

**BUSINESS CASES WITH DATA SCIENCE**

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

***Wonderful wines of the world market segmentation***

Group S – Business Case #1

Antonio Carvalho, 20200642

Frederico Rodrigues, 20200583

Gonçalo Carvalho, 20200664

Manuel Borges, 20200596

February 2021

INDEX

[1. INTRODUCTION 1](#_Toc65252625)

[2. BUSINESS UNDERSTANDING 2](#_Toc65252626)

[2.1. Background 2](#_Toc65252627)

[2.2. Business Objectives 3](#_Toc65252628)

[2.3. Business Success criteria 3](#_Toc65252629)

[2.4. Situation assessment 3](#_Toc65252630)

[2.4.1. Risks and Contingencies 3](#_Toc65252631)

[2.5. Determine Data Mining goals 3](#_Toc65252632)

[3. PREDICTIVE ANALYTICS PROCESS 5](#_Toc65252633)

[3.1. Data understanding 5](#_Toc65252634)

[3.2. Data preparation 5](#_Toc65252635)

[3.3. Modeling 5](#_Toc65252636)

[3.4. Evaluation 5](#_Toc65252637)

[4. RESULTS EVALUATION 6](#_Toc65252638)

[5. DEPLOYMENT AND MAINTENANCE PLANS 7](#_Toc65252639)

[6. CONCLUSIONS 8](#_Toc65252640)

[6.1. Considerations for model improvement 8](#_Toc65252641)

[7. REFERENCES 9](#_Toc65252642)

[8. APPENDIX (OPTIONAL) 10](#_Toc65252643)

# 

# INTRODUCTION

This report was written to analyze and cluster the dataset provided by the wonderful Wines of the World. While reaching this goal, the data handling was processed to extract the maximum value and minimum redundancy.

Wonderful Wines of the World is in the wine industry for seven years. Four years ago, they started building a database with all their customers' purchases and various characteristics, like income, age, frequency, and some others. The wine industry has much competition between producers, distribution, and retails - Thousands of different wines around the world are produced and consumed every day, and according to wine specialists - it's a very sensitive business once its customers should have a very personalized and detailed approach in the sense of being able to know the right information about what place, price, product, and promotion to use to each customer. This strategy, known as Marketing mix, is the one WWW wants us, as Data Scientist consultants, to prepare to improve their relationship with the clients, reach new ones, and distinguish which ones to prioritize.

Marketing mix refers to the set of actions an organization uses to promote its brand. As explained above, this marketing methodology is commonly grounded in 4Ps of marketing – Price, Product, Place, and Promotion – the best way to *put the right product in the right place at the right price with the proper promotion*. This way becomes easier for wonderful wines of the world to meet their customer's needs and demands.

With this in mind, our project was established to support Wonderful Wines of the world in creating a customer segmentation with the commitment of distinguishing and characterize clients. Understand how WWW can reach its groups of clients through these four principles is our priority in this report.

# BUSINESS UNDERSTANDING

## Background

To properly understand the project and its business, it's crucial that the first phase starts by understanding the background. Wonderful wines of the world is a seven-year-old international company, also known as WWW, that focuses its business on selling wines from all over the world—characterized by distributing from the producer to the consumer different and interesting wines through several channels worldwide.

With a robust website, ten small stores in major cities around the USA, new catalogs sent to its customers every six weeks, and a telephone retail channel, WWW have a wide variety of approaching its clients. Added to the fact of having aggressive promotions actions in wine and food magazines. Therefore, one of this company's priorities is to understand better their client's behavior with lots and different clients.

Chart, bar chart

Description automatically generated

Figure 2.1 – number of purchases in the past 18 months

Chart

Description automatically generated

figure 2.2 – Income per household

Throw-out these two figures, we can quickly understand a bit more about customers' income and purchases. In figure 2.1. we see a considerable drop of purchases from the fifth item to the six – passing from more than 500 clients purchasing five items in the last 18 months to just more or less 200 clients buying six items. Regarding figure 2.2. there is a regular income per household from WWW's customer's sample.

## Business Objectives

WWW pretends to know more about the different types of customers and their behavior to improve their relationship. According to the company - "*Our customers are highly involved in the wine industry and have sufficient money to indulge their passion for wine*" - however, they don't have specific knowledge about customers once until now all actions were based on feedback from salespeople and intuition.

Our objective as a business consultant is to improve and understand how WWW's customers behave and its clients' different characteristics to design a better and efficient marketing approach. For this analyze was provided by the company a dataset with almost 3% of all clients with several variables/characteristics about them. This marketing approach has been made with different marketing tools and perspectives.

## Business Success criteria

Our main criteria to evaluate the success of the project in a more overall and broadway is to provide the most useful information about each type of customer to reduce uncertainty surrounding costumers, including everything about them, from purchases to wine preferences, which is crucial for better business performance.

In a more objective and specific method, we also measure success from three parameters: 1. decrease the churn rate by identifying potential clients lost and the ones about to stop buying. 2. XXXXXXXXXXXXXX 3. XXXXXXXXXX

## Situation assessment

As already mentioned above, this project was made available by the wonderful wines of the World organization two documents: 1. An excel with all data from WWW's costumers with 10.000 observations (purchases from clients in the last 18 months) and 30 variables (containing income, purchases, wine types, between lot more). This excel was extracted from a dataset with a time span of 4 years with around 350.000 customers' purchases. 2. A metadata file with the general context and all variables explained.

For analyzing and segmenting the customers, our team of three Data Scientists consultants has used Python programming language on a Jupyter notebook file containing all phases of the customer segmentation development. For all tasks, including everything from choosing and changing variables to creating clusters and final analysis, we used resources from the libraries: Pandas, NumPy, SciPy, and Sklearn for processing data, Matplotlib, Seaborn, and Plotly for data visualization. CONFIRMMMMM

### Risks and Contingencies

After carefully analyzing the resources through different techniques and python tools, it was considered as a risk a small number of outliers that were removed and some redundant features XXXXXXXXXXXX

## Determine Data Mining goals

State project objectives in technical terms. NÃO ESQUECER TECHNICAL TERMS

Data mining goals and objectives have been developed while analyzing the resources provided by WWW and in coherence with the business goals. in accordance with *Wonderful wines of the world* needs and requirements the main Data mining objectives established in a first phase of Data exploration and preparation were:

1. Check for incoherencies in the variables – by visualizing individually the most important variables age, income, visits to the website and number of purchases.
2. Compute the correlation between variables – if high, small or normal.
3. XXXXXXXXXXXX

After the data preparation is ready, the data mining process proceeds to the modelling phase where the goal is to arrange and come up with a precise and detailed group of clusters to segment WWW´s customers. For this, in a more technical explanation, we have …. XXXXXXXXX

# PREDICTIVE ANALYTICS PROCESS

Describe only the significant 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

Bla, bla.

## Data preparation

Bla, bla.

## 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.

Once there is no possibility of exactly knowing real world consumers, potential clients and how customers’ segmentation will be treated and approach, our group has prepared a fictious, but precise, way of describing each type of client in a more interesting method using Personas – personas is a collective image of a segment of a company target, each persona doesn´t represent the entire target but represents a fraction of a target audience.

As a result of the segmentation made and explained above, we can visualize different clients with singular characteristics and tastes who have bought from WWW in the past 18 months. Having this into account, the following personas have been created:



1. **CLUSTER 0 -** **Willian Gates**, 77 years-old, from Jacksonville, Florida. William has a master’s degree in management, has a incredibly high income, and likes to buy wines very frequently to WWW, and therefore he has much value for the company. He doesn´t pay attention to promotions, and he doesn´t even like to buy wines online. Its favorite wine type is Dry white, and he dislikes unusual wines.



1. **Cluster 1 - Marry Lee**, 37 years old, has a Ph.D. in International Law and works in a Lawyer's firm in San Francisco, California. She has a strong preference for Dry Red wines, being practically the only type of wine she consumes. Marry is enthusiastic about buying wines online and visiting WWW Website. Sometimes she enjoys taking advantage of discounts.
2.  **Cluster 2 - Henry Jones**, 25 years old, a student from New York. As an ordinary bachelor student with a very low income, Henry can´t buy wines frequently, but when he buys, he doesn´t really care if the wine is on discount. He loves all types of wines except Dry Red, probably because it’s the most expensive one, and he almost always buys it online.

1. **Cluster 3 - Mark Smith**, 55 years, Professor in Atlanta, Georgia, has a usual education with a pretty satisfactory salary, he likes to buy wines frequently but doesn´t show to give much importance to discounts. He likes every type of wine but specially Dry Red.

# 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.

# APPENDIX (OPTIONAL)