Assessing the nexus of label content and consumer reviews: insights from premium wines on Vivino

Camila Rocuant Fuentes[1] [0009-0002-0584-685X],

Paulo Rita[1] [0000-0001-6050-9958],

Nuno António[1,2*] [0000-0002-4801-2487] nantonio@novaims.unl.pt

[1] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal

[2] CITUR – Centro de Investigação, Desenvolvimento e Inovação em Turismo, Faro, Portugal

*Corresponding author

This is the accepted author manuscript of the following article published by EMERALD in *Internet Research:*

Rocuant Fuentes, C. C., Rita, P., & António, N. (2025). Assessing the Nexus of Label Content and Consumer Reviews: Insights from Portuguese Premium Wines on Vivino. *International Journal of Wine Business Research*. https://doi.org/10.1108/IJWBR-08-2024-0044

ASSESSING THE NEXUS OF LABEL CONTENT AND CONSUMER REVIEWS: INSIGHTS FROM PORTUGUESE PREMIUM WINES ON VIVINO

Abstract

Purpose: This study explores the relationship between wine label content and consumer reviews on Vivino, focusing on how label information compares to consumer perceptions and purchase decisions.

Methodology: The research employs text mining, image mining, and topic modeling techniques to analyze a dataset of 444 highly rated Portuguese wines and 993,945 consumer reviews from Vivino. It examines the overlap between topics discussed in consumer reviews and the information provided on wine labels.

Findings: The findings reveal that wine labels have a limited similarity with consumer reviews. While consumer reviews primarily emphasize sensory characteristics like fruitiness and tannins, wine labels focus more on descriptive and technical information.

Originality: This research provides new insights into the dynamic relationship between wine labeling and consumer behavior. By analyzing the alignment between traditional wine labels and online consumer reviews, the study offers valuable implications for wineries seeking to optimize label design to match consumer preferences, particularly sensory attributes better. This study extends existing research and highlights critical gaps in how wine labels convey essential sensory information to consumers.

Keywords

Consumer Reviews; Image Mining; Latent Dirichlet Allocation (LDA); Text Mining; Text Similarity; Wine Labels

1. Introduction

Wine has played an important role in connecting people and food throughout history and continues to do so today. Portugal exemplifies a country where wine is an integral part of daily life. It stands out for its high alcohol consumption, with 52.5 liters per capita (Instituto Nacional de Estatística Portugal, 2023; OECD, 2022). Portugal ranks as the fifth-largest wine exporter in Europe, with growth rates of 11.7% in recent years (Instituto Da Vinha e Do Vinho, 2021).

Portugal is renowned for its flourishing wine industry, particularly in the Douro and Alentejo regions. These regions are famous for producing exceptional wines and are integral to Portugal's wine heritage (Instituto Da Vinha e Do Vinho, 2023). They embody a long and prosperous history of winemaking traditions.

Selecting a suitable wine constitutes an intricate decision-making process for consumers, who often seek support from diverse informational sources, including wine guides, reviews, advertisements, other consumers, and wine labels (Sherman & Tuten, 2011).

Notably, labeling is crucial in the decision-making process associated with wine purchases. Label design and product information such as grape variety, brand name, and price influence consumer decisions (Sherman & Tuten, 2011). European shoppers place over 30% of importance on label design when buying wine, indicating that a well-designed label can convey a sense of quality and prestige, which is especially important for high-end wines (Wine Intelligence, 2019).

Wine packaging conveys quality, providing consumers with relevant information about the wine, encompassing grape variety, alcohol content, and aging processes (Mueller et al., 2010). As a result, attributes and information on the label are pivotal in the purchasing process. This importance is underscored by consumers' inability to discern quality attributes, color, and aroma without purchasing and tasting the wine (Barber & Almanza, 2007).

For consumers, wine labels are a relevant element in their purchase decisions (Madureira & Nunes, 2013). The region of origin stands out as the most influential factor, often perceived as a quality marker, particularly by less experienced buyers. Beyond this, label details play an important role, providing practical information such as food pairings for occasional shoppers and deeper historical context for more knowledgeable consumers, effectively guiding their

e-WOM: Electronic Word of Mouth
LDA: Latent Dirichlet Allocation
NLTK: Natural Language Toolkit
OCR: Optical Character Recognition
REST API: Representational State Transfer Application Programming Interface

choices. Research also indicates that back wine labels influence consumers' choices, suggesting that the information presented in that specific type of label plays a role in decision-making (Annunziata et al., 2015; Vecchio et al., 2018).

Alongside these traditional influences, the evolving landscape of consumer behavior in wine introduces electronic word-of-mouth (e-WOM) through social media reviews. Consumers actively share insights about wine products and brands through online platforms (Alzate et al., 2022). This form of e-WOM is increasingly influential, especially in the post-COVID-19 era, as 74% of consumers now consider online comments when making purchase decisions (Stackla, 2021). The shift in consumer behavior reflects engagement with wine brands on social media platforms (Szolnoki et al., 2014). These authors state that consumer interaction on social media impacts the brand's profile and sales, with high engagement in wineries. Interestingly, Gastaldello et al. (2021) discuss how COVID-19 impacted consumer behavior, leading to continued support for local wine producers, growth in virtual experiences, and a sustained increase in online shopping even after the pandemic.

In this context, online reviews of Portuguese wines, particularly Port wine, and discussions surrounding wine tourism in the Douro Valley have been analyzed on Twitter, focusing on consumers' experiences. The analysis indicated that most sentiments were positive, although Taylor Port wine garnered some negative feedback. Prominent topics in the conversations centered on the characteristics of vineyards, farms, and the region (Pérez-Rodríguez et al., 2022). Furthermore, the study advocates developing a more advanced methodology to extract valuable insights from social networks beyond Twitter.

Regarding the relationship between wine labels and consumer reviews, Jaud & Melnyk (2020) indicate that including images alongside text in labels can enhance consumer perceptions, such as liking, taste expectations, and purchase intentions, compared to labels with text alone. However, this research also points out the need for further investigation into factors like wine price and peer reviews, as these can influence how consumers respond to label designs. Previous research has highlighted the limitations regarding the consideration of the relationship between wine labels and reviews, as well as the analysis of specific price points in both reviews and labels (Jaud & Melnyk, 2020; Lam et al., 2019).

This study is motivated by the desire to explore the potential similarities between traditional wine labels and contemporary online reviews. While distinct, these two sources are significant

in the consumer's wine journey. Therefore, this study uncovers their potential interconnections without suggesting that they directly influence each other. Valuable insights from this exploration can benefit wineries by enhancing consumer engagement and helping scholars navigate the evolving landscape of consumer behavior in the wine industry.

To accomplish this research, the study utilizes Vivino as a data source. Vivino is the largest global wine marketplace and the most downloaded wine app (Vivino, 2022a). This online platform connects millions of wine enthusiasts, enabling them to explore, rate, and review wines. By utilizing crowdsourced information, it tailors personalized wine recommendations for users.

The study uses image and text mining techniques, including topic modeling and text similarity analysis. Using these techniques, the main topics users highlight in their reviews are identified, and key information from wine labels is extracted. The study uncovers patterns in extensive datasets. Moreover, by considering wine prices together with consumer reviews, this study intends to close gaps in existing research (Jaud & Melnyk, 2020; Lam et al., 2019; Pérez-Rodríguez et al., 2022), addressing the following research questions:

1. Which do users emphasize as primary topics in wine reviews?
2. What are the main themes found in the information presented on wine labels?
3. To what extent do key wine purchase drivers, such as varietal, region, and social media ratings, change the correspondence between information on wine labels and user reviews?

The roadmap for this research involves conducting a comprehensive literature review on wine labels and consumer perceptions, outlining the chosen methodology, analyzing user reviews and labeling data, discussing findings, and concluding with actionable insights for wineries and academia.

2. Literature review

Exploring the nexus between wine labels and social media reviews involves examining how labeling and online reviews shape consumer perceptions within the wine industry.

2.1 Packaging attributes and consumer perceptions

Packaging influences consumer decisions and is recognized as a pivotal instrument in marketing (Löfgren & Witell, 2005). Different packaging attributes were examined for their impact on consumer perception, revealing the significance of technical, ergonomic, and communicative attributes in shaping purchasing decisions. This underscores the need for innovative and appealing packaging that aligns with customers' needs.

Yeo et al. (2020) showed that product packaging influences purchase intentions in different ways, depending on product characteristics, consumer involvement, and purchase time. Kumar Agariya et al. (2012) categorized packaging attributes into visual and verbal domains, revealing that visual attributes affect low-involvement consumers with limited time, while verbal elements influence highly involved consumers without time constraints. Veflen et al. (2023) further expanded on this by linking sensory packaging cues, such as shapes and colors, to taste perceptions in food products.

In the wine industry, visual elements in packaging influence purchasing decisions, alongside the rising significance of information elements (Silayoi & Speece, 2004). In this industry, packaging, branding, and consumer experience shape the purchase decisions of consumers, particularly those less experienced (Mueller & Szolnoki, 2010).

Furthermore, the visual elements of wine bottles not only influence consumer perceptions and purchasing decisions but also reflect the brand's identity and storytelling in the competitive wine market (Vranken et al., 2024). In particular, Hallez et al. (2023) discovered that simple back-label designs can alleviate consumer apprehensions during the purchase process. This aligns with consumer preferences, as front-label cues, including origin and brand, significantly influence purchase decisions. Additionally, back-label cues, encompassing wine style and description, hold notable sway (Barber & Almanza, 2007).

Regarding features that shape consumer behavior in wine, extrinsic cues such as photos and brand names in labels are recognized for drawing attention and reinforcing brand recognition, leading to longer gaze durations (Chiang Lu, 2021). On the other hand, label color and wine style descriptions also influence perceptions. Black-and-white labels for red wines convey

uniqueness, while white wines allow for more flexible designs without significantly impacting consumer choices (Saket Joshi, 2022).

Delving deeper into luxury wines, the consumption of high-end tiers is driven by a complex interplay of product perception, pricing dynamics, information accessibility, consumer identity, and regional influences (Wright et al., 2023). These consumers are notably influenced by factors such as quality, value, trendiness, word of mouth, and especially customization in labeling (Barber, 2010; Leban et al., 2020).

Emerging data from Vivino suggests that consumer ratings on its platform now significantly influence pricing more than expert scores (Bazen et al., 2023). This shift indicates that consumer ratings may precede expert opinions in the wine market. Previous Vivino research also showed users tend to rate based on their vintage and region, exhibiting a preference for local wines (Kotonya et al., 2018). This underscores the multifaceted influence of packaging and recognition on consumer preferences and purchasing behaviors.

2.2 Online Reviews Shaping Consumer Preferences

Online reviews guide consumer purchases, with many relying on product reviews before finalizing a purchase (Calheiros et al., 2017; Latuny et al., 2021). These reviews influence consumer behavior, particularly for high-priced products. Chen et al. (2011) noted that consumers read more reviews for expensive products, and star ratings have a pronounced impact on expensive items within the same category.

Online reviews of wine provide valuable insights that assist consumers in making informed purchasing decisions. They may also offer greater utility than conventional label information, such as vintage and price (Yang et al., 2022).

Ferreira et al. (2023) argue that the descriptive language in wine reviews cultivates a shared lexicon within the wine community, facilitating nuanced discussions and enhancing comprehension among enthusiasts. Gunasekar et al. (2022) observe that knowledgeable wine reviewers emphasize sensory descriptors, including flavor and aroma, over general descriptive attributes. This contributes to a more comprehensive understanding of each wine's unique qualities. Notably, in the luxury tier, wines tend to be described with more elaborate language (Ramirez, 2010).

Rizo et al. (2022) studied consumers' emotional responses and expectations to like and purchase with and without Vivino reviews in wine. Reviews and ratings heavily influenced

these variables, with individual reviews having a higher impact than ratings (Moro & Rita, 2022).

2.3 Wine Regulations in the European Union

The regulatory landscape for wine labeling in the European Union has undergone significant changes from 2021 to 2023, impacting how information is conveyed to consumers. Before December 2023, wine labels were required to indicate the Protected Designation of Origin (PDO) or Protected Geographical Indication (PGI). For aromatized wine products, mandatory information included a nutrition declaration and an ingredient list, as outlined by Regulation (EU) No 1169/2011 (O.J., 2011), stipulating that ingredient lists could not be displayed alongside marketing or sales information. In the context of this study, Portugal has 30 PDOs and 10 PGIs registered (eAmbrosia, 2023).

As of the end of 2023, each wine bottle must feature an e-label, specifically a QR code linking to a dedicated webpage providing comprehensive details on ingredients, allergens, energy, and nutritional content (O.J., 2023). This webpage must not redirect the user to marketing or sales information and must include translations into all official EU languages. However, information regarding intolerances, allergies, and energy must still be printed on the physical wine label.

Considering the latest regulations described, Table 1 organizes the mandatory and non-mandatory information by label type.

The regulations introduce a structured approach to information presentation. The front label focuses on essential consumer details, such as the varietal, PDO/PGI designation, and alcoholic strength. In contrast, the back label incorporates more detailed information, including energy values and allergen declarations via e-label. Optional content remains a valuable tool for producers to engage consumers without breaching regulatory requirements. These new regulations seem to prompt the wine industry to focus more on back labels, as they concentrate most of the information.

However, it is important to note that the data used in this study were collected before these recent regulatory changes were implemented. Thus, the findings reflect the labeling landscape before the new requirements came into effect.

Table 1. Summary of wine labeling information according to new Regulation (EU) 2021/2117

Label type	Mandatory information	Non-mandatory information
Front label	• Varietal • PDO/PGI designation • Alcoholic strength by volume (%) • Origin • Net content (volume)	• Awards, promotional content
Back label	• Energy value (kcal) • Allergens (e-label) • List of ingredients (e-label) • Nutritional declaration (e-label)	• Winemaking technique details • Recommendations (e.g., food pairing, serving temperature) • Tasting notes
Non-specified	• Name of bottler, producer, vendor	

3. Methodology

To explore the relationship between the textual content of wine labels and corresponding reviews on Vivino, methods such as image mining, text mining, text preprocessing, topic modeling, and text similarity were employed to analyze consumer reviews and wine label information. The research design in Figure 1 is detailed in the following sections:

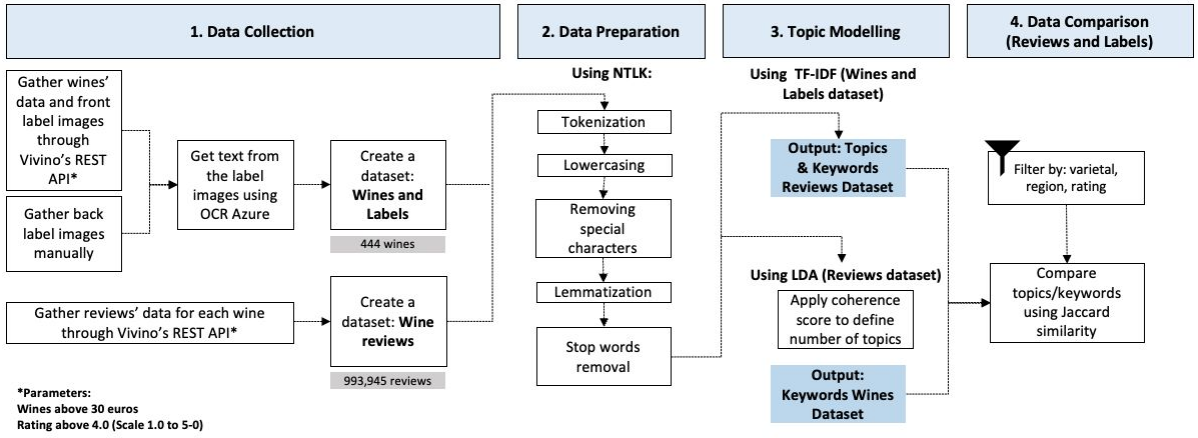

Figure 1. Research design

3.1 Data collection

This study collected and analyzed a dataset of Portuguese wines from Vivino, consisting of 444 wines and 993,945 consumer reviews, at the beginning of 2023.

Although the sample size and number of collected back-labels were limited, this manual process was constrained by the availability of labels online. The dataset was acquired through a Representational State Transfer Application Programming Interface (REST API) using a GET request, which facilitates communication through HTTP requests to execute database functions (IBM, 2021).

The extraction parameters specifically focus on wines rated 4.0 or above. This decision stems from the observation that consumers seek feedback on higher-priced products due to the greater monetary risk (Chen et al., 2011). Maintaining high quality and meeting consumer expectations is critical in the premium segment.

Vivino's (2020) rating system identifies wines with a 4.0 rating or higher as belonging to the top 15% of wines globally, with an average market price of 30 euros (Vivino, 2022b). As such, limiting the analysis to wines rated above 4.0 and priced over 30 euros ensures a focus

on premium offerings aligned with the study's scope within the high-end Portuguese wine market.

Wine attributes extracted from Vivino using the API included the wine name, price, average rating, user reviews, region, winery, varietal, body description, acidity, sweetness, tannin, vintage (year), and front label image URL. Different vintages were considered for comparison.

Table 2 summarizes the dataset, showing that 80.9% of the sample is red wine varieties and 19.1% is white wine.

Table 2. Summary statistics of the main variables in the dataset employed in the study. [2]

Variable	count	unique	top				frequency
wine_name	993,945	444	Cartuxa Évora Colheita Tinto				23,452
region	993,945	31	Douro				549,963
winery	993,945	243	Cartuxa				243,720
varietal	993,945	2	Red				977,942
year	957,566	71	"2011"				108,995
body_description	993,945	9	Very full-bodied				395,482
username	993,592	26,772	Miguel Silva				1,617
		mean	**std**	**min**	**25%**	**50%**	**75%**	**max**
price (euros)	993,945	137.3	112.2	30	49	85	180	480
reviews	993,945	18,612	19,774	28	6,617	11,195	18,583	72,323
acidity	985,329	3	0.1	2.3	2.9	3	3.1	4.5
sweetness	985,329	1.9	0.1	1	1.7	1.9	2	3
tannin	971,488	3.3	0.2	2.6	3.1	3.4	3.6	4
average_rating	993,945	4.4	0.5	1	4	4.5	5	5
user_reviews	993,910	722	1,529	0	31	190	796	20,514

[2] Top: Most frequent terms in the dataset.
Count: Number of distinct observations.
Acidity, Sweetness, Tannin: Evaluations made by reviewers on a scale from 1 to 5.
Review Rating: Ranges from 1 to 5.
The URL for the front label image is not included in Table 2 due to its extensive length; however, it was utilized in this study to extract the front label images.

Table 2 features 243 unique wineries across 31 regions, with Douro being the most popular. This dataset includes wines from both PDO and PGI categories, highlighting the diversity of the Portuguese landscape.

The study includes wines from 1942 to 2022, with 2011 receiving the highest number of reviews. Approximately one-third of the wines were classified as "very full-bodied."

The dataset comprises wines with an average price of 137.3 euros, and in terms of flavor, the wines exhibit an average acidity of 3.0, sweetness of 1.9, and tannin of 3.3. Regarding reviews, the average rating is 4.4, and 26,772 users have written reviews. These users are highly active, with an average of 722 reviews and 420 followers each, indicating they are more experienced enthusiasts or experts than typical consumers. Front label URLs were included; thus, this study was complemented by a manual collection of back labels.

Due to Vivino only having front labels available online, the authors manually collected 189 back-label image URLs corresponding to 42.6% of the selected wines, utilizing methods such as online searches and in-store visits. The texts from each front and back label image URL were extracted using the Optical Character Recognition (OCR) technique. OCR converts an image of text into a machine-readable format (Klostermann et al., 2018; Leotta et al., 2022). The conversion was done using the Azure Computer Vision OCR tool, as in previous research (Leotta et al., 2022), following a process observed in Figure 2:

Figure 2. Graphical representation of a case example where Azure Computer Vision OCR extracts text from a wine label image containing mixed languages and writing styles.

A subset of the extracted texts was manually verified to ensure accuracy by comparing the OCR outputs with the original label images. Twenty manual samples were run, and the accuracy rate was 100% in all cases.

Once the information was obtained, the resulting text from each label was merged with the rest of the dataset. Reviews were scrutinized using the same mechanism, and only Portuguese reviews were extracted to simplify the label comparison. Review attributes included text, rating, time, review likes, and user information, including user id, username, followers, and following. The dataset provides a comprehensive collection of Portuguese wines with corresponding label images and consumer reviews.

3.2 Topic modeling and data preparation

Topic modeling is employed to analyze consumer reviews and identify key themes. This unsupervised method is used to identify concealed topics in large texts (Li & Lei, 2021). Data preprocessing is required to apply topic modeling (Gurusamy & Kannan S, 2014; Zhang, 2019). For the reviews corpus, this process is accomplished using a Natural Language Toolkit (NLTK) (Bird et al., 2013), including i) Tokenization, ii) Lowercasing, iii) Removal of special characters, iv) Lemmatization and v) Stop word removal.

Once the data is cleaned, topic modeling is applied. Among the various topic modeling techniques, Latent Dirichlet Allocation (LDA) is chosen because it is the most widely used in marketing (Maier et al., 2018).
LDA is a stochastic generation model that generates topics from multiple documents (Blei et al., 2003), calculating the probability of different words being grouped within a specific topic. It assumes that topics are independent, while the words inside each topic are dependent and follow a Dirichlet distribution (Blei et al., 2003).

For this analysis, LDA is applied in Python using Gensim, an open-source library for natural language processing (Rehurek, 2009), and a coherence score is used to define the number of topics, suggested as the most accurate method by Röder et al. (2015). A higher score is assigned, indicating the interpretability and relevance of each word.
For this research, coherence scores are tested for every LDA application. The resulting topic names are based on the keywords obtained for each of them.

In parallel, text from the front and back labels is preprocessed using the NTLK package. Unfortunately, due to the brevity of the wine label text, LDA is not suitable. Therefore, Term Frequency-Inverse Document Frequency (TF-IDF) is a more effective method.

TF-IDF evaluates word importance within documents, making it practical for short text analysis such as wine labels and ensuring more reliable insights. This technique is commonly used in natural language processing and information retrieval to measure term importance (Ghosh & Desarkar, 2018). As a result, it provides the main keywords in wine labels to be compared with the words within the review topics obtained by LDA.

3.3 Comparison between text in labels and reviews

Finally, text similarity contrasts the common topics in wine labels and reviews. This method assesses the extent of comparability between two texts and can cluster documents by content (Huang, 2008). Various techniques have been employed to understand text similarity. For instance, Zhang (2019) used Jaccard similarity to compare primary topics derived from LDA in Airbnb and hotel reviews.

Jaccard similarity quantifies the overlap between two sets. Its values range from 0 to 1, with higher values indicating greater similarity (Wahyuningsih et al., 2021). This research employs this measure to examine the similarity between the primary keywords in wine labels and those in consumer reviews.

Subsample analyses were conducted to explore further the key factors influencing wine purchases, including wine varietal, region, rating, and type of label (front-back). The outlined processes described in sections 3.1, 3.2, and 3.3 were also applied to these subsets.

4. Results and Discussion

The investigation identified the primary matters discussed by reviewers when evaluating Portuguese wines on Vivino. The analysis revealed that sensory characteristics, particularly fruitiness, tannins, acidity, aroma, and body, were most frequently mentioned, along with references to the regions Douro and Alentejo. The topics discussed by users were categorized into three main groups: Sensory wine characteristics, Douro wines, and Alentejo wines, as seen in Table 3.

Table 3. LDA topic dimensions for consumer reviews in Vivino

Topic	Keywords
Sensory wine characteristics (46.7%)	Fruta, Bom, Taninos, Boca, Acidez, Aroma, Final, Notas, Madeira, Corpo *(fruit, good, tannin, mouth, acidity, smell, end, notes, wood, body)*
Douro wines (33.5%)	Bom, Excelente, melhor, Ano, Pouco, Douro, Sabor, Grande, Equilibrado, Suave *(good, excellent, better, year, little, taste, big, balanced, soft)*
Alentejo wines (19.8%)	Alentejo, Sempre, Dia, Casa, Prova, Amigo, Saude, Top, Reserva, Quinta *(Alentejo, always, day, home, test, friend, health/ cheers, top, reserve, farm)*

Exploration of the "Sensory wine characteristics" topic revealed a consistent focus on sensory qualities, highlighting their ongoing importance in consumer discussions, "Douro wines" reviews emphasized quality, flavor, and balance. In contrast, "Alentejo wines" reviews focused on consistent quality and associations with home and friendship.

A temporal analysis over eleven years showed that sensory attributes remained the main topic. Wines from the Douro region were widely discussed until 2019, and Alentejo wines gained relevance after 2020.

Douro wines were more commonly discussed across all price tiers, while Alentejo wines were primarily mentioned in higher-priced segments.

Analyzing wine labels, the TF-IDF methodology identified primary keywords associated with the detailed portrayal of wines, including region, vineyard, and geographical indications, as seen in Table 4.

Table 4. Top 10 terms in the text of wine labels for the whole dataset, obtained through TF-IDF analysis.

Keywords	Frequency
Douro	480
quinta *(farm)*	334
produzido *(produced)*	209
produto *(product)*	176
reserva *(reserve)*	173
vinha *(vineyard)*	167
origem *(origin)*	134
denominação *(denomination)*	125
nacional *(national)*	118
controlada *(controlled)*	112

The term “Douro” is strongly associated with the renowned Portuguese wine-producing area, while “Quinta” *(farm)* and “produzido” *(produced)* relate to vineyards and production processes. The frequent terms “reserva” *(reserve)* and “vinha” *(vineyard)* emphasize geographical origins, while “denominação” *(denomination)* and “origem” *(origin)* suggest compliance with European Union labeling regulations (Ministério da Agricultura & Instituto dos Vinhos do Douro e do Porto, 2010).

The study contrasts the words in consumer reviews (Table 3) with those on wine labels (Table 4), with results for similarity shown in .

Table **5**.

Table 5. Jaccard similarity measured by topic

Topic	**Jaccard Similarity**
Sensory wine characteristics	0.00
Douro wines	0.08
Alentejo wines	0.11
Total Similarity	**0.05**

The findings indicate limited similarity between review topics and wine label information. Notably, the “Sensory characteristics” topic has no words in common with wine labels. Conversely, "Alentejo wines" exhibit the highest similarity, at 11.1%. Overall, the technical information on wine labels differs from user-generated content, primarily focusing on perceptual wine qualities.

Given the importance of varietal, region, and social media ratings in wine purchases, the study explored how these factors relate to the relationship between label information and user reviews.

4.1 Varietal

The varietal analysis revealed that 80.9% of wines are red, constituting 98% of reviews, while white wines comprised 19.1% of wines and 2% of reviews. LDA identified three topics for both red and white wine reviews, as seen in Table 6.

Table 6. LDA topic dimensions for red and white wines consumer reviews in Vivino

Red wines		White wines	
Topic	**Keywords**	**Topic**	**Keywords**
Sensory characteristics (50.7%)	Fruta, Bom, Taninos, Boca, Aroma, Acidez, Final, Corpo, Madeira, Notas *(fruit, good, tannin, mouth, smell, acidity, smell, body, wood, notes)*	**Sensory characteristics (49.6%)**	Bom, Aroma, Corpo, Acidez, Boca, Fruta, Final, Notas, Excelente, Nariz *(good, smell, body, acidity, mouth, fruit, end, notes, excellent, nose)*
Douro red wines (30.1%)	Bom, Excelente, Melhor, Douro, Preço, Sabor, Suave, Equilibrado, Encorpado, Qualidade *(good, excellent, better, price, taste, soft, balanced, full-bodied, quality)*	**Sensory and visual characteristics (25.7%)**	Bom, Acidez, Aroma, Final, Cor, Agradavel, Longo, Fruta, Boca, Amarelo *(good, acidity, smell, end, color, pleasant, long, fruit, mouth, yellow)*
Alentejo red wines (19.2%)	Ano, Garrafa, Grande, Gostei, Prova, Custo, Pouco, Alentejo, Quinta, Tempo *(year, bottle, big, like, test, cost, little, Alentejo, farm, time)*	**Aging traditional white wines (24.7%)**	Ano, Pouco, Palavras, Tradicional, Contudo, Saiu, Guarda, Tapada, Coelheiros, Fantástico *(year, little, words, traditional, however, left, guard, cover, Coelheiros, fantastic)*

In terms of sensory characteristics, red wines focus on descriptors like "fruta" *(fruit)*, "taninos" *(tannins)*, and "acidez" *(acidity)*. White wines, however, focus on "aroma" *(smell)*, "corpo" *(body)*, and "acidez" *(acidity)*. Douro red wines highlighted terms related to the region, while Alentejo red wines highlight elements like "ano" *(year)* and "custo" *(cost)*. White wines, particularly those associated with aging and satisfaction, focus on factors like "ano" (year) and "fantástico" (fantastic). The analysis suggests that red wine reviewers consider regions as important, while white wine reviewers have a more nuanced understanding of sensory and aging attributes.

For wine labels, the main topics remain broadly similar between the two varietals, as shown in Table 7.

Table 7. Top 10 terms in the text of red (left) and white (right) wine labels obtained through TF-IDF analysis

Keywords red wines	Frequency	Keywords white wines	Frequency
Douro	436	Douro	44
quinta *(farm)*	299	produto *(product)*	37
produzido *(produced)*	174	quinta *(farm)*	35
vinha *(vineyard)*	151	produzido *(produced)*	35
reserva *(reserve)*	149	origem *(origin)*	34
controlada (controlled)	141	denominação *(denomination)*	29
produto *(product)*	139	controlada (controlled)	29
nacional *(national)*	116	Dão	27
origem *(origin)*	100	reserva *(reserve)*	24
denominação *(denomination)*	96	vinha *(vineyard)*	18

The analysis reveals that "Douro" appears prominently in both datasets, suggesting regional association. Keywords "Quinta" *(farm)*, "produzido" *(produced)*, "reserva" *(reserve),* "controlado" *(controlled)*, and "produto" *(product)* are frequent in both datasets. Despite this overlap, differences arise in white wines, also highlighting Dão, a specific wine region in Portugal.

Comparing the topics discussed by users with the most frequent words on the labels results in the findings seen in Table 8.

Table 8. Jaccard similarity measured by topic for red and white wines

Red wines		White wines	
Topic	**Similarity**	**Topic**	**Similarity**
Sensory characteristics	0.00	Sensory characteristics	0.00
Douro red wines	0.03	Sensory and visual characteristics	0.00
Alentejo red wines	0.13	Aging traditional white wines	0.00
Total Similarity	**0.06**	**Total Similarity**	**0.00**

The comparison indicated that the similarity of reviews and labels is low for both varietals. Notably, white wine reviews showed no shared words with wine labels, with all similarity scores at zero. The "Sensory characteristics" topic for red wines exhibited no similarity with label information, while the "Alentejo wines" topic displayed the highest similarity.

The results suggest a disconnect between sensory characteristics in user reviews and the information on wine labels, highlighting an opportunity for producers to enhance the communication of sensory attributes on the packaging for both varietals.

4.2 Region

Given the prominence of the Douro and Alentejo regions in the Portuguese wine industry (Instituto Da Vinha e Do Vinho, 2023), a detailed analysis was conducted.

Wines from Douro account for 42.8% of the total sample and 55% of the reviews, while Alentejo wines constitute 24.5% of the dataset and 21.8% of the reviews. The remaining wines were from other regions of Portugal.

LDA identified three distinct review groups for each region, emphasizing "Sensory characteristics" as the primary topic, as seen in .

Table 9.

Table 9. LDA topic dimensions for Douro wines consumer reviews in Vivino

Douro wines		**Alentejo wines**	
Topic	**Keywords**	**Topic**	**Keywords**
Sensory characteristics (52.1%)	Bom, Fruta, Taninos, Boca, Acidez, Final, Aroma, Corpo, Notas, Madeira *(good, fruit, tannin, mouth, acidity, end, smell, body, notes, wood)*	**Sensory characteristics (59.6%)**	Fruta, Bom, Taninos, Boca, Acidez, Aroma, Final, Madeira, Notas, Cor *(fruit, good, tannin, mouth, acidity, smell, end, wood, notes, color)*
Red wines from Douro (21.3%)	Touriga, Ano, Nacional, Frances, Grande, Bom, Tinta, Quinta, Vinha, Garrafa *(Touriga, year, national, french, big, good, red, farm, vineyard, bottle)*	**White wines from Alentejo (18.0%)**	Bom, Excelente, Sabor, Melhor, Encorpado, Branco, Equilibrado, Pouco, Preço, Ótimo *(good, excellent, taste, best, full-bodied, white, balanced, little, price, great)*
Best value and aging Douro wines (26.6%)	Bom, Melhor, Excelente, Preço, Qualidade, Vale, Beber, Suave, Ano, Tempo *(good, better, excellent, price, quality, worth, drink, soft, year, time)*	**Best-aging Alentejo wines (22.4%)**	Ano, Grande, Garrafa, Melhor, Esporão, Tempo, Bom, Prova, Adega, Amigo *(year, big, bottle, better, spur, time, good, test, wine cellar, friend)*

Both regions emphasize descriptors for sensory characteristics, such as "Bom" *(good)*, "Fruta" *(fruit)*, "Taninos" *(tannins)*, and "Aroma" *(aroma)*. However, Douro wines highlight "Corpo" *(body)* more prominently, while Alentejo wines focus slightly more on "Madeira" *(wood)*.

Red wines from Douro showcase grape varieties like “Touriga Nacional”, while white wines from Alentejo receive praise for being "Encorpado" *(full-bodied)* and "Equilibrado" *(balanced)*. Both regions share commonalities in terms of descriptors like "Bom" *(good)* and "Melhor" *(better),* emphasizing positive attributes associated with wines.

Consumers from both regions consistently emphasized “sensory characteristics” as the main topic. This finding aligns with previous analyses, suggesting a noteworthy trend.

The best value, aging attributes, and varietal topics followed. Reviews of Douro wines mainly discussed red wines, while those from Alentejo favored white wines.
Consequently, wine labels from both regions were analyzed, as seen in Table 10.

Table 10. Top 10 terms in the text of Douro (left) and Alentejo (right) wine labels obtained through TF-IDF analysis.

Keywords Douro wines	Frequency	Keywords Alentejo wines	Frequency
quinta *(farm)*	215	regional *(regional)*	66
controlada (controlled)	127	produzido *(produced)*	56
produzido *(produced)*	102	tinto *(red)*	39
produto *(product)*	95	reserva *(reserve)*	39
reserva *(reserve)*	94	vinha *(vineyard)*	38
vinha *(vineyard)*	91	nacional *(national)*	33
tinto *(red)*	89	produto *(product)*	30
denominação *(denomination)*	74	geográfica *(geographic)*	30
origem *(origin)*	74	ano *(year)*	28
Porto	66	quinta *(farm)*	28

Analyzing Douro and Alentejo's prevalent terms on wine labels reveals distinctive patterns. The Douro labels frequently feature "Quinta" *(farm)*, "controlado" *(controlled)*, "produzido" *(produced)*, and "reserva" *(reserve)*, emphasizing vineyards, controlled production, and reserved quality. Similarly, Alentejo wines often include "regional" *(regional)*, "produzido" *(produced)*, and "reserva" *(reserve)*. Around 56% of top keywords in Alentejo wines also appear in Douro labels, reflecting shared industry practices and conventions in wine labeling.

However, unique terms, such as "Porto" for Douro and "geografica" *(geographic)* for Alentejo, emphasize distinct regional attributes. This blend of shared and unique terminology highlights the nuanced identities that Douro and Alentejo communicate through their labels.

Consequently, the similarity between wine labels and user reviews was studied, with findings presented in Table 11.

Table 11. Jaccard similarity measured by topic for Douro and Alentejo wines

Douro wines		**Alentejo wines**	
Topic	**Similarity**	**Topic**	**Similarity**
Sensory characteristics	0.00	Sensory characteristics	0.00
Red wines from Douro	0.17	White wines from Alentejo	0.00
Best value and aging Douro wines	0.06	Best value and aging Alentejo wines	0.09
Total Similarity	**0.05**	**Total Similarity**	**0.02**

The results reveal minor differences between the Douro and Alentejo regions. The “sensory characteristics” topic continues to show no similarity with wine labels. This discrepancy signifies an opportunity for wine producers to incorporate sensory details into packaging, enhancing product information.

Upon further analysis, reviews from the “Red wines from Douro” group show the highest similarity with label information, while “Alentejo white wines” show no similarity. This suggests gaps in consumer awareness regarding technical attributes between the two groups. Additionally, the topic "best value and aging" on labels demonstrates low similarity in both regions, suggesting labels should better communicate these attributes.

This study accentuates the role of regional identity in shaping wine evaluations. It reveals that red wines from Douro should improve their labels by providing clearer information on aging potential and value. In contrast, Alentejo white wines must improve their communication by focusing on technical and sensory characteristics on labels.

4.3 Ratings

The study categorized wines by ratings. According to Vivino (2020), the highest-ranked wines, with ratings above 4.6, constitute the top five percentile and account for 24.4% of total reviews. The remaining wines have ratings of 4.6 and below.

Three distinct review groups emerge for each category, with “Sensory wine characteristics” remaining dominant. In addition, “Wines Harvested with Touriga Nacional” was notable among top-rated wines, aligning with Portugal's recognition of Touriga National as its finest red grape variety (Duarte et al., 2008). A detailed topic description is in Table 12.

Table 12. LDA topic dimensions for highest rated (above rating 4.6) wines consumer reviews in Vivino

Highest-rated wines (rating above 4.6)		The rest of the wines	
Topic	**Keywords**	**Topic**	**Keywords**
Sensory characteristics (47.9%)	Fruta, Bom, Taninos, Boca, Final, Aroma, Acidez, Madeira, Excelente, Longo *(fruit, good, tannin, mouth, end, smell, acidity, wood, excellent, long)*	**Sensory characteristics (56.3%)**	Bom, Fruta, Boca, Taninos, Acidez, Aroma, Final, Corpo, Notas, Madeira *(good, fruit, mouth, tannin, acidity, smell, end, body, notes, wood)*
Top Douro wines (29.2%)	Melhor, Bom, Douro, Ano, Tudo, top, Excelente, Grande, Preço, fantastico *(better, good, Douro, year, everything, top, excellent, big, price, fantastic)*	**Best price-quality wines (24.5%)**	Bom, Ano, Preço, Qualidade, Pouco, Excelente, Garrafa, Tempo, Grande, Melhor *(good, year, price, quality, little, excellent, bottle, time, big, better)*
Touriga Nacional harvest wines (22.8%)	Barca, Jovem, Ano, Melhor, Touriga, Velha, Nova, Amigo, Grande, Nacional *(ferry, young, year, better, Touriga, old, new, friend, big, national)*	**White wines from Douro (19.2%)**	Douro, Bom, Excelente, Melhor, Branco, Grande, Maravilhoso, Belo, Acompanhar, Amigo *(Douro, good, excellent, better, white, big, wonderful, beautiful, accompany, friend)*

Comparison between the highest-rated wines and the rest reveals similarities in sensory characteristics. Both categories describe wines as possessing "Fruta" *(fruit)*, "Bom" *(good)*, "Taninos" *(tannins)*, "Boca" *(mouth)*, "Aroma" *(smell)*, and "Acidez" *(acidity)*. Top-rated wines slightly emphasize more specific and complex flavor descriptors like “Touriga Nacional”, “Velha” *(old)* "Excelente" *(excellent)*, “Fantastico” (fantastic), “Top” and "Longo" (long). Meanwhile, the other wines showcase general terms, such as price-quality or white wines, instead of a specific grape variety.

.

Table 13 displays the information on the wine labels for both groups.

Table 13. Top 10 terms in the text of highest ranked wine labels (left) and wines rated below 4.6 (right), obtained through TF-IDF analysis.

Keywords highest rated wines	Frequency	Keywords rest of the wines	Frequency
Douro	26	Douro	454
quinta *(farm)*	20	quinta *(farm)*	314
vinha *(vineyard)*	19	produzido *(produced)*	200
controlado (controlled)	10	reserva *(reserve)*	168
produto *(product)*	10	produto *(product)*	166
produzido *(produced)*	9	controlado (controlled)	160
grande *(big)*	8	tinto *(red)*	149
Dão	8	vinha *(vineyard)*	148
tinto *(red)*	7	origem *(origin)*	128
nacional *(national)*	6	denominação *(denomination)*	119

For the highest-rated wines, terms like "Douro", "Quinta" (farm), and "Vinha" (vineyard) are notable, suggesting an emphasis on the origin of wine. The "Dão" in this category indicates regional distinction. In contrast, the topic “Rest of the wines” terms such as “denominação” (denomination), “origem” (origin) and “Reserva" (reserve) suggest broader regulatory information. Both categories share terms like “Douro”, "Quinta" (farm), "Vinha" (vineyard), “Produzido” (produced) and “Produto” (product), indicating a shared focus on terminology based on industry usage..

Table 14 compares the main words on labels and topics from user reviews.

Table 14. Jaccard similarity by topic for all wines, segmented by rating.

Highest rated wines		Rest of the wines	
Topic	**Similarity**	**Topic**	**Similarity**
Sensory characteristics	0.00	Sensory characteristics	0.00
Top Douro Wines	0.06	Best price-quality wines	0.06
Touriga Nacional harvest wines	0.06	White wines from Douro	0.09
Total Similarity	**0.03**	**Total Similarity**	**0.03**

Despite the small sample size of highly rated wines, results indicate that similarity between label information and reviews is consistently low across different ratings. Closer examination shows that the “sensory characteristics” topic results align with previous patterns. At the same

time, “Top Douro wines” exhibit lower similarity than “White wines from Douro”, particularly for wines with lower scores.

The topic of “Touriga Nacional” exhibits higher similarity than average for top-rated wines., likely due to its technical nature.

These findings suggest that the similarity between label information and reviews remains consistent across different ratings. Producers should consider providing more detailed information about grape variety on labels, as this was prominent in reviews for highly rated wines. Producers should align label information with consumer expectations, especially regarding sensory attributes, regardless of the wine's rating.

A summary of the main findings regarding the different factors affecting wine purchase drivers' is found in
Table **15**.

Table 15. Summary of Jaccard similarity measured by wine purchase driver

Wine Purchase driver	**Jaccard Similarity**	
Varietal	**Red wines**	**White wines**
	0.06	0.00
Region	**Douro wines**	**Alentejo wines**
	0.05	0.02
Rating	**Highest-rated wines**	**Rest of the wines**
	0.03	0.03

The results reveal that label information terms are present in user reviews only in a limited way, even under these conditions. Although red and Douro wines exhibit a slightly higher degree of similarity than their counterparts, wine producers still have plenty of room to integrate the insights obtained from analyzing each of these drivers. Particularly noteworthy is the treatment of sensory attributes, which constitutes a salient topic in every analysis and consistently demonstrates no discernible relationship with the information presented on wine labels, regardless of the specific case under consideration.

4.4 The Role of front and back label information

In the final analysis, the study categorized wine labels by type. Comprehensive front-label data was obtained, while back-label information was manually collected for 42.6% of the wines, which accounted for 57.1% of the consumer reviews.

The dataset was divided into wines with front and back label information, with key terms exhibited in Table 16.

Table 16. Top 10 terms in the text of front wine label (left) and back label (right), obtained through TF-IDF analysis.

Keywords front label	Frequency	Keywords back label	Frequency
Douro	296	Douro	92
quinta *(farm)*	201	quinta *(farm)*	66
reserva *(reserve)*	170	produzido *(produced)*	56
vinha *(vineyard)*	137	produto *(product)*	32
origem *(origin)*	102	vinha *(vineyard)*	29
controlada (controlled)	96	nacional *(national)*	28
denominação (denomination)	76	origem *(origin)*	27
grande *(big)*	73	garantia (guarantee)	25
regional *(regional)*	72	Porto	22
produzido *(produced)*	70	denominação *(denomination)*	19

Front labels seem strategically designed to highlight regional and quality indicators, while back labels delve into production specifics and assurances. Terms such as “Douro”, “Quinta” *(farm)*, and “Reserva” *(reserve)* are present on front labels, while back labels feature terms like “Garantia” *(guarantee),* and “Porto.”

It was observed that wineries repeat around 70% of the information on the front label on the back. These shared attributes typically relate to the wine’s origin, including region, vineyard, and denomination. Both labels often include information about wine type, production year, and producer and winemaker names.

The findings from the similarity analysis for each label are presented in Table 17.

Table 17. Jaccard similarity is measured by topic for wines by type of label.

Topic	Front label similarity	Back label similarity
Sensory characteristics	0.00	0.00
Douro wines	0.08	0.09
Alentejo wines	0.14	0.06
Total Similarity	**0.05**	**0.03**

The findings reveal that user review similarities scored low for both labels. Reviews focusing on sensory wine characteristics remain unaffected by label information. However, similarities exist between front and back labels for "Douro wines" and "Alentejo wines". Overall, the similarity between labels and consumer reviews is deemed irrelevant, with all scores below 10%.

Based on these results, it is recommended that wineries balance the technical information on both labels to provide a holistic view of the wine and gain resemblance to consumer reviews.

5. Conclusions

The comprehensive analysis of user reviews on the Vivino platform, combined with examining wine labels, offered valuable insights into the similarities between user perceptions and label information.

Results reveal an important divergence between wine labels and consumer reviews, which aligns with the literature suggesting that consumers primarily focus on the sensory characteristics of wines. In contrast, wine labels prioritize descriptive attributes, such as regional cues and vineyard descriptives (Silayoi & Speece, 2004). These insights provide valuable contributions to academia, offering a foundation for further exploration of the nuanced interaction among consumer preferences, regional identity, and label information.

Furthermore, considering the limitations exposed in previous research (Jaud & Melnyk, 2020; Lam et al., 2019), the study of high-priced wines delved into the segmentation based on key characteristics, such as varietal, region, and rating. The results indicated that such categorizations did not substantially affect the similarity between wine label information and user reviews. However, nuanced insights emerged for specific varietals; red wine evaluations exhibited a pronounced inclination towards regional attributes, whereas white wine discussions demonstrated a more nuanced understanding beyond sensory and geographical

attributes. This divergence underscores wineries' importance in tailoring their communication strategies to accommodate varietal-specific nuances in consumer expectations.

Subsequent examination of regional differences revealed that Douro and Alentejo wine labels showed low similarity with reviews, highlighting the importance of effective labeling strategies in marketing and promoting Portuguese wines.

The study underscores the need for wineries to tailor communication strategies in response to these nuances. This could include promoting wine alongside food pairings, collaborations, or marketing claims related to sensory attributes, which aligns with the existing literature but highlights gaps where traditional labels fall short in addressing consumer preferences (Gunasekar et al., 2022; Yang et al., 2022). Notably, producers are encouraged to adopt consumer-centric language that aligns with user reviews' nuanced expressions of appreciation, words such as “taninos” (tannin), “aroma” (smell), “acidez” (acidity), “fruta” (fruit), “corpo” (body), “madeira” (wood) frequently appeared in reviews. Still, it was absent from wine labels, suggesting their potential inclusion.

Aligning marketing efforts with evolving sensorial preferences can enhance consumer resonance and brand loyalty. Understanding what aspects of wine labels resonate with users can inform producers and marketers about the elements that attract potential buyers.

Exploring the similarity of both wine labels concerning Vivino user discussions presents an opportunity for wineries to use the back label to highlight some sensorial characteristics. By leveraging the available space, wineries can better capture consumer attention and align with their expectations.

Furthermore, the methodology employed in this research can be extended to analyze user-generated content in other fast-moving consumer goods industries, enabling investigations into the relationship between packaging information and reviews. This research challenges prevailing assumptions regarding consumer reliance on wine labels as the primary source for informed purchase decisions. The findings indicate that consumer reviews prioritize attributes, such as taste descriptors and sensory qualities, that labels may not consistently represent (Calheiros et al., 2017; Chen et al., 2011). Thus, this research contributes to the literature by revealing the potential limitations of traditional wine labels in conveying the most significant information to consumers.

A potential implication for online platforms, such as Vivino, is to consider providing a more comprehensive view of the wine. While only the front label is currently displayed, this research underscores the importance of the back label. Vivino could explore ways to allow users to access the back label or incorporate key sensory terms from consumer reviews alongside the front label, bridging the gap between the reviews and the product information on the page.

This study offers new methods for understanding label information and its similarity with consumer perceptions in the premium wine industry in Portugal and provides insights for practitioners.

5.1 Limitations and Future Research

Since the dataset utilized in this study focused exclusively on a portion of Portuguese wines, it limits the generalizability of the findings to encompass the broader market. Future research should consider analyzing a more diverse range of wines regarding price points, geographies, and ratings to understand the broader similarity of label information and consumer reviews. Including a broader range of ratings could provide a more balanced view of the overall wine landscape. Future research could address this limitation by incorporating a stratified sampling method that includes both high and low-rated wines, thereby capturing a more comprehensive array of consumer opinions and preferences.

Another limitation is that less than half of the wines in the dataset contain information only from the front label rather than the back one. Thus, back-label information needed to be captured manually. This requirement introduces the possibility of human error and subjectivity. Future studies should explore automated methods for collecting back-label information to ensure accuracy and efficiency.

Another limitation of this study is its exclusive focus on the textual components of wine labels, overlooking the potential impact of other labeling attributes, such as colors, designs, and packaging styles. Furthermore, the study does not consider the role of image icons, such as certification labels, which may also shape consumer decision-making. Future research could address these aspects to provide a more comprehensive understanding of how both textual and visual elements compare to online reviews.

The inclusion of regulatory information on wine labels, mandated for legal reasons, is a limitation of this study, especially given the changes in regulations that took effect in 2023. Therefore, finding a way to neutralize or exclude these terms and focus solely on comparing

user reviews with non-regulatory language on wine labels could be a valuable direction for future research.

Finally, exploring other factors influencing wine purchases, such as reviews from social media or sources beyond Vivino, could provide a more comprehensive understanding of consumer online reviews while reducing potential biases associated with relying solely on one platform.

Data availability

Although public, the data used in this research is owned by Vivino. Thus, it cannot be publicly shared by third entities. However, the authors welcome other researchers' inquiries to study the same data.

References

Alzate, M., Arce-Urriza, M., & Cebollada, J. (2022). Mining the text of online consumer reviews to analyze brand image and brand positioning. *Journal of Retailing and Consumer Services*, *67*. https://doi.org/10.1016/j.jretconser.2022.102989

Annunziata, A., Pomarici, E., Vecchio, R., & Mariani, A. (2015). European consumers' interest toward nutritional information on wine labeling: A cross-country analysis. *BIO Web of Conferences*, *5*. https://doi.org/10.1051/bioconf/20150504003

Barber, N. (2010). "Green" wine packaging: targeting environmental consumers. *International Journal of Wine Business Research*, *22*(4). https://doi.org/10.1108/17511061011092447

Barber, N., & Almanza, B. A. (2007). Influence of wine packaging on consumers' decision to purchase. *Journal of Foodservice Business Research*, *9*(4). https://doi.org/10.1300/J369v09n04_06

Bazen, S., Cardebat, J. M., & Dubois, M. (2023). The role of customer and expert ratings in a hedonic analysis of French red wine prices: from gurus to geeks? *Applied Economics*. https://doi.org/10.1080/00036846.2023.2257036

Bird, S., Klein, E., & Loper, E. (2013). *NLTK - Natural Language Toolkit*. Http://Www.Nltk.Org/.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(4–5). https://doi.org/10.1016/b978-0-12-411519-4.00006-9

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling. *Journal of Hospitality Marketing and Management*, *26*(7). https://doi.org/10.1080/19368623.2017.1310075

Chen, Y., Fay, S., & Wang, Q. (2011). The Role of Marketing in Social Media: How Online Consumer Reviews Evolve. *Journal of Interactive Marketing*, *25*(2). https://doi.org/10.1016/j.intmar.2011.01.003

Duarte, F. L., Teixeira, a, Costa, a, Ramos, P., Schuller, D., & Gomes, a C. (2008). Yeast Diversity Related With Touriga Nacional Grape Variety. *Environmental Research*, *1*.

eAmbrosia. (2023). *EU geographical indications register*. https://ec.europa.eu/agriculture/eambrosia/geographical-indications-register/

Fernandes Ferreira Madureira, T. C., & Simões de Sousa Nunes, F. J. (2013). Relevant attributes of Portuguese wines: Matching regions and consumer's involvement level. *International Journal of Wine Business Research*, *25*(1). https://doi.org/10.1108/17511061311317318

Ferreira, C., Robertson, J., Lam, J., & Vella, J. (2023). Expert reviews uncorked: Contrasting the differences in the language used in online reviews of white and red wine. *Journal of Wine Research*, *34*(2). https://doi.org/10.1080/09571264.2023.2190086

Gastaldello, G., Mozzato, D., & Rossetto, L. (2021). Drinking Covid-19 away: wine consumption during the first lockdown in Italy. *Bio-Based and Applied Economics*, *10*(3). https://doi.org/10.36253/bae-10044

Ghosh, S., & Desarkar, M. S. (2018). Class Specific TF-IDF Boosting for Short-text Classification: Application to Short-texts Generated during Disasters. *The Web Conference 2018 - Companion of the World Wide Web Conference, WWW 2018*. https://doi.org/10.1145/3184558.3191621

Gunasekar, S., Das, P., Dixit, S. K., Mandal, S., & R Mehta, S. (2022). Wine-experienscape and tourist satisfaction: through the lens of online reviews. *Journal of Foodservice Business Research*, *25*(6). https://doi.org/10.1080/15378020.2021.2006039

Gurusamy, V., & Kannan S. (2014). Preprocessing Techniques for Text Mining. *International Journal of Computer Science & Communication Networks*, *5*(1).

Hallez, L., Vansteenbeeck, H., Boen, F., & Smits, T. (2023). Persuasive packaging? The impact of packaging color and claims on young consumers' perceptions of product healthiness, sustainability and tastiness. *Appetite*, *182*. https://doi.org/10.1016/j.appet.2022.106433

Huang, A. (2008). Similarity measures for text document clustering. *New Zealand Computer Science Research Student Conference, NZCSRSC 2008 - Proceedings*.

IBM. (2021). *What is a REST API?* https://www.ibm.com/sg-en/topics/rest-apis

Instituto Da Vinha e Do Vinho. (2021). *Fórum anual vinhos de Portugal: Exportações*.

Instituto Da Vinha e Do Vinho. (2023). *Evolução da Produção Nacional de Vinho por Região Vitivinícola*. https://www.ivv.gov.pt/np4/163.html

Instituto Nacional de Estatística Portugal. (2023, December 22). *Human consumption of wine per capita (l/ inhab.)*. Https://Www.Ine.Pt/Xportal/Xmain?Xpid=INE&xpgid=ine_indicadores&indOcorrCod=0000178&contexto=bd&selTab=tab2&xlang=EN. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0000178&contexto=bd&selTab=tab2&xlang=EN

Jaud, D. A., & Melnyk, V. (2020). The effect of text-only versus text-and-image wine labels on liking, taste and purchase intentions. The mediating role of affective fluency. *Journal of Retailing and Consumer Services*, *53*. https://doi.org/10.1016/j.jretconser.2019.101964

Klostermann, J., Plumeyer, A., Böger, D., & Decker, R. (2018). Extracting brand information from social networks: Integrating image, text, and social tagging data. *International Journal of Research in Marketing*, *35*(4). https://doi.org/10.1016/j.ijresmar.2018.08.002

Kotonya, N., de Cristofaro, P., & de Cristofaro, E. (2018). Of wines and reviews: Measuring and modeling the vivino wine social network. *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018*. https://doi.org/10.1109/ASONAM.2018.8508776

Kumar Agariya, A., Johari, A., Sharma, H. K., S Chandraul, U. N., & Singh, D. (2012). The Role of Packaging in Brand Communication. *International Journal of Scientific & Engineering Research*, *3*(1).

Lam, J., Lambrechts, M., Pitt, C., & Afsharipour, A. (2019). When writing about wine: how ratings impact reviews. *Journal of Wine Research*, *30*(4). https://doi.org/10.1080/09571264.2019.1684250

Latuny, W., Lawalata, V. O., Pailin, D. B., & Ohoirenan, R. (2021). Sentiment Analysis of Consumers for Determining the Packaging Features of Eucalyptus Oil Products. *Jurnal Ilmiah Teknik Industri*, *20*(1). https://doi.org/10.23917/jiti.v20i1.13461

Leban, M., Seo, Y., & Voyer, B. G. (2020). Transformational effects of social media lurking practices on luxury consumption. *Journal of Business Research*, *116*. https://doi.org/10.1016/j.jbusres.2019.09.010

Leotta, M., Mori, F., & Ribaudo, M. (2022). Evaluating the effectiveness of automatic image captioning for web accessibility. *Universal Access in the Information Society*. https://doi.org/10.1007/s10209-022-00906-7

Li, X., & Lei, L. (2021). A bibliometric analysis of topic modelling studies (2000–2017). *Journal of Information Science*, *47*(2). https://doi.org/10.1177/0165551519877049

Löfgren, M., & Witell, L. (2005). Kano's Theory of Attractive Quality and Packaging. *Quality Management Journal*, *12*(3). https://doi.org/10.1080/10686967.2005.11919257

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA

Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Ministério da Agricultura, do D. R. e das P., & Instituto dos Vinhos do Douro e do Porto, I. P. (2010). *Regulamento de protecção e apresentação das denominações de origem e indicação geográfica da região demarcada do Douro e das categorias especiais de vinho do Porto* (242/2010; pp. 11988–11997). https://dre.pt/dre/detalhe/regulamento/242-2010-988313.

Moro, S., & Rita, P. (2022). Tasting the Port wine cellar experience: what features please the most? *Journal of Wine Research*, *33*(2). https://doi.org/10.1080/09571264.2022.2081140

Mueller, S., Lockshin, L., Saltman, Y., & Blanford, J. (2010). Message on a bottle: The relative influence of wine back label information on wine choice. *Food Quality and Preference*, *21*(1). https://doi.org/10.1016/j.foodqual.2009.07.004

Mueller, S., & Szolnoki, G. (2010). The relative influence of packaging, labelling, branding and sensory attributes on liking and purchase intent: Consumers differ in their responsiveness. *Food Quality and Preference*, *21*(7). https://doi.org/10.1016/j.foodqual.2010.07.011

OECD. (2022, November 30). *Key Findings Alcohol Portugal*. https://www.oecd.org/portugal/Preventing-Harmful-Alcohol-Use-Key-Findings-PORTUGAL.pdf

O.J. (2011). REGULATION (EU) No 1169/2011 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL . *Official Journal of the European Union*. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:304:0018:0063:en:PDF

O.J. (2023). Implementation of new EU wine labelling provisions following the amendment of Regulation (EU) No 1308/2013 of the European Parliament and of the Council and Commission Delegated Regulation (EU) 2019/33 . *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C_202301190

Pérez-Rodríguez, G., Baptista, J. P., Igrejas, G., Fdez-Riverola, F., & Lourenço, A. (2022). Use Social Media Knowledge for Exploring the Portuguese Wine Industry: Following Talks and Perceptions? *Scientific Programming*, *2022*. https://doi.org/10.1155/2022/2912770

Ramirez, C. D. (2010). Do Tasting Notes Add Value? Evidence from Napa Wines. *Journal of Wine Economics*, *5*(1). https://doi.org/10.1017/s1931436100001425

Rehurek, R. (2009). *Gensim - Topic modelling for humans*. https://radimrehurek.com/gensim/index.html

Rizo, A., Bartu, A., Laguna, L., & Tarrega, A. (2022). Effect of an opinion app on expectations and emotional responses of young consumers toward white wines. *Food Quality and Preference*, *103*. https://doi.org/10.1016/j.foodqual.2022.104706

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*. https://doi.org/10.1145/2684822.2685324

Sherman, S., & Tuten, T. (2011). Message on a bottle: the wine label's influence. *International Journal of Wine Business Research*, *23*(3). https://doi.org/10.1108/17511061111163050

Silayoi, P., & Speece, M. (2004). Packaging and purchase decisions: An exploratory study on the impact of involvement level and time pressure. In *British Food Journal* (Vol. 106, Issue 8). https://doi.org/10.1108/00070700410553602

Stackla. (2021). *Post Pandemic Shifts in Consumer Shopping Habits*. https://www.nosto.com/blog/report-post-pandemic-shifts-in-consumer-shopping-habits-authenticity-personalization-and-the-power-of-ugc/#form_section

Szolnoki, G., Taits, D., Nagel, M., & Fortunato, A. (2014). Using social media in the wine business: An exploratory study from Germany. *International Journal of Wine Business Research*, *26*(2). https://doi.org/10.1108/IJWBR-09-2013-0031

Vecchio, R., Annunziata, A., & Mariani, A. (2018). Is more better? Insights on consumers' preferences for nutritional information on wine labelling. *Nutrients*, *10*(11). https://doi.org/10.3390/nu10111667

Veflen, N., Velasco, C., & Kraggerud, H. (2023). Signalling taste through packaging: The effects of shape and colour on consumers' perceptions of cheeses. *Food Quality and Preference*, *104*. https://doi.org/10.1016/j.foodqual.2022.104742

Vivino. (2020, October 20). *The Vivino Wine Rating System: Credibility of The Crowd*. https://www.vivino.com/wine-news/vivino-5-star-rating-system

Vivino. (2022a). *About Vivino*. https://www.vivino.com/about

Vivino. (2022b). *How much does a good bottle of wine cost?* https://www.vivino.com/wine-news/how-much-does-a-good-bottle-of-wine-cost#:~:text=An%20average%20bottle%20of%20red,costs%20%2432.48%20USD%2C%20on%20average

Vranken, S., Kurten, S., & Beullens, K. (2024). Time to 'Wine': A Content Analysis Investigating How Social MediaInfluencers Refer to Alcohol Use in Instagram's Feed Posts and Stories. *Mass Communication and Society*. https://doi.org/10.1080/15205436.2024.2309337

Wahyuningsih, T., Henderi, & Winarno. (2021). Text Mining an Automatic Short Answer Grading (ASAG), Comparison of Three Methods of Cosine Similarity, Jaccard Similarity and Dice's Coefficient. *Journal of Applied Data Sciences*, *2*(2). https://doi.org/10.47738/jads.v2i2.31

Wine Intelligence. (2019). *Importance of label design when choosing wine bottles worldwide* . https://www.statista.com/statistics/1166175/importance-of-wine-label-design-worldwide-us/

Wright, D. K., Yoon, H., Morrison, A. M., & Šegota, T. (2023). Drinking in style? Literature review of luxury wine consumption. *British Food Journal*, *125*(2). https://doi.org/10.1108/BFJ-06-2021-0661

Yang, C., Barth, J., Katumullage, D., & Cao, J. (2022). Wine Review Descriptors as Quality Predictors: Evidence from Language Processing Techniques. *Journal of Wine Economics*, *17*(1). https://doi.org/10.1017/jwe.2022.3

Yeo, S. F., Khoo, Y. H., Tan, C. L., & Lim, K. B. (2020). Product packaging: Impact on customers' purchase intention. *International Journal of Business and Society*, *21*(2).

Zhang, J. (2019). Listening to the Consumer: Exploring Review Topics on Airbnb and Their Impact on Listing Performance. *Journal of Marketing Theory and Practice*, *27*(4). https://doi.org/10.1080/10696679.2019.1644953