REFERENCES 11](#_Toc65173073)

[8. APPENDIX (OPTIONAL) 13](#_Toc65173074)

# 

# INTRODUCTION

We were challenged to analyse the dataset provided by a wine business - Wonderful Wines of the World (WWW) - who aims to better understand its customers’ characteristics and segmentation and to improve the set-up of the business in order to gain new customers and retain the current ones. This is the first study targeting this business’s clients, looking at the 10.000 customers that purchased over the past 18 months.

The present report intends to give an overview of the process behind the analysis, which is made following the CRISP-DM reference model.

In the kick-off meeting, the WWW managers set the following three key questions that we aim to get answers to by the end of the report:

* Which characteristics best distinguish the customers.
* Which and how many customer segments there are in the provided database.
* How can the business reach new and existing customers from each segment and which ones should be prioritized.

[[1]](#footnote-1)

# BUSINESS UNDERSTANDING

## Background

Wonderful Wines of the World (WWW) has been present in the wine market for 7 years. The company aims to provide customers with a premium selection of wine and wine accessories.

The key persons in this business are the owner (Fernando Bação) and the managers (João Fonseca and David Silva). Fernando is interested on increasing wine and accessories selling. João and David are looking on the actions needed to get the outcome the owner expects.

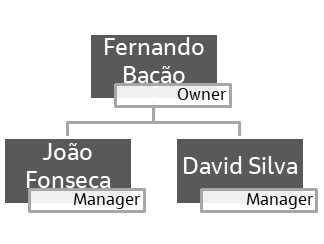


Figure 2.1 - Organizational chart.

During the 7 years of existence, this company has been using a marketing strategy based on previous experience on how to sell more: send a catalog (which is renewed every 6 weeks) to the 350,000 customers on the list (from the past 4 years) and expect the customer to approach the company to buy wine and accessories. Also, from previous analysis, the company already knows that most of its customers are wine lovers who have no financial constrains to get good quality wine.

The customers currently have three different ways of purchasing wine and accessories from WWW: in person (through one of the ten stores WWW has in major cities around the USA), by telephone (through the catalog) or online (on WWW’s web site).

The problem area this business is now trying to achieve is how can they improve the wine and accessories selling (not only to existing, but also new customers) using the knowledge on the current customers and this is why they have reached Data4Business Consulting.

## Business Objectives

The customer’s primary objective is to increase wine and accessories selling by understanding the following:

* How can we group the customers according to key characteristics and what are those characteristics.
* How many different segments of customers are there in the database from the past 18 months.
* How can the we apply the knowledge on the existing customers to reach new (and existing) customers from each segment and which ones should be prioritized.

## Business Success criteria

The expected outcome will be well defined customers’ segments which can make possible to build a customized marketing strategy. The success of the proposed task will be evaluated by WWW’s owner and managers.

## Situation assessment

This project will be conducted by the following team of four data scientists from Data4Business Consulting:

* Débora Santos
* Diana Furtado
* Pedro Medeiros
* Rebeca Pinheiro

We have been provided by the WWW’s IT team with a database of the customers who purchased in the last 18 months, composed by 10,000 customers and a metadata file on this dataset.

### Terminology

???????

### Risks and contingencies

Table 2.1 identifies a list of risks and contingency proposed.

|  |  |
| --- | --- |
| **Risk** | **Contingency** |
| Redundant/ non-relevant features | Work with remaining features or ask for different variables |
| Text | Number |

Table 2.1 – Risks and contingency

If applicable, describe also risks and contingencies, terminology, and costs and benefits.

## Determine Data Mining goals

Segment customers according to their willingness to purchase wine and accessories, considering their demographic and social information (age, years of education, presence or absence of children, income, etc), their 18 months’ records of commercial information (purchases, complaints, websites visits, etc).

## Project Plan

XXXXXXXXXXXXXXX

# PREDICTIVE ANALYTICS PROCESS

In this section we go through the process of understanding and preparing the data for modelling, the modelling itself and different algorithms used and, finally, the results evaluation.

Describe only the major steps involved in the process. Do not replicate what is already described in the Notebook. If necessary reference the reader to the Notebook.

## Data understanding

At this stage we analysed the data to get a better understanding of the features, what do they mean, how are they distributed, if there is noise, missing and/or duplicated values we should process, which features are relevant for the objective and which features are redundant.

The dataset has 10,000 observations (customers) and 29 features, from which 10 categorical and 19 numerical.

Pandas profile

## Data preparation

In the project plan we have decided to use K-means algorithm for clustering and for this reason we have decided to focus on cleaning only the numeric variables data, since those are the ones contributing to the model.

Looking at the dataset we concluded there were no missing nor duplicated values.

To check for the presence of outliers on the numeric variables we looked at the box and whiskers plots for each numeric feature, Figure 3.1 below, and concluded that features *Freq*, *LTV*, *Sweetred* and *Dessert* seem to have outliers.

BOX PLOTS APENAS DOS FEATURES Q VAMOS EXCLUIR OUTLIERS

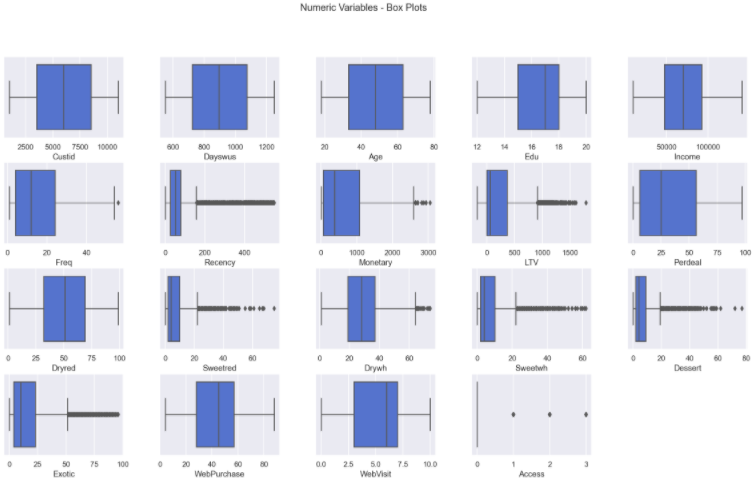
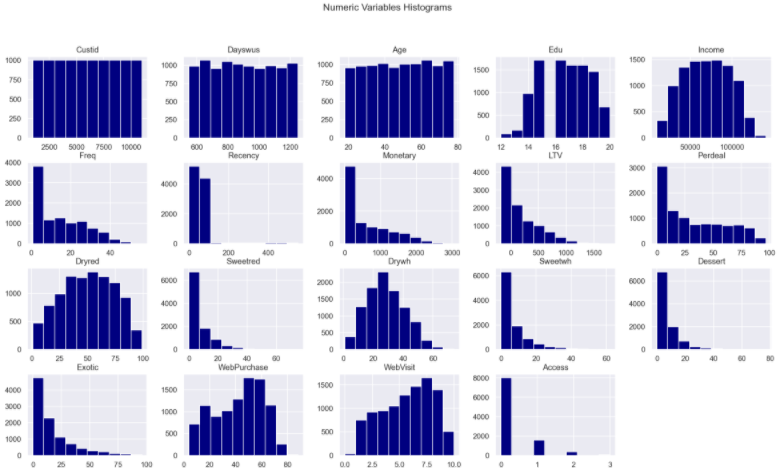


Figure 3.1 - Box and whiskers plot for numeric features.

We then looked at the numeric variables distribution to ensure

What did we conclude from the pairwise relationship between variables visualization? + Not to include following graph

Aiming to use K-means we needed to ensure the numeric features follow a normal distribution. For this reason we have used histograms to verify its distribution and skewness and concluded that…..



RobustScaler is based on percentiles and therefore not influenced by a few numbers of very large marginal outliers.[14].

## Modeling

Bla, bla.

## Evaluation

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

# RESULTS EVALUATION

Describe the degree to which the model meets the business objectives. If that cannot be done without the application of the model in a real environment, describe how could that be done.

Assess the data mining results in respect to the business success criteria.

# DEPLOYMENT AND MAINTENANCE PLANS

Describe how the strategy to deploy the model into production (necessary steps, persons involved, systems that may require changes, etc.).

State how after deployment the model’s performance should be monitored and maintained.

# CONCLUSIONS

Final remarks on the project.

## Considerations for model improvement

Bla, Bla

# REFERENCES

Author, A. A., Author, B. B., & Author, C. C. (Year). Title of article. *Title of Periodical, volume number* (issue number), pages.

[1] Chapman, P, Clinton, J, Kerber, R., Khabaza, T., Reinartz, T, Shearer, C. & Wirth, R. (2000). *CRISP-DM 1.0*, CRISP-DM consortium

[2] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html

# APPENDIX (OPTIONAL)

1. Example of a footnote. [↑](#footnote-ref-1)