

Increasing online shop revenues with web scraping: a case study for the wine sector

Web scraping
for online wine
sales

Oriol Jorge

Department of Computer Science, University of Lleida, Lleida, Spain

Adria Pons, Josep Rius and Carla Vintró

Department of Business Administration, University of Lleida, Lleida, Spain, and

Jordi Mateo and Jordi Vilaplana

Department of Computer Science, University of Lleida, Lleida, Spain

3383

Received 16 July 2019
Revised 22 November 2019
Accepted 11 January 2020

Abstract

Purpose – Wine has been produced for thousands of years and nowadays we have seen a spread in the wine culture. E-commerce sales of wine have increased considerably and online customer's satisfaction is influenced by quality and price. This paper presents a case study of the company "QuieroVinos, S.L.", an online wine shop founded in 2015 that sells Spanish wines in two main marketplaces.

Design/methodology/approach – With the final target of increasing the company profits it has been designed and developed an application to track the prices of competitors for a set of products. This information will be used to set the product prices in order to offer the products both competitively and profitably in each Marketplace. This application must check, by tacking into account information such as the product cost or the minimum product margin, if it is possible to decrease the price in order to reach the top cheapest position and as a consequence, increase the sales.

Findings – The application improved in a notorious way the company's results in terms of sales and shipping costs. It must be said that without the use of the presented application, performing the price comparison process within each one of the marketplaces would have taken a long time. Moreover, as prices change very frequently, the obtained information has a very limited time value, and the competitors prices should be analyzed daily in order to take accurate decisions regarding the company's price policy.

Originality/value – Although the application has been designed for the wine sector and the two named marketplace, it could be exported to other sectors. For that, it should be implemented new modules to collect information regarding the competitor's price of the products selling on each corresponding marketplace.

Keywords Wine, Web scraping, Marketplace, Price, Revenues, Cloud

Paper type Case study

1. Introduction

E-commerce can be defined as "the process of buying and selling products or services using electronic data transmission via the internet and the www" (Grandon and Pearson, 2004), and it constitutes an important tool for online relationships between customers and companies (Verma *et al.*, 2015). Companies can intercept a bigger marketplace and moreover, online channels increase customer profitability and reduce the cost to serve them (Gensler *et al.*, 2012). Customers can easily gather and compare information of different products and prices

© Oriol Jorge, Adria Pons, Josep Rius, Carla Vintró, Jordi Mateo and Jordi Vilaplana. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial & non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at: <http://creativecommons.org/licenses/by/4.0/legalcode>

This work was supported by the Ministerio de Economía y Competitividad under contract TIN2017-84553-C2-2-R. Also, the authors are members of the research group 2017-SGR363, funded by the Generalitat de Catalunya.



British Food Journal
Vol. 122 No. 11, 2020
pp. 3383-3401
Emerald Publishing Limited
0007-070X
DOI 10.1108/BFJ-07-2019-0522

from different vendors without geographical limitations and with less time and effort, and consequently, the consumer search costs become lower for online shoppers (Smith, 2002).

Different authors have researched about e-commerce and have identified the main consumer motivators and impediments. (Fenech and O'Cass, 2001) found that perceived usefulness and attitude are important factors that positively affect online shopping. (Jarvenpaa and Todd, 1996) argue that the possibility to shop anytime and from anywhere is one of the most important advantages of e-commerce. (Alba *et al.*, 1997) mention the global choice as one advantage, (Krause, 1998) mention time savings both in the searching process and the transactions, and (Bhatt and Emdad, 2001) mention the quick comparisons of offerings and prices. Other authors have identified different barriers to e-commerce. For instance, (Kangis and Rankin, 1996) mention difficulties in quality evaluation, (Hoffman *et al.*, 1999) mention the lack of trust in virtual sellers, and (Udo, 2001) mention security risks in transactions. According to (Swaminathan *et al.*, 1999) consumers who value convenience are more likely to buy online, while those who value social interactions are less interested in e-shopping. However, (Kangis and Rankin, 1996) state that perceptions from consumers on the benefits and disadvantages of e-commerce differ across product categories. For a more extensive review on the main potential drivers and barriers for consumer adoption of e-commerce see (Anckar, 2003).

The existing literature on e-commerce highlights that online customer's satisfaction, and therefore repurchase intention, is influenced by cognitive aspects and affective online experience related to shopping websites (Rose *et al.*, 2012) like easy-access to information, or easy transaction. Moreover, online customer's satisfaction is also influenced by product and service quality and by price. Hence, the perception of the company product and service quality is often related to the overall reputation of the company (DiRusso *et al.*, 2011). Therefore, online shoppers search for different dimensions. Some consumers are brand loyal and tend to limit their searches to specific brands, while others search for the best price.

In the early years of e-commerce, it was expected that online competition would lead to one market-driven price for any given product, but despite this prediction, high levels of price differences still exist on the internet (Bailey and Bakos, 1997). Price differences may stem from market inefficiencies or from seller characteristics such as sellers' quality (DiRusso *et al.*, 2011) and may influence purchasing decisions of consumers. Hence, price-sensitive consumers tend to search for potential savings and this forces retailers to be more vigilant about their pricing strategies and to provide competitive prices in the marketplace. In this sense, consumers are increasingly using Internet shopping agents that provide access to price, product, and retailer information from multiple competitors (Lindsey-Mullikin and Grewal, 2006). Two dominant models for price aggregation sites are online marketplaces and shop bots. Shop bots are an organized list of products with links back to the original vendor sites to get additional information and to complete the transaction, whereas online marketplaces are more structured sites with specific rules for presenting product information and they host the transaction (DiRusso *et al.*, 2011). A good example of an online marketplace is Amazon Marketplace. The Amazon marketplace requires its sellers to provide a standard and accurate set of information and therefore, it exhibits a high degree of information governance (Ritala *et al.*, 2014). With marketplaces vendors can access to billions of global consumers and can perform more efficient transactions. Moreover, marketplaces are an interesting option to start on-line sales since companies only need to care about logistics management. The vendor receives an e-mail with the items to be shipped and the address of the customer. The sale and payment are all managed in the marketplace. For companies operating in marketplaces it is essential to find the right pricing strategy since it can be a decisive factor for online buyers.

Marketplace pricing may be influenced by various factors including marginal costs, competition, network effects, provider differentiation, transaction size and volume, quality

vs. quantity, and who pays the bill. Different authors have studied the topic of pricing and marketplaces. (DiRusso *et al.*, 2011) analyzed how differences in information governance affect online prices and price differences. The authors collected price data for ten electronics products from Amazon Marketplace and used a regression model to analyze their data. (Steenkamp, 1988) analyzed the relationship between price and quality in the marketplace through an empirical investigation that involved 413 product tests and 6,580 brands. The authors concluded that price is often a poor market signal of quality, however (Rao, 1984) argues that the price decision is perhaps the most significant among the decisions of the marketing mix (strategy) for a branded product. And (Jiang, 2002) developed a model of price search behavior in the electronic marketplace. As stated by (Kannan, 2001) the pricing of products and services sold over the internet channel is becoming more dynamic in part as a response to consumer use of price-comparison bots and marketplaces.

E-commerce has grown exponentially since mid 1990s, when the internet became a real option for consumers (Szolnoki *et al.*, 2016), but certain products, such as wine, have had more difficulties in achieving profitability in e-commerce sales. Despite this, the wine industry also looked at the opportunities to sell its products online and managed to find solutions to overcome the initial difficulties. In this respect, marketplaces represent a business opportunity in the online sale of wine and currently a large volume of wine sales are made through this distribution channel. Under this context, this article aims to develop a price-comparison for the wine sector, and more specifically for the marketplaces of the company “Quierovinos, S.L.”, an online wine shop that offers a selection of Spanish wines. All the information required is gathered by using a Web Scrapping Script that updates data continuously and helps to define the optimal price and therefore to position products, under competitive and profitability criteria, in each marketplace. As stated by (Milev, 2017) a Web Scrapping is a method to extract characters, words or sentences from web pages in order to be treated and analyzed afterwards. That is, a Web Scrapping is a technological solution that extracts data from web sites in a fast, efficient and automated way, offering data in a format that is structured and easy to use (Castrillo-Fernández, 2015). A Web Scraper first analyzes the structure of the page from where data will be extracted and then, these data are exported in JSON format and are subsequently stored in a PostgreSQL database. The final objective of this research project is to improve the benefits of the products of the company “Quierovinos, S.L.” in the marketplaces where it operates.

2. The wine industry and e-commerce

The knowledge of wine and grapes is part of the cultural history of mankind. Grapevine is the oldest cultivated plant. According to the Bible, Noah planted grapevine and made wine after the Flood; King Salomon considered wine to be the second greatest pleasure for mankind; and Egyptians used preparations of natural remedies mixed with wine against different diseases (Fehér *et al.*, 2007). During the last decades we have seen an increase in the consumption of wine due to different factors (Hall *et al.*, 2004). Wine may have potential health benefits. For instance, according to different studies, moderate consumption of wine has been linked to a reduction in the risk of death by heart disease and heart attack by 30–50% (Higgins and Llanos, 2015). Wine is a cultural product. Wine tourism and wine experiences are an important business and contribute to national economies beyond the sales of the product itself (Pitt, 2017). And most European cultures consider wine a refined choice.

The “World of Atlas of Wine” divides wine producing countries into two worlds. The Old World countries are traditional wine producing countries around the Mediterranean area and include Greece, France, Italy, Spain, Germany, Portugal, Austria and Hungary. The New World countries are wine producing countries settled after the European colonial expansion and include the United States, Australia, New Zealand, Chile, South Africa and Argentina

(Li *et al.*, 2018). The European production accounts for more than 60% of the world production. Specifically, Spain has a profound tradition in the production of wine and is one of the European leaders of this sector.

Wines are not simple products. They are produced from many regions and varieties. They have complex production requirements and the complexity extends from production to marketing and retail. Unlike other products, aging is a positive factor (Cox, 2009). The wine industry is one of the most globalized worldwide. According to the wine economist (Veseth, 2013) there are more than 15,000 wine brands sold worldwide, and there is commonly a wide price range of branded wine offerings in the average supermarket.

In a study conducted by (Hall *et al.*, 2004), the authors identified motivational factors and some age differences for wine consumption. In this line, (Levine and Pownall 2004) found that 60% of the wine consumption corresponds to the group aged 35–64 years. And (Hussain *et al.*, 2007) concluded that wine knowledge is the best predictor of consumption. Consumers value quality but also consider price to be a determinant factor in the purchasing decision. For instance, (Lockshin *et al.*, 1993) analyzed the effect of price and oak flavor on the quality perceptions of consumers and found that consumers judged wines mainly by price regardless of the oak level. Similarly, (Gergaud and Livat, 2007) found that price represents a substitute for umbrella branding where consumers are not very knowledgeable about wine, and price is then the major quality signal. (Lockshin and Corsi, 2012) developed an extensive review on consumer behaviour for wine. (Quester and Smart, 1996) state that consumer behavior can be explained in terms of product involvement, and (Bruwer and Li, 2007) argue that involvement with wine as a product can be influenced by the role wine plays in the consumers' lifestyle.

Before, wine could only be purchased in local wineries or in stores. Currently it can also be purchased online. As stated by (Dressler, 2018), online offers for wine are increasing as the general consumer attitude to buy online increases. (Quinton and Harridge-March 2008) analyzed a sample of store and online wine buyers and found that it is relevant that the online service instills trust for the first-time buyer.

Online wine shops offer bigger ranges of wines from small and big wineries than physical stores, and provide added assistance to the customer including suggestions, advice and helpful information for a successful purchase. "QuieroVinos, S.L." is an online wine shop founded in 2015 and established in the region of Catalonia (Spain) that sells Spanish wines. It operates in different marketplaces and the two main ones are Uvinum (Galtés, 2019) and Vivino (Llewellyn, 2019). Uvinum was founded in 2009, and it is one of the European leaders in the wine sector. It has more than 90,000 products in its portfolio and more than 100,000 registered customers. Currently, it operates in 14 countries, including France, the United Kingdom, Spain and Italy, with a turnover exceeding of 10 million euros. Vivino was founded in 2009 and it is considered to be the app number 1 in the wine sector. Currently its database consists of 1.3 million wines. The wine tag app Vivino provides information about the wines, different price options and users can post comments on the wines.

There is currently no application that tracks within the Marketplaces Uvinum and Vivino to get the best price offer for each product. However, since price is an important factor in the online purchasing decision of wine, there are diverse price comparisons available in the internet that allow to find the best price for a specific product, but unlike marketplaces they link the consumer directly to the vendor's website to make the purchase and this may intimidate the customer and discourage sales. The most common price comparisons in the Spanish wine sector are NoSoloVino, Comparan-doVinos and VinosWine. These price comparisons are used in the research presented to identify the wine online buyers. Considering that one of the biggest lever that online shops have to acquire and retain demand is pricing, without forgetting the importance to build a sustainable business as well as the need to invest in understanding the pricing dynamics, sooner rather than later, and become methodical about it, the authors briefly review price policies as part of marketing mix in the following section.

3. The role of pricing in the marketing mix for e-commerce

Business environment has changed profoundly since when, in 1953, Neil Borden for the first time introduced the term “marketing mix” in his speech at the American Marketing Association, and since (McCharty, 1964) defined the 4 Ps (product, price, place and promotion) marketing mix as a combination of all the factors which managers may leverage to satisfy market needs. However, after almost fifty years, (Dominici, 2009) conclude that, despite the controversies between many studies, the basic construction of 4 Ps is still valid and, with some extension and adjustment for On-line companies is still the core of operative decisions.

After product, pricing plays a key role in the marketing mix. The reason for this importance is that where the rest of the elements of the marketing mix are cost generators, price is a source of income and profits. Through pricing, online companies manage to support the cost of maintenance, the cost of distribution, and the cost of promotion.

According to economic theory, market transactions involve search, price discovery and settlement (Lee and Clark, 1996). Before dot-com boom price discovery was the process of determining, via mechanisms such as auctions or negotiation, the prices at which demand and supply “clear” (Bakos, 1998). However, nowadays price discovery is done through online marketplaces and shop bots. Buyers are expected to continue searching for products that meet their needs at the lowest prices, until the costs of additional search exceed the potential gains (Rothschild, 1974). That is probably the reason why according to Internet Retailer, price-monitoring technology vendor Ugam recorded 9,715 price changes of electronics, toys and household goods on Amazon.com during the holiday season from 24 November to 14 December (Rueter, 2014). Amazon has established a record, surpassing the frequency- volatility of the prices of its competitors - such retail giants as Best Buy Co., Target Corp., Wal-Mart Stores Inc. and Toys “R” Us Inc. According to the vice- president of marketing, Amazon is able to change the price of the product up to 10 times per day. Approximately the price of 20% of all online products is changing daily, and the price of the most running products is updated every few minutes, according to the vice-president of product and business development strategy in the price monitoring company Decide.com, which recently has been acquired by eBay Inc. Thus, the price in e-commerce is highly dynamic and depends on market conditions and the pricing strategies can have an individual character for each user, and it is possible due to BigData technologies (Pogorelova et al., 2016).

4. Methodology

With the final target of increasing the company profits it has been designed and developed an application to track the prices of competitors for a set of products.

This information will be used to set the product prices in order to offer the products both competitively and profitably in each Marketplace. This application must check, by taking into account information such as the product cost or the minimum product margin, if it is possible to decrease the price in order to reach the top cheapest position and as a consequence, increase sales.

Figure 1 shows the basic structure and the necessary architecture for the application developed in this work.

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 JSOM[3] file and imported to a local database. An automatic process is executed once a day to keep this

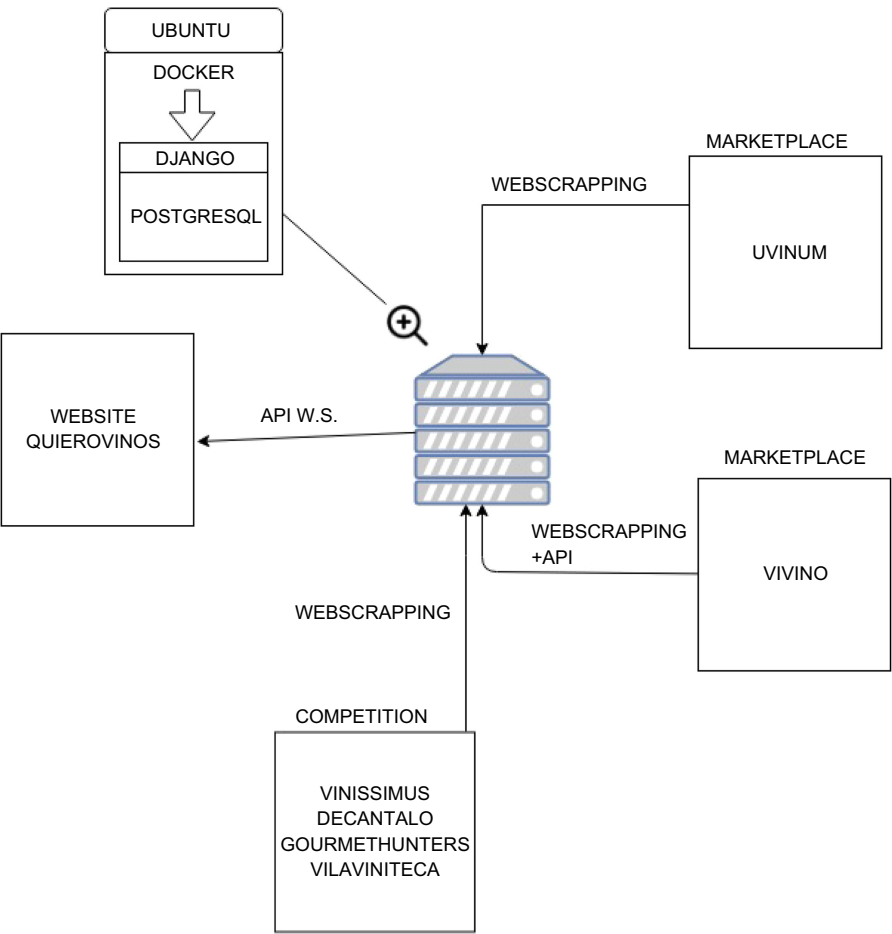


Figure 1.
System architecture

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 Scrapping* 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:

- (1) The *MinimumMargin* is the minimum allowed margin that has been defined for each product and from which the application will decide if its possible to decrease the price. This margin has been calculated by the company sales manager and it is the minimum expected ratio that makes the sale of each product profitable.
- (2) The *PurchasePrice* is the purchase price of the product provided by the company.

The following formula shows the calculation preformed by the application for each product in order to know if the price can be modified. The *MinimumPriceSale* is the price that we should apply in order to be the cheapest seller on the corresponding Marketplace, reaching the top cheapest position for that product. Therefore, if the result of the formula is greater than *MinimumMargin* it will mean that there is enough margin to apply the new price on that product. Other-wise, the *MinimumMargin* will have to be respected and will limit the price reduction.

$$((\text{MinimumPriceSale} - \text{PurchasePrice}) / \text{MinimumPriceSale}) * 100 > \text{MinimumMargin}$$

4.1 Operational details

In this section, the most important algorithms are introduced in order to explain how the application manages to obtain the list of products to analyze and the selling details from these products according to the competitors on each Marketplace.

First, in order to get the list of the products sold by the company on each Marketplace, a tool called *Selenium* ([Seleniumhq, 2019](#)) has been used to access each Marketplace *BackOffice* [5]. With this tool, one can simulate a command line navigation through a hidden browser, performing similar actions to what a real user would do. In this way, together with the company's credentials for each Marketplace, a computer script can get the entire list of products published on that Marketplace together with extra useful information such as the internal identifier, public buying links, etc. . . and save it on the server's filesystem. This task is performed by [Algorithm 1](#). For example, in line 2 the URL that links to private Marketplace Backoffice is defined. Then, from line 3 to 11, we can see how the Algorithm goes through the entire list of products and save in a *Json* file both the public *Uvinum* buying link and the internal reference of the product.

Algorithm 1. Get Links Uvinum

- (1) **Parse(self):**
- (2) $URL \leftarrow \text{https://manager.vcst.net/affiliates/products-catalog-list}$
- (3) **while** $currPage \leq lastPage$ **do**
- (4) $Uvinum\ Links \leftarrow getAllProductLinks$
- (5) **for** link in $uvinum_links$ **do**
- (6) $linkjuvinum \leftarrow link.find(productj\ url)$
- (7) $reference \leftarrow link.find(reference)$
- (8) $links.append(linkjuvinum, reference)$
- (9) **end for**
- (10) $currPage \leftarrow currPage + 1$
- (11) **end while**
- (12) **return** $links$

[Algorithm 2](#) has two main functions. First, setting up the location from where the shipment costs are calculated. This feature has been implemented because Uvinum Marketplace calculates the shipping costs depending on the country from which you access the website. Secondly, getting the product list with their corresponding links from the generated file by the previous Algorithm in order to process this information and proceed to the next Algorithm (lines 3 to 5).

Algorithm 2. Get Sales Uvinum

- (1) **getSales(self,links):**
- (2) *changeLocation()*
- (3) **for** link in links **do**
- (4) *getWineShops(link, ref)*
- (5) **end for**
- (6) **return** json

[Algorithm 3](#) is in charge of accessing the product url to obtain the best selling price, the name of the store that sells it at that price and its cost shipment (lines 2–6). On line 7, the obtained data is stored in a file *JSON* format. Moreover, (lines 8–15) it gets the same information (price, seller name and shipping cost) for other possible sellers of the same product in order to have a complete list of competitors.

Algorithm 3. Get WineShops

- (1) **getWineShops(self,web,ref):**
- (2) *web*←*getWeb(www.uvinum.es/product/price)*
- (3) *best*←*web.find(price)*
- (4) *price*←*best.find(price)*
- (5) *name*←*best.find(name)*
- (6) *shipping*←*web.find(shipping)*
- (7) *addWineShopAndSale(name, price, ref, shipping)*
- (8) **if** *web.find(othersellers)* **then**
- (9) *wineshops*←*web.findAll(wineshop)*
- (10) **for** wineshop in wineshops **do**
- (11) *price*←*wineshop.find(price)*
- (12) *name*←*wineshop.find(name)*
- (13) *shipping*←*wineshop.find(shipping)*
- (14) *addWineShopAndSale(name, price, ref, shipping)*
- (15) **end for**
- (16) **end if**

Finally, [Algorithm 4](#) collects all data from all competitors selling a given product and persists that information into a file.

Algorithm 4. Add WineShop And Sale

- (1) **addWineShopAndSale(self, name, price, ref, shipping):**
- (2) **if** *!wineshop.find(wineshops_file)* **then**
- (3) *wineshops_file* ← *wineshop.save(name, shipping)*
- (4) **end if**
- (5) *sales_file* ← *sale.save(ref, name, price, shipping)*

Web scraping
for online wine
sales

3391

4.2 Front-end

In this section, the front-end side of the application is introduced, detailing how the data is presented to the end user and explaining the main operations that can be performed.

First, the most important visualized data for each analyzed product are:

- (1) Reference: Internal product reference.
- (2) Name: Product name.
- (3) Winery: The product winery.
- (4) Margin (%): Current margin of that product for the company. [Marketplace]
- (5) Margin (€): Amount in euros that the product price should be decreased in order to be the cheapest in each Marketplace.
- (6) PVPC (€): Most competitive product price. It is the cheaper product price of each Marketplace minus X cents. This X value is calculated by means of simulation (see [section 5](#)). The default value is $X=1$.
- (7) MU (%): Margin of the product if we decrease the product price to be the cheapest in each Marketplace.
- (8) Can Reduce?: shows three values: YES, NO, CONSIDER. In the current case study, the minimum margin has been set to 30% since it has been defined by company as the margin from which the product is profitable. If by decreasing the product price the resulting margin is still more than 30%, the column will display *YES* in a green color, along with a button that will automatically update the new price of the product in the corresponding Marketplace and in the own company's online shop. Other-wise, if the resulting margin is between 25 and 30% the application will display the text *CONSIDER* meaning that company can take into account this product and try to negotiate a better purchase price with its supplier to sell the product in a competitive way. Finally, for a resulting margin lower than 25% a red *NO* will be displayed indicating that there is no chance for that product to be sold as the cheapest among the competitors.

[Figure 2](#) is an example of how this data is presented to the final user of the application. At the top of the screenshot it can be seen a summary of the results for each Marketplace. The number in green represents the amount of products the price of which can be decreased in order to be the cheapest seller and respecting the minimum margin defined by the company. The number in yellow represents the amount of products that can be taken into consideration since they are close to the desired margin. The number in red means the amount of products where the company is not competitive, that is, the margin that should be applied to be the cheapest seller is below 25%. And the number in blue represents the amount of products that are already sold at the cheapest price for each Marketplace.

BFJ
122,11

3392

Figure 2.
Product list

0.4 €

30.0 %

winnum

22

133

542

1521

82

161

412

1317

4 vino

Products

WineShops

MarketPlaces

Sales

Show 25 rows

Excel

Column visibility

PDF

Search:

Reference %	Name	Category %	Winery	% Margin %	Uvium margin %	pypv %	%MU %	Can reduce? %	Vino margin %	pypv %	%MV %	Can reduce? %
AN0131	Rigol Brut Nature	Cava	Rigol	60.47%	-0.45	4.5	44.30%	Can reduce? %	-1.16	3.79	18.79%	No
AN0221	Mionetto Prosecco	Otros Espumosos	Mionetto	41.35%	-0.27	9.72	41.35%	Can reduce? %	-2.68	7.04	0.34%	No
CB0016	Bertha Max Gran Reserva 2007	Cava	Bertha	46.53%	-2.40	29.6	35.18%	Can reduce? %	-2.40	29.6	35.18%	Can reduce? %
CB0576	Glorioso Crianza Magnum 2014	Vino Tinto	Bodegas Palacio	32.73%	-0.02	12.22	32.50%	Can reduce? %	-0.02	12.22	32.50%	Can reduce? %
CB0636	Amaren Blanco Fermentado 2016	Vino Blanco	Bodegas Amaren	30.28%	-0.02	12.87	30.07%	Can reduce? %	+0.01	12.9	30.38%	No
CB0637	Petit Saù Blanc 2017	Vino Blanco	Mas Blanchi Jové	67.71%	-0.01	9.2	67.52%	Can reduce? %	-0.07	9.14	66.37%	Can reduce? %
CB0703	Jaspi Negre 2016	Vino Tinto	Coca I Filó	61.59%	-1.41	7.99	36.04%	Can reduce? %	-1.41	7.99	36.04%	Can reduce? %
CB0707	Tocat de L'Ala Blanc 2017	Vino Blanco	Celler Rog Paradís	48.23%	-0.02	9.88	47.91%	Can reduce? %	-0.01	9.89	48.07%	Can reduce? %

Figure 3 shows for each product the list of all its sellers along with its price, the shipping cost, the available stock and the number of bottles sold.

4.3 Back-end

The web application has been installed on a Virtual Machine (VM) in public Cloud provider using a Software as Service (SaaS) system. Table I shows the main hardware and software features of the VM. As we can see, the computational requirements for running the application are very low. Even though, in Section 5 we will see how this configuration for the VM is enough to run the application smoothly.

For this project, the VM has been installed over a Docker (Vaughan-Nichols, 2019) container. This technology has many advantages such as the portability in terms of the application migration to another data-center or regarding the self-sufficiency since the Docker contains the necessary libraries, files and configuration files to deploy and run the application. Summarizing, this technology has a better performance and optimizes the computational resources better than other virtualization tools, helps developers to quickly create ready-made container applications and makes the management and deployment of applications easy.

From the software development point of view, *Python 3* has been chosen as a programming language since it is one of the most used when it comes to making web applications with *web scraping* requirements. Also, the *Django* framework has been chosen because it provides the developer with a wide set of libraries and tools suited to perform many different tasks that make the development process easier and quicker.

Finally, *PostgreSQL* has been chosen for storing the data since it is one of the most common database engine.

5. Results

In this section we will discuss how the use of this application increased the company's profits from different perspectives. In particular, we will analyze the evolution of the company's incomes, the amount of bottles sold and the number of shipments made, from one month before the application was released to production (January 2019) and compared to the same period of the previous year.

On the other hand, the scalability of the application is also analyzed, comparing the global application's performance by splitting the workload into several virtual machines.

5.1 Revenues

According to the data provided by Quiervinos.com, Figures 4 and 5 show the total incomes evolution of the company from January 2019 to May 2019 compared to the same period of the previous year 2018. Taking into account that the application was deployed on February 2019, it can be easily appreciated that it has an important positive effect over the company's revenues.

For example, on May 2019, one month after the application deployment, the incomes increased by 48% on Uvinum Marketplace and by 169% on Vivino compared to the same month of the previous year. In April the incomes raised 69% and 130% respectively and on May the increment went up to 170% on Uvinum and 57% on Vivino.

If we take a look to the number of Bottles sold, the behaviour is more or less the same. Figures 6 and 7 show a similar evolution to the incomes for the same analyzed period and compared also with the previous year. That confirms the good performance of the application and a huge positive impact over the company's revenues.

BFJ
122,11

3394

Figure 3.
Sales list

Show 25 rows

Excel

PDF

Column visibility

Products

WineShops

MarketPlaces

Sales

Add new sale

Search: ardanza

ID	WineShop	MarketPlace	Price	Shipping	Total	Stock	Bottles Sold
[JA0200] Viña Ardanza Reserva 2009							
185008	Quiervinos	Vivino	17.94	4.95	22.89	87	156
185009	Tu tienda del Vino	Vivino	17.95	5.99	23.94	723	46
185010	Decantare	Vivino	18.75	5.99	24.74	24	3
185011	Campoluz Enoteca	Vivino	19.45	5.99	25.44	50	0
185012	Bodeboca	Vivino	19.67	5.99	25.66	210	3
185013	GrauOnline	Vivino	19.89	5.99	25.88	1533	1
185014	TomeVinos	Vivino	19.9	5.99	25.89	730	2
185015	Vino Espana	Vivino	20.98	5.99	26.97	1559	2
185016	WineisSocial	Vivino	21.5	5.99	27.49	23	0
185017	El Sumillier	Vivino	22.9	5.99	28.89	0	0

5.2 Shipments

The shipping costs are an important aspect of the selling process and one of the main reasons of shopping cart abandonment (abandonment rate). Also, the shipping cost for a company are usually related to the monthly volume of shipments, being more competitive when the number of shipments increase. Thus, Figures 8 and 9 show the great improvement in the following months after the deployment of the application when it comes to the performed shipments. In the last analyzed month, for example, the raise reached more than 300% on Uvinum and more than 200% on Vivino compared to the same month of the previous year. That important increase allows the company to negotiate better rates with its shipping providers.

Web scraping
for online wine
sales

3395

SO	Ubuntu 16.04x64
Core	Intel Xeon Virtual CPU 0.5 VCores
RAM	1GB DDR3
Disk	5GB SSD

Table I.
Server configuration

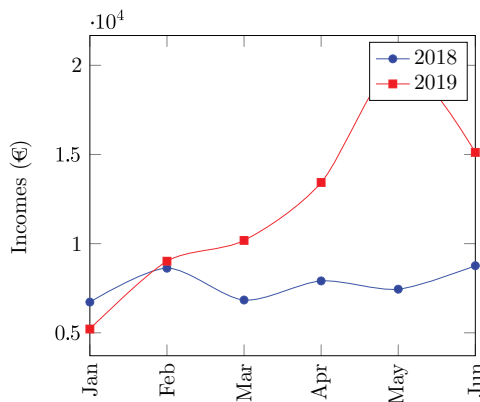


Figure 4.
Incomes from Uvinum

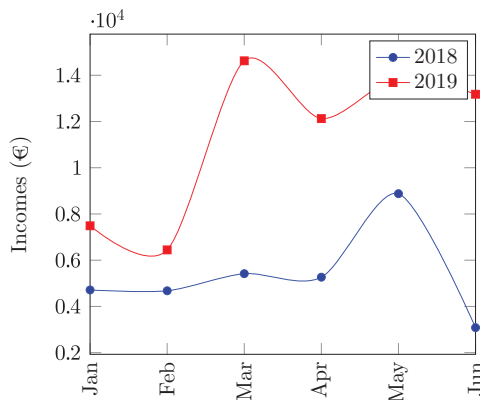


Figure 5.
Incomes from Vivino

Figure 6.
Bottles sold on Uvinum

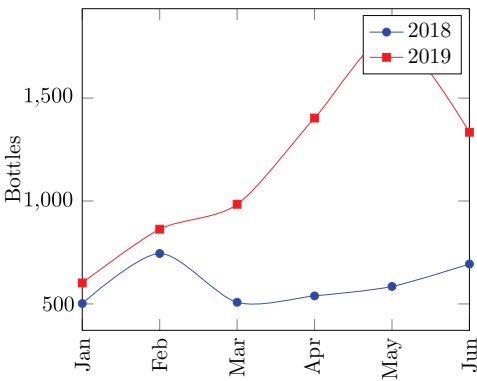


Figure 7.
Bottles sold on Vivino

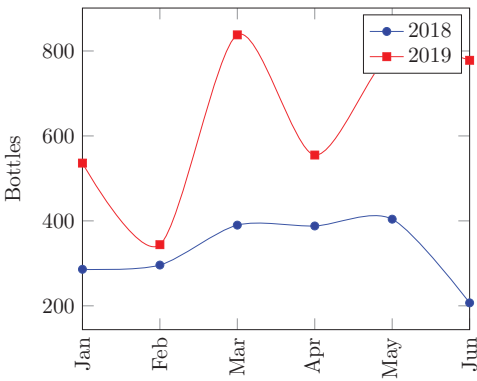
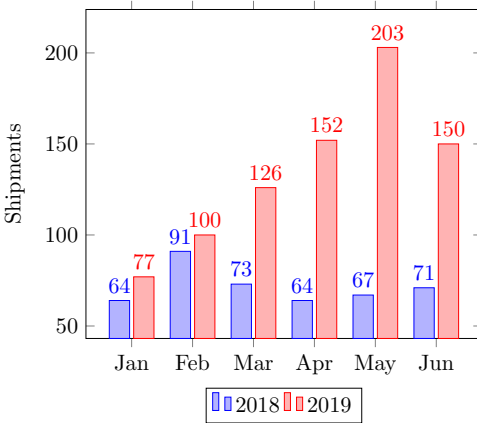


Figure 8.
Uvinum shipments



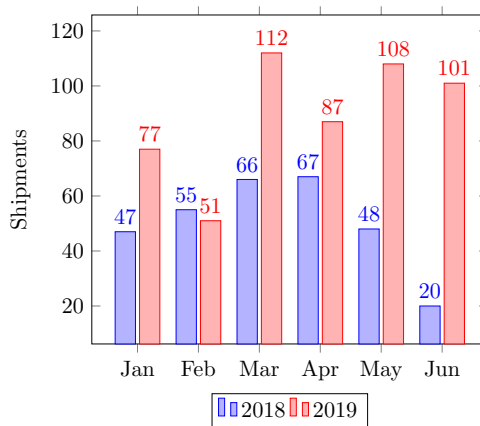


Figure 9.
Vivino shipments

5.3 Scalability

In order to study the scalability of the application, 16 identical Virtual Machine (VM) have been created in a cloud environment. All of them with the same resource configuration has the one used for production and detailed in [Section 4.3](#).

Also, the application has been modified in order to split the workload between the different instances running on each VM. The basic idea consists in splitting the file containing the list of products and the corresponding buying links into several files. Once we have these files, each one of them with a subset of products, they will be processed separately for the same application but running in parallel on different servers. To do so, a main server node has been defined in order to manage the task of partition the original file with the list of products, to send via *SFTP* protocol the files with the subset of products to each node and to collect the result *JSON* with the results of the task performed by each VM. Finally, the main server will be also in charge of putting together all the results and save the final data to the database.

[Figure 10](#) shows the execution time of the same execution but with a different number of VM. It can be seen that the running time decreases almost proportional as the number of servers increases. That is, the execution time with 16 servers, will be approximately 16 times less than the execution time in a single server. Despite this, the reduction is not exactly

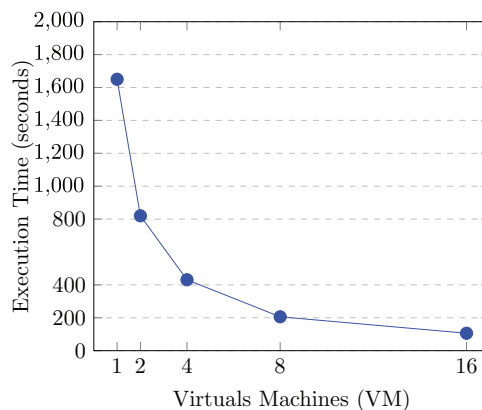


Figure 10.
Total execution time

proportional due to the time needed to split the main file plus the communication time required to send and collect the data between the main server and the others.

6. Discussion

As we have seen in [Section 5](#) the application improved in a notorious way the company's results in terms of sales and shipping costs. It must be said that without the use of the presented application, performing the price comparison process within each one of the marketplaces would have taken a long time. Moreover, as prices change very frequently, the obtained information has a very limited time value, and the competitors prices should be analyzed daily in order to take accurate decisions regarding the company's price policy.

By using our application, this process takes place in less than an hour with 1VM and less than 2 min with 16VM. Once all data is collected, the company manager can have a global and a very intuitive vision of the results. That allows him to quickly perform price changes without having to worry about margins, minimum product prices, etc. Also, using the application to define the price policy, avoid possible human errors.

Finally, we would like to comment why the results obtained in *Uvinum* are better than the ones obtained in *Vivino*. This could be attributed to the fact that, during the first quarter of the analyzed year, *Uvinum*, made several aggressive campaigns of free shipping for a certain amount spent in the marketplace.

7. Conclusions and future work

From the deployment of the application, sales through the marketplaces have increased 104% in average. Although it only has been under production for a few months, the obtained results allows us to be very optimistic and we are looking forward to see the evolution of the upcoming months.

Also, we would like to point out that this application could be used by other specialized online stores which use *Uvinum*, and *Vivino* marketplaces to sell their products. Although the application has been designed for the wine sector and the two named marketplace, it could be exported to other sectors. For that, it should be implemented new modules to collect information regarding the competitor's price of the products selling on each corresponding marketplace.

One of the main faced problems during the development has been the limitations and the high levels of protection of both marketplaces. For example, if too many requests where made in a short period of time, their security systems ended up restricting our requests for a specified period of time. To avoid this, we had to slow down our application so they do not detect massive data reading. Even that was a very effective, easy to implement, and quick solution, it has a negative effect over the execution time required to perform the data collection process. For that reason, other alternatives such as making the requests through *proxy* servers will be also studied in order to skip that protection. For the moment, this solution has not been implemented since the required time to perform the data collection is not a crucial factor for us.

Notes

1. API: an application programming interface is set of subroutine definitions, communication protocols, and tools for communication among various software components.
2. HTTP: internet communication protocols.
3. JSON: open-standard file format that uses human-readable text to transmit data objects consisting of attribute-value pairs and array data types.

4. Django framework: computer set of tools to ease create complex, database-driven websites.
5. Backoffice: private space of the marketplaces where the company administrator can manage the list of products published on it as well as the received buying orders. To access, each company have a unique and personal user and password.

Web scraping
for online wine
sales

References

3399

- Alba, J.W., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R.S. and Wood, S. (1997), "Interactive home shopping: consumer, retailer, and manufacturer incentives to participate in electronic marketplaces", *Journal of Marketing*, Vol. 61, pp. 38-53.
- Anckar, B. (2003), "Drivers and inhibitors to e-commerce adoption: exploring the rationality of consumer behavior in the electronic marketplace", Vol. 1, pp. 41-58.
- Bailey, J. and Bakos, Y. (1997), "An exploratory study of the emerging role of electronic intermediaries", *International Journal of Electronic Commerce*, Vol. 1, pp. 7-20.
- Bakos, J.Y. (1998), "The emerging role of electronic marketplaces on the internet", *Communications of the ACM*, Vol. 41 No. 8, pp. 35-42, doi: [10.1145/280324.280330](https://doi.org/10.1145/280324.280330).
- Bhatt, G.D. and Emdad, A.F. (2001), "An analysis of the virtual value chain in electronic commerce", *Logistics Information Management*, Vol. 14 Nos 1/2, pp. 78-85, doi: [10.1108/0957126050110362465](https://doi.org/10.1108/0957126050110362465).
- Bruwer, D.J. and Li, E. (2007), "Wine-related lifestyle (wrl) market segmentation: demographic and behavioural factors", *Journal of Wine Research*, Vol. 18 No. 1, pp. 19-34, doi: [10.1080/09571260701526865](https://doi.org/10.1080/09571260701526865).
- Castrillo-Fernández, O. (2015), "Web scraping: applications and tools", European-dataportal.eu available at: https://www.europeandataportal.eu/sites/default/files/2015_web_scraping_applications_and_tools.pdf.
- Cox, D. (2009), "Predicting consumption, wine involvement and perceived quality of Australian red wine", *Journal of Wine Research*, Vol. 20 No. 3, pp. 209-229, doi: [10.1080/09571260903450963](https://doi.org/10.1080/09571260903450963).
- DiRusso, D.J., Mudambi, S.M. and Schuff, D. (2011), "Determinants of prices in an online marketplace", *Journal of Product and Brand Management*, Vol. 20 No. 5, pp. 420-428, doi: [10.1108/10610421111157946](https://doi.org/10.1108/10610421111157946).
- Dominici, G. (2009), "From marketing mix to e-marketing mix: a literature overview and classification", *Journal of the Academy of Marketing Science*, Vol. 4 No. 9, pp. 17-24. available at: <https://ssrn.com/abstract=1961974>.
- Dressler, M. (2018), "The German wine market: a comprehensive strategic and economic analysis", *Beverages*, Vol. 4, p. 92.
- Fehér, J., Lengyel, G. and Lugasi, A. (2007), "The cultural history of wine - theoretical background to wine therapy", *Central European Journal of Medicine*, Vol. 2 No. 4, pp. 379-391.
- Fenech, T. and O'Cass, A. (2001), "Internet users' adoption of Web retailing: user and product dimensions", *Journal of Product and Brand Management*, Vol. 10 No. 6, pp. 361-381, doi: [10.1108/EUM000000000006207](https://doi.org/10.1108/EUM000000000006207).
- Galtés, M. (2019), "Pernod ricard compra uinum para crecer en comercio online", La Vanguardia, available at: <https://www.lavanguardia.com/economia/emprendedores/20180804/451201682513/pernod-ricard-uvinum.html>.
- Gensler, S., Peter, V. and Böhm, M. (2012), "Understanding consumers' multichannel choices across the different stages of the buying process", *Marketing Letters*, Vol. 23 No. 4, pp. 987-1003.
- Gergaud, O. and Livat, F. (2007), "How do consumers use signals to assess quality?", Working Paper No. 3, American Association of Wine Economists, April, pp. 1-22.
- Grandon, E.E. and Michael Pearson, J. (2004), "Electronic commerce adoption: an empirical study of small and medium US businesses", *Information and Management*, Vol. 42 No.1, pp. 197-216, available at: <http://www.sciencedirect.com/science/article/pii/S0378720604000205>.

- Hall, J., Binney, W. and Barry O'Mahony, G. (2004), "Age related motivational segmentation of wine consumption in a hospitality setting", *International Journal of Wine Marketing*, Vol. 16 No. 3, pp. 29-43, doi: [10.1108/eb008777](https://doi.org/10.1108/eb008777).
- Higgins, L. and Llanos, E. (2015), "A healthy indulgence? Wine consumers and the health benefits of wine", *Wine Economics and Policy*, Vol. 4 No. 1, pp. 3-11.
- Hoffman, D., Novak, T. and Peralta, M.A. (1999), "Building consumer trust online", *Communication of the ACM*, Vol. 42, pp. 80-85.
- Hussain, M., Cholette, S. and Castaldi, R. (2007), "Determinants of wine consumption of US consumers: an econometric analysis", *International Journal of Wine Business Research*, Vol. 19, pp. 49-62.
- Jarvenpaa, S.L. and Todd, P.A. (1996), "Consumer reactions to electronic shopping on the world wide web", *International Journal of Electronic Commerce*, Vol. 1 No. 2, pp. 59-88. doi: [10.1080/10864415.1996.11518283](https://doi.org/10.1080/10864415.1996.11518283).
- Jiang, P. (2002), "A model of price search behavior in electronic marketplace", *Internet Research*, Vol. 12 No. 2, pp. 181-190, doi: [10.1108/10662240210422530](https://doi.org/10.1108/10662240210422530).
- Kangis, P. and Rankin, K. (1996), "Interactive services: how to identify and target the new markets", *Journal of Marketing Practice: Applied Marketing Science*, Vol. 2 No. 3, pp. 44-67, doi: [10.1108/EUM0000000004298](https://doi.org/10.1108/EUM0000000004298).
- Kannan, P.K. and Kopalle, P.K. (2001), "Dynamic pricing on the internet: importance and implications for consumer behavior", *International Journal of Electronic Commerce*, Vol. 5 No. 3, pp. 63-83, doi: [10.1080/10864415.2001.11044211](https://doi.org/10.1080/10864415.2001.11044211).
- Krause, J. (1998), *Electronic Commerce: Geschäftsfelder der Zukunft Heute Nutzen*, Carl Hanser Verlag, München.
- Lee, H.G. and Clark, T.H. (1996), "Impacts of the electronic marketplace on transaction cost and market structure", *International Journal of Electronic Commerce*, Vol. 1 No. 1, pp. 127-149, available at: <http://www.jstor.org/stable/27750803>.
- Levine, M. and Pownall, S. (2004), "Wine industry outlook: consumption demographics. Market segmentation, trends and opportunities", *Wine Industry Outlook Conference*, Darling Harbour, Sydney, Australia.
- Li, H., Wang, H., Li, H., Goodman, S., van der Lee, P., Xu, Z., Fortunato, A. and Yang, P. (2018), "The worlds of wine: old, new and ancient", *Wine Economics and Policy*, Vol. 7 No. 2, pp. 178-182, available at: <http://www.sciencedirect.com/science/article/pii/S2212977418300619>.
- Lindsey-Mullikin, J. and Grewal, D. (2006), "Imperfect information: the persistence of price dispersion on the web", *Journal of the Academy of Marketing Science*, Vol. 34 No. 2, p. 236, doi: [10.1177/0092070305283366](https://doi.org/10.1177/0092070305283366).
- Llewellyn, J. (2019), "Wine app vivino receives \$10.3 million in follow-on series a funding", *Prnewswire.com* available at: <https://www.prnewswire.com/news-releases/wine-app-vivino-receives-103-million-in-follow-on-series-a-funding-215816691.html>.
- Lockshin, L. and Corsi, A.M. (2012), "Consumer behaviour for wine 2.0: a review since 2003 and future directions", *Wine Economics and Policy*, Vol. 1 No. 1, pp. 2-23, available at: <http://www.sciencedirect.com/science/article/pii/S2212977412000075>.
- Lockshin, S., Lawrence and Rhodus, W.T. (1993), "The effect of price and oak flavor on perceived wine quality", *International Journal of Wine Marketing*, Vol. 5, pp. 13-25.
- McCharty, E.J. and Irwin, R.D. (1964), *Basic Marketing: A Managerial Approach*, 2nd ed.
- Milev, P. (2017), "Conceptual approach for development of web scraping application for tracking information", *Economic Alternatives* No. 3, pp. 475-485, available at: <https://ideas.repec.org/a/nwe/eajour/y2017i3p475-485.html>.
- Pitt, L. (2017), "Ten reasons why wine is a magical marketing product", *Journal of Wine Research*, Vol. 28 No. 4, pp. 255-258, doi: [10.1080/09571264.2017.1395386](https://doi.org/10.1080/09571264.2017.1395386).

- Pogorelova, E., Yakhneeva, I., Anna, A. and Prokubovskaya (2016), "Marketing mix for e-commerce", *International Journal Of Environmental and Science Education*, Vol. 11 No. 14, pp. 6744-6759, available at: <https://ssrn.com/abstract=2848818>.
- Quester, P.G. and Smart, J. (1996), "Product involvement in consumer wine purchases: its demographic determinants and influence on choice attributes", *International Journal of Wine Marketing*, Vol. 8 No. 3, pp. 37-56, doi: [10.1108/eb008661](https://doi.org/10.1108/eb008661).
- Quinton, S. and Harridge-March, S. (2008), "Trust and online wine purchasing: insights into UK consumer behaviour", *International Journal of Wine Business Research*, Vol. 20, pp. 68-85.
- Rao, V.R. (1984), "Pricing research in marketing: the state of the art", *The Journal of Business*, Vol. 57 No. 1, pp. S39-S60, available at: www.jstor.org/stable/2352922.
- Ritala, P., Golnam, A. and Wegmann, A. (2014), "Coopetition-based business models: the case of Amazon.com", *Industrial Marketing Management*, Vol. 43 No. 2, pp. 236-249. Special Issue on Co-opetition Cooperation and Competition, available at: <http://www.sciencedirect.com/science/article/pii/S0019850113002150>.
- Rose, S.N., Clark, M., Samouel, P. and Hair, N. (2012), "Online customer experience in e-retailing: an empirical model of antecedents and outcomes", *Journal of Retailing*, Vol. 88 No. 2, pp. 308-322, available at: <http://www.sciencedirect.com/science/article/pii/S0022435912000243>.
- Rothschild, M. (1974), "A two-armed bandit theory of market pricing", *Journal of Economic Theory*, Vol. 9 No. 2, pp. 185-202.
- Rueter, T. (2014), "The price is right - then it's not", available at: <https://www.internetretailer.com/2014/08/04/price-right-then-its-not>.
- Seleniumhq (2019), "Introduction — Selenium documentation", Seleniumhq.org [online] available at: https://www.seleniumhq.org/docs/01_introducing_selenium.jsp.
- Smith, M.D. (2002), "The impact of shopbots on electronic markets", *Journal of the Academy of Marketing Science*, Vol. 30 No. 4, pp. 446, doi: [10.1177/009207002236916](https://doi.org/10.1177/009207002236916).
- Steenkamp, J.B. (1988), "The relationship between price and quality in the marketplace", *De Economist*, Vol. 136, pp. 491-507.
- Swaminathan, V., Lepkowska-White, E. and Bharat Rao, R. (1999), "Browsers or buyers in cyberspace? An investigation of factors influencing electronic exchange", *Journal of Computer-Mediated Communication*, Vol. 5.
- Szolnoki, G., Thach, L. and Kolb, D (2016), *Current Status of Global Wine Ecommerce and Social Media*, Palgrave Macmillan US, New York, NY, pp. 1-12, doi: [10.1057/9781137602985_1](https://doi.org/10.1057/9781137602985_1).
- Udo, G. (2001), "Privacy and security concerns as major barriers for e-commerce: a survey study", *Information Management and Computer Security*, Vol. 9, pp. 165-174.
- Vaughan-Nichols, S. (2019), *What Is Docker and Why Is it So Darn Popular?*, ZDNet available at: <https://www.zdnet.com/article/what-is-docker-and-why-is-it-so-darn-popular/>.
- Verma, V., Sharma, D. and Sheth, J. (2015), "Does relationship marketing matter in online retailing? A meta-analytic approach", *Journal of the Academy of Marketing Science*, Vol. 44, pp. 206-217.
- Veseth, M. (2013), *Extreme Wine: Searching the World for the Best, the Worst, the Outrageously Cheap, the Insanely Overpriced, and the Undiscovered*, Rowman and Littlefield Publishers, Lanham, MD.

Corresponding author

Josep Rius can be contacted at: josep.riustorrento@udl.cat

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com