**SMM750 Group Assignment**

**Group 5:**

Linh Nguyen ()

Soumya Ogoti (220045527)

Wenxu Tian (210059418)

Aparna Viswanathan (220004767)

Fan Xia (200039922)

1. Business scope and methodology

During the pandemic period, global beverage market has been under a negative impact due to closure of restaurants, bars, and clubs, as well as a sharp decrease in tourism sector. However, the boost in e-commerce sales has partly offset the adverse trend in wine industry (Wittwer & Anderson, 2021).

To establish the preliminary understanding of the wine market, the team investigated online wine consumer purchasing behaviour. The reasons for online wine shopping can be categorised into three aspects: **cheaper price**, **detailed product description**, and **more options available** in online shops (Bonn et al., 2016; From The Vine, 2022). In this case, we determine that our analysis on wine market will be based on information surrounding wine details (including wine type, country, ABV and year), price (for a bottle of 75 cL), and reviews (including number of reviews and scores) which is going to be collected from competitors’ websites for analysis.

To be clearer, the key objectives are to: (1) identify the most popular wines sold across the companies and the key features of best-selling wines; (2) the most frequent price range for wines in order to learn the product portfolio and business focus (e.g. how much should be invested on mass-market wines and niche-market wines respectively in our e-commerce shop). Therefore, web scraping is utilised to access data and BeautifulSoup and Selenium are applied as tools for scraping data from websites.

Given the fact that this online wine shop plans to target a broad market scope with various number of wines, the team carefully selected four online retailers which provide wines in a variety of selections with detailed product descriptions which can be useful for analysis. The four companies are **Laithwaites**, **Virgin Wines**, **Decantalo** and wine section page on **Morrisons** website. To be more specific, the team selected Morrisons wine as a good representative of online wine retailer of mass-market wines as the wines at Morrisons target consumers with a limited budget. What is more, Virgin Wines is also taken into account due to the fact that unlike Morrisons, Virgin Wine not only has cheap wines but also sell middle-end wines which are priced over £100. Laithwaites, which provides wines from £2 to £1,400 is also considered to be valuable in this case as the fine wines in its product offering will increase the analysis accuracy. Last but not least, apart from all three British brands mentioned above, other markets outside of the UK is also considered. A website called Decantalo which is based on another significant wine market, Spain, is analysed on the ground that it offers products to an international market.

1. Variable Description

In this section, a table of variable description is presented to explain the variables in more details.

Table 1 Variable Description

|  |  |  |
| --- | --- | --- |
| **Variable** | **Description** | **Unit** |
| name | Name of the wine |  |
| wine\_type | Type of the wine, e.g. red wine, white wine, sparkling wine, rosé wine etc. |  |
| country | Country of the wine where it was produced |  |
| country\_code | Country code of the wine where it was produced | Country code |
| logprice | Log price of the wine per bottle | Price in GBP (£) |
| scale\_price | Scaled price of the wine to 75 cL | Price in GBP (£) |
| year | Year when the grapes for the wine were harvested | Years |
| score | Rating score of the wine | Lowest (0) to highest (5) rating |
| num\_review | Number of reviews of the wine | No. of reviews |
| abv | Alcohol by volume, which measures alcohol content of wine | Percentage (%) |
| age | Age of the wine which is calculated by deducing the year from current year (2022) | Years |
| size(cL) | Volume of the wine | cL |

In terms of ratings, for wines with less than 5 reviews, the ratings are excluded from our analysis as the number of reviews are not sufficient to justify. In this case, all the wines with less than 5 reviews have a rating score of 0. Furthermore, since the dataset for our analysis is quite large, log of price is taken to visualise the price distribution of different wines. As the bottle sizes vary among wines, in order to compare price of wines on a fair basis, the team scaled price of the wine to 75 cL.

Besides, the age of wines is calculated by deducting year when grapes was harvested from the current year (2022). For the wine type, apart from the main types such as red, white, sparkling and rose, there are some other wine types such as orange and sherry which account for a relatively small percentage. Therefore, other wine types except red, white, sparkling and rose are categorised as “others” for further analysis.

1. Results and Analysis

Missing values for price are dropped etc.

1. Conclusion and Evaluations

During the data scraping process, there are four main problems we have encountered: **IP blocking**, **slow and unstable load speed**, **different webpage structures**, and **poor quality of data**. In terms of IP blocking, one example is that for some website such as Waitrose Cellar and Ocado, they are highly secured which did not allow web scraping. This was identified at the initial research stage and as a result, Morrisons was chosen as its alternative. To deal with the loading content issue caused by a large number of access requests, Python time sleep function is utilised to add delay while executing a program. Another challenge we faced was that different websites have different web page structures. For example, for Laithwaites website, year of the wine is included in name of wine while for the rest of the other websites, it is available in product description. Therefore, it requires customisation on the code to get all the information needed. With regard to the problem of poor data quality, there exists some formatting issue such as unnecessary spaces or special characters between words. The spaces and special characters are removed to obtain the exact value for data.

While processing and analysing the data, the team detected that some wines have very few reviews which can be quite subjective to users. After checking the dataset, results showed that the number of reviews for most wines fell into the range between 0 to 20. In this case the team decided to only include wine scores with more than 5 reviews to eliminate subjective reviews.

However, there’s a limitation in analysing the country of wines. As Decantalo is a Spanish website and we observed that over 50% of the wines sold on this platform are produced in Spain. Therefore, the results of countries can be biased.

**References**

Bonn, M.A., Kim, W.G., Kang, S. and Cho, M., 2016. Purchasing wine online: The effects of social influence, perceived usefulness, perceived ease of use, and wine involvement. *Journal of Hospitality Marketing & Management*, *25*(7), pp.841-869.

From The Vine, 2022. Shop Smart: Reasons To Buy Your Wine Online. [online] Available at: https://www.wtso.com/blog/shop-smart-reasons-to-buy-your-wine-online/ [Accessed 31 Oct. 2022].

Tanford, J.A., 2006. E-commerce in Wine. *JL Econ. & Pol'y*, *3*, p.275.

Wittwer, G. and Anderson, K., 2021. COVID-19 and global beverage markets: Implications for wine. *Journal of Wine Economics*, *16*(2), pp.117-130.

Draft:

<https://www.mdpi.com/2306-5710/4/4/87/htm>

<https://www.researchgate.net/publication/46509070_Market_Segment_Analysis_to_Target_Young_Adult_Wine_Drinkers>

<https://www.emerald.com/insight/content/doi/10.1108/17511060810864624/full/html>

Among the four retailers, Decantalo is a Spanish company which targets an international market and a

Additionally, analysis of Decantalo is considered to be valuable as the company offers a wide range of fine wines with a value of over £1,000 and it also target a wider market not only the UK but also internationally. The prioritisation of these four sources is also ranked by the importance and complexity of gathering the data for the analysis. Laithwaite and Virgin Wines are put at the top priority as it has a broader product range in wine and details of the wine can be easily accessed. After information from Laithwaite and Virgin Wines is obtained, Morrisons and Decantalo are added to the list to increase the accuracy and sufficiency of the information gathering process.

One of the main limitations in our evaluation process is that due to the limited time and sources, our analysis only evaluate type, country, ABV and year of the wine while some variables such as aroma and grape variety can also affect consumers selections. However, this type of information is not always available, and it can also be hard to be categorised or converted to a numerical value for our analysis. What is more, as cost varies among different wines and it can be obtained from the websites, profit margins cannot be analysed for further implications. This problem also applies to customers as we failed to analyse the customer group (such as sex, age etc.) and their purchasing behaviour (which is associated with a number of factors such as salary, willingness to spend on wine etc.). Delivery cost is another factor we ignored in this research as different retailers offer different delivery options which lead to a various number of pricing methods. Hence, it is very challenging to analyse the delivery costs under this circumstance. Last not least, promotions and deal offer are also not taken into consideration in the analysis.

that it lists most of the middle-end wines. In addition to Morrisons and Virgin Wines, analysis of Laithwaites, which provides 1,200 wines up to £1,400, is also considered to be valuable in this case as the company provides fine wines which appears to be a good option for niche-market consumers.

The key objectives to gather the information from our competitors’ websites is to identify the best-selling wines, the most popular wine type, country and year among customers, as well as set reasonable prices for a variety of wines.

Firstly, an HTTP request was sent to URL of the webpages which the team plan to access. After all relevant information of the wines is obtained, t

*“In the first stage, the application connects to the company’s server in order to obtain all the published products. To do so, the company has provided us with access to its API [1]. By means of a HTTP[2] request and using the provided credentials, the application can access the server Website Quierovinos to obtain all the products with their corresponding attributes. All information obtained is exported to a JSON[3] file and imported to a local database. An automatic process is executed once a day to keep this information up to date. The frequency of this execution can be changed according to the company’s requirements. Also, all the information obtained is displayed through a private web application developed with Django framework[4] that allows the manager of the company to read the processed information in a friendly way and help him on performing several actions. Once all this information (purchase price, margin, shipping price...) provided by the company of all the products is stored on the local database, the second stage consists of obtaining data from the Marketplaces about competitors prices for the analyzed products. To do so, we use the Web Scraping technique to get all interesting data from the Marketplaces websites regarding our target products. This information is exported into a JSON file and stored into the local database extending the previous data.*

*Next, once all this data is collected and stored together in the local database, another computer process analyzes each product to determine if it is possible to reduce the price based on the following parameters defined previously by the company manager”*

In addition, years, types, and countries of all wines are studied to find out the top 5 popular years, the most popular wine type, and the most popular countries across the four companies