Research on Arbitrage Opportunities for Wine from Norway to the United Kingdom

Group 4:
Giulia Capuano, Artem Demin, Alex Efthymiou, Theodor Harket & Yuanheng Li

Arbitrage opportunities arise from differences in price among markets. The wine business in Norway can be characterised as a monopoly market in which the average cost of everyday wines is higher than that in the United Kingdom (UK). In contrast, the average price of premium wines is lower than that in the UK market. Therefore, we can exploit this arbitrage opportunity by buying premium wines from Norway and importing them to the UK.

1. Research Plan

Government-owned monopoly, Vinmonopolet controls the beverages business in Norway and has the exclusive right to sell alcoholic drinks with over 4.75% alcohol by volume. The monopoly was founded to prevent the citizens' overconsumption of alcohol and to remove the over-profit from selling strong beverages. As the monopoly in the business, Vinmonopolet has strong pricing power over suppliers, which limits the suppliers' profits lower than that in other markets. The pricing method used by Vinmonopolet is entirely transparent; the retail price is based on the principle that each product should cover its costs and giving Vinmonopolet a markup of NOK 8.90 (equivalent to 74 pence) per litre plus a maximum of 22% of the purchase price. The percentage markup gradually declines for more expensive wines; the maximum markup is NOK 110 (equivalent to £ 9.16) per unit. With a markup calculated as mentioned above, the final prices for customers are higher in common wines (higher markup and taxes than in the UK). The retail prices for premium wines are much lower than that in the UK (high pricing power over suppliers and low markup and taxes compared to that in the UK).

We can exploit this arbitrage opportunity by buying expensive wines from Norway and importing them to the UK. To maximise our profits in this arbitrage opportunity, we decided to consider the following questions:

1. What countries and regions produce the most popular premium wines for UK customers?
2. What is the most preferred category of wines for UK customers? (i.e., red, white, rose, sparkling, fortified)
3. What are the prices of these wines in the Norwegian market?
4. What are the prices of these wines in the UK market?

To answer these questions, we decided to analyse the following data sources.

2. Data Sources

Tweets are analysed to determine the preferred types of wines in the UK market. At the same time, Google Trend is exploited to study the geographic distribution of countries and regions where the most popular premium wines for UK customers are produced. Two different websites are also used as the source of prices: The Vinmonopolet e-shop displays our costs to buy these wines from the Norwegian market, and Vivino, an international online wine marketplace with a database of more than 12 million wines, exhibits the possible sales prices of the most popular premium wines in our targeted market are produced. Also, Vivino is one of the largest e-commerce wine marketplaces with a highly engaging community, actively setting ratings and sharing reviews with others. Vivino presents an extensive list of wines from many regions, offering consumers a wide choice for any occasion. Figure 1 shows the pipeline of this work.

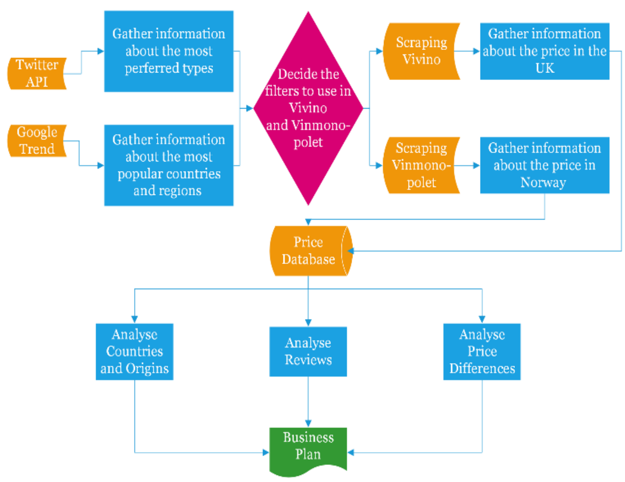

Figure 1. Workflow

3. Data Gathering

The Twitter API is used to gather the number of tweets in English about the main wine types. We also use the Google Trend API to analyse the search trend of wine-producing regions. The two websites mentioned above, Vivino and Vinmonopolet, are dynamic, so we use Selenium for scraping.

3.1. Twitter Analysis

To determine the ranking of different types of wine based on the UK population's preferences, we decided to look at tweets originating from the UK in the past week mentioning the main categories of wine – red, white, rose, and sparkling. We used the Twitter API and the Python package Tweepy to do so. As there is a limit to the number of tweets one can request per day, we decided to request 1000 tweets for our search. The first problem we faced was filtering the tweets by geolocation. There are two options here: one is a location input by the user when posting the tweet, and the other is the user's geolocation at the time they are making the tweet, access to which requires the user's permission. Both options returned very few tweets from the UK in the last week. Hence, we decided to remove the location filter and assume that all English speakers have the same preferences.

We requested 1000 original tweets (meaning no retweets) in English from the past week regarding a combination of keywords. More specifically, we searched for "red wine OR white wine OR sparkling wine OR rose wine". Since the order of the key terms in the request affects the resulting tweets, we repeated the request four times, with each category in the first position once, and summed the number of tweets to obtain the ranking of the categories. Based on this basic, high-level analysis, we found that red wine was tweeted about the most, followed by white, sparkling, and rose (See Figure). We assumed this ranking indicated preferences and used this to guide how many products within each category of wine we should stock.

Looking at the tweets we got, there were instances where tweets did not precisely mention what we were looking for, i.e., they said the words red and wine, but not sequentially, and had a different context. When iterating through the tweets to count the number of times each category was mentioned, we ensured in our code such cases did not yield a count, hence cleaning the data in this way. A further point to consider is that we only measured how many times a wine type was mentioned, but this is not a great measure of preference. For example, some tweets said more than one type of wine because they were comparing them, indicating a preference for one, but our method meant both categories got the point. Further, we did not identify the sentiment of the tweet which would have given us valuable insights into preferences.

3.2. Google Trend Analysis

The use of Google Trends was made to analyse the wine market in the UK broadly. Specifically, the purpose of the analysis was to understand the market's preference regarding the type of wine, between red, white, rose and sparkling, within the most popular regions that produce wine in Europe.

Before scraping the websites Vivino and Vinmonopolet and receiving the dataset with the most popular wines and regions they are produced, we decided to research red, white, rose and sparkling wine associated with the most common areas in Europe where wine is produced. Through this, we could get a clearer idea of the preferences of our target market and consequently deliver a service that would fulfil a gap in the current market.

The chosen wine type plus the region were selected using the Google Trend API ("red wine + x region", "white wine + x region", "rose wine + x region", and "sparkling wine + x region") to determine the most popular type of wine. Additionally, a time frame of one year was added to obtain recent and still relevant data to analyse, and of course, the research was limited to the UK.

In each dataset created, an additional column of "total wines" was added to give an overview of the region's importance in general and determine the most popular wine-producer areas.

Finally, regarding the type of wine, the results were pretty evident from the beginning of the analysis since "red wine" was always the most researched term within all the regions, compared to the other type of wines.

3.3. Web Scraping Vinmonopolet

Vinmonopolet is, as abovementioned, a dynamic website, so we used Selenium to scrape information from the website. Using the techniques learned in previous lectures, we scraped over 5000 wines, focusing on all types of wines costing £60 or more, to identify opportunities to profit from premium wines.

We created eight variables (see Appendix for lexica): Name of the wine (Producer + Name), Year of production, Price in GBP, Centilitres in each bottle, Producer country, Category, Producer region, and Ordering id. All the aforementioned were thoroughly considered and included because of their necessity to analyse and increase the profitability of our venture e-commerce.

When scraping Vinmonopolet, a Norwegian page, we came across several issues; the major problems are discussed here. The primary concern is language, which introduces two problems: translated words and special characters. Special characters in terms of letters are not included in the UK alphabet. We solved this by creating a .csv-file, including all translations necessary, translating the words at the end of the scraping algorithm.

The second problem is that Vinmonopolet includes the producer in its full wine name, so we could not separate the producer and wine name, troubling the matching of this dataset and the Vivino dataset, as the last separates them.

3.4. Web Scraping Vivino (UK)

Vivino is a dynamic website as well. The scraping procedure for Vivino is very much like the one mentioned above. We intended to scrape wine in the premium sector with a price above £50 from Vivino. Several problems were faced during the scraping:

1. The price configuration was presented using a drag slider with no input boxes.
2. The price wasn't always linked to the same HTML element.
3. The automatic window scroll was often stuck when the number of resulting wine cards exceeded 1000.

These problems were solved using the below-listed approaches:

1. We found a Selenium function called "Drag and Drop", which can be used to move price sliders
2. All possible elements linked to the price of wine have been identified.
3. To avoid computational issues, the scraping data have been split into parts based on price intervals and wine categories. The red wine category (the largest one) within each price interval is scraped separately from the other wine categories. This separation of scraping data ensures the smooth work of the Selenium scroll function.

4. Data Processing & Visualisation

4.1. Most Preferred Wine

Using the data gathered with the Twitter API, we created a pie chart displaying the percentage of how many times different types of wines are mentioned in recent tweets (fig. 2.a). As shown, the most said wine is "red", followed by "white", "sparkling", and "rose".

4.2. Most Popular Wine Regions

The datasets of each region from Google Trend were combined, and the line chart shown was created to showcase the research results that support our business's core idea. We made a Time-Series chart and four pie charts. The line chart displays the search trend of different wine regions over time (fig. 2.c), and the pie charts illustrate the distribution of searches among the four most popular wine regions (fig. 2.b). The line chart can visually highlight the most active region

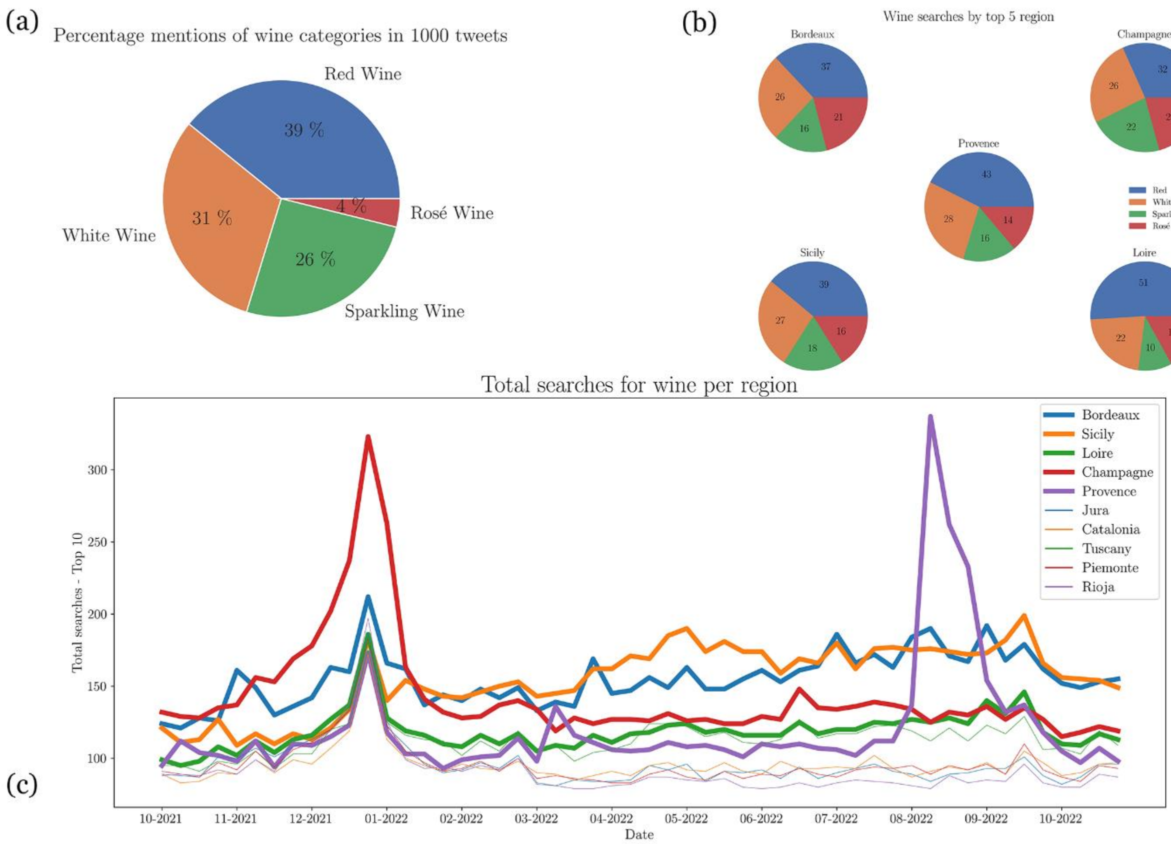

Figure 2. Twitter and Google Trend Analysis

in wine production, such as Champagne, which was mainly researched around the end of 2021 and 2022, indicating it is trendy during festive times; Provence, with a significant peak during summer 2022; and Bordeaux and Sicily which remain constantly higher than the average.

4.3. Geographical Distribution of Regions of Premium Wines sold in Vinmonopolet

The data gathered from Vinmonopolet include the countries and regions of the premium wines. We created a map showing the most mentioned regions in the Vinmonopolet dataset (fig. 3a). We used the World Cities Database to plot the geographical figures and visualise the geographical distributions of wines. As the map highlights, France represents the most significant wine producer, followed by the Northern part of Italy.

4.4. Years with Most Highly Rated Wines

One of the advantages of Vivino over Vinmonopolet is Vivino displays the rating of the wines on the webpage. Because of this, we can create a bar chart showing the number of reviews for wines produced between 1975-2021 and whose average rating is higher than 3.8. The charts' bars are coloured based on the average rating of all expensive wines made that year. (fig 3. c)

5. Combined Database Analysis

The two data frames, gathered from Vinmonopolet and Vivino, are combined for price comparison. The steps for completing this task are as follows:

1. Create a double for-loop to iterate over all wines in both datasets.
2. Use a package called SequenceMatcher that takes in two strings, and compares their similarity, returning a percentage.
    a. Vinmonopolet always includes both producer and wine name, Vivino does not always include both.
    b. Therefore, we put a minimum similarity percentage of 85%.
3. If we find wines that are 85% similar or more, we check if their year is the same (a crucial aspect to consider in the wine industry).
4. Last, we check if the price in Norway is lower than the market price.
5. If the wine passes all the checks, we store it in the dictionary as an arbitrage opportunity.

To visualise the arbitrage database, we also created a map visualisation displaying the wine regions that contain most wines for arbitrage. (fig. 3.b)

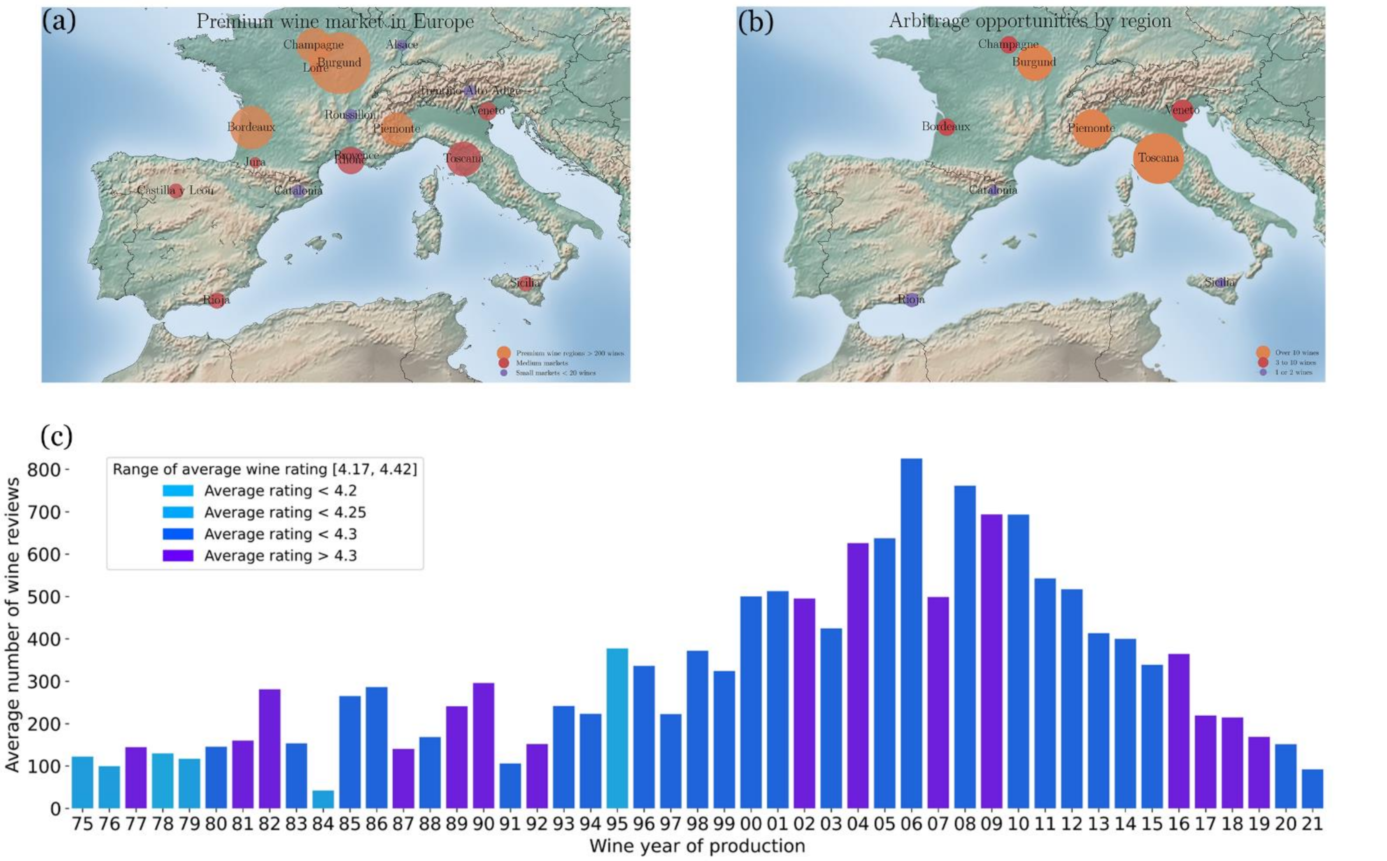

Figure 3. Vinmonopolet, Vivino, and Combined Dataset Analysis

	name	vpolet_price	vivino_price	region	country	ratings	Price Difference	Profit %
0	Ceretto Barolo Brunate 2015	69.55	227.43	Piemonte	Italy	4.3	157.88	222.69
1	Reyneke Cabernet Sauvignon 2017	77.38	225.00	Coastal Region	Sør-Afrika	4.6	147.62	186.90
2	De Montille Volnay 1er Cru Les Taillepieds 2018	106.09	260.00	Burgund	France	4.3	153.91	142.25
3	Silver Oak Alexander Valley Cabernet Sauvignon...	66.87	155.00	California	USA	4.6	88.13	127.31
4	Nin-Ortiz Nit de Nin Mas d'en Caçador 2018	77.09	167.07	Catalonia	Spain	4.2	89.98	112.83
5	Matrot Mersault-Blagny 1er Cru 2018	66.09	135.00	Burgund	France	4.4	68.91	99.73
6	Fossacolle Brunello di Montalcino Riserva 2012	73.04	144.95	Toscana	Italy	4.4	71.91	94.35
7	Pod. La Vigna Brunello di Montalcino Riserva 2012	73.91	144.95	Toscana	Italy	4.4	71.04	92.06
8	Ch. D'Yquem Sauternes 2009	196.52	372.68	Bordeaux	France	4.7	176.16	88.11
9	Casanova di Neri Tenuta Nuova Brunello di Mont...	86.51	165.00	Toscana	Italy	4.5	78.49	87.26

Table 1. Wines with Highest Progfit Margin

6. Arbitrage Opportunities and Future Perspectives

6.1. Arbitrage Opportunities

The combined dataset gives us a list of wines whose prices are higher in the UK market but lower in the Norwegian market. The wine import tax in the UK and the shipping cost from Norway to the UK should be less than £3 per bottle (GOV.UK, 2021). We can then calculate the price difference and the net profit margin of arbitraging this bottle. Table 1. shows the 10 wines with the highest price margin, where the profit range of the top 10 wines varies from 222.66% to 87.26%. According to the data gathered in the analysis, within the European region, Italy is the producing country in which the arbitrage opportunity is highest, with 4 wines within the 10 most profitable wines, and France is second with 3 wines.

Consequently, we can know what kind of wine we should focus on in the future. We can see from Figure 2 that the most preferred types of wines by Twitter users are red, white, and sparkling. Also, the most searched regions on Google are Bordeaux, Champagne, Provence, Sicily, and Loire. From Figure 3, we can see that the most significant wine region in Europe is Burgundy, and the most welcomed years for wines (later than 2010) are from 2016 to 2019. Therefore, we can focus on premium red, white, or sparkling wines produced by these regions in the mentioned years to make sure our products are welcomed in the UK market. Finally, the results of the previous analysis match the outcome of the market analysis as a whole.

6.2. Future Perspectives

For the start of the business, our focus was to exploit the arbitrage opportunity between the European prices (analysed with Vivino) and the more reasonable prices the Norwegian market offers. However, as Table 1 highlights, the new markets, such as the US and South Africa, are slowly gaining much popularity, so one of the expansion possibilities is to include the new market's wine in our catalogue to accommodate every customer's taste and acquire international customers as well.

7. References

1. GOV.UK (2021). Trade with Norway. [online] GOV.UK. Available at: https://www.gov.uk/guidance/trade-with-norway#tariffs-on-goods-from-1-december-2021 [Accessed 1 Nov. 2022].

2. Vinmonopolet (n.d.). *Vinmonopolet | Hjemmeside*. [online] www.vinmonopolet.no. Available at: https://www.vinmonopolet.no/ [Accessed 3 Nov. 2022].

3. Vivino (n.d.). *Vivino (United Kingdom) - Buy the Right Wine*. [online] www.vivino.com. Available at: https://www.vivino.com/GB/en [Accessed 3 Nov. 2022].

8. Lexica

8.1. *Lexicon for Twitter Data Collection*

This lexicon applies to the following files:

- df_tweets_red.csv
- df_tweets_white.csv
- df_tweets_sparkling.csv
- df_tweets_rose.csv

Variable	***Explanation***
user	*Individual's Twitter account username*
red_wine	*A variable indicating how many times "red wine" was mentioned in the tweet*
white_wine	*A variable indicating how many times "white wine" was mentioned in the tweet*
sparkling_wine	*A variable indicating how many times "sparkling wine" was mentioned in the tweet*
rose_wine	*A variable indicating how many times "rose wine" was mentioned in the tweet*
content	*Full text of the tweet*

8.2. *Lexicon for Google Trend Market Research*

This lexicon applies to the following files:

- Castilla.csv
- Jura.csv
- Alsace.csv
- Catalonia.csv
- Sicily.csv
- Tuscany.csv
- Loire.csv
- Piemonte.csv
- Bordeaux.csv
- Veneto.csv
- Trentino.csv
- Rhône.csv
- Roussillon.csv
- Rioja.csv
- Champagne.csv
- Burgund.csv
- Provence.csv

Variable	***Explanation***
Red wine	*It is associated with the number of searches red wine has been searched on Google*
White wine	*It is associated with the number of searches for white wine has been searched on Google*
Rose wine	*It is associated with the number of searches rose wine has been searched on Google*
Sparkling wine	*It is associated with the number of searches sparkling wine has been searched on Google*
Date	*The exact date on which the search on Google was made*
Region	*Refers to the wine region*
Total_wines	*Refers to the total number of wines searched on Google per day, regardless of the type of wine*

8.3. *Lexicon for Vinmonopolet Scraping*

- Wines_vinmonopolet.csv

Variable	***Explanation***
Name	*Combination of producer and wine where the general writing is producer + wine*
Year	*Year of production. If no year is included, we set the default value to Non-Vintage*
Price	*Price of wine in GBP*
Cl	*Centilitres in the bottle*
Country	*In which country is the wine produced*
Category	*The category of alcohol. (Red wine, white wine, sparkling wine, rose wine, liquor etc.)*
Region	*In which region is the wine produced*
ID	*The ID of the wine in Vinmonopolet*

- Translations.csv

Variable	***Explanation***
Norwegian	*Norwegian translation of the word*
English	*English translation of the word*

- Worldcities.csv

Variable	Explanation
City	*Name of city*
City_ascii	*Name of the city using ASCII-characters*
Lat	*Latitude of city*
Lng	*Longitude of city*
Capital	*Is it a capital or not (primary for capital)*
CapitalEurope	*Capital of a European wine country with more than 20 wines*

8.4. *Lexicon for Vivino Scraping*

- Wines_Vivino.csv

Variable	Explanation
WineBrand	*The producer of wine*
WineName	*The name of the wine*
WinePrice	*Price of wine in GBP*
WineRating	*Rating of wine on a scale of 0 to 5*
WineReviewCount	*Number of reviews associated with wine*
WineLocation	*The wine country of origin*
WineYear	*Year of production. If no year is included, we set the default value to Non-Vintage*

8.5. *Lexicon for Combining Datasets*

- Arbitrage_wines.csv

Variable	Explanation
Name	*Combination of producer and wine where the general writing is producer + wine*
VPolet_Price	*Price of wine in GBP buying from Vinmonopolet*
Vivino_Price	*Price of wine in GBP buying from Vivino*
Region	*In which region is the wine produced*
Country	*In which country is the wine produced*
Ratings	*Average user rating*