**Stellenbosch 2023 Abstract Submission**

**Title**
Exploring wine region sensory characteristics through text analysis: Insights from online reviews

**I want to submit an abstract for:**
Conference Presentation

**Corresponding Author**
Hanj Odendaal

**E-Mail Corresponding Author**
hanjo.odendaal@hotmail.com

**Affiliation**
Department of Economics, University of Stellenbosch, Stellenbosch, Western Cape South Africa

**Keywords**
- Text Analysis
- Structural Topic modelling
- Sensory Science
- Big Data

**Research Question**
I explore the nexus of traditional sensory analysis and computational linguistics by analyzing a text corpus of over 500,000 wine reviews.

**Methods**
Computational linguistics, Structural Topic Modeling

**Results**
I find that different wine regions were associated with distinct perceived flavor profiles and topics

**Abstract**

# Introduction

Wine tasting is a complex sensory experience, with numerous factors such as soil, climate, varietal, vinification processes, and the winemakers' expertise having an influence on the taster's perception of flavors and aromas. Traditionally, wine tasters use a set of standard flavor descriptors to characterize wines through what is known as sensory analysis or Sensory Science (SenSci). This process is time-consuming, expensive, can results in highly variable outcomes and may suffer from small sample bias [@francis2015application]. With the advent of text analysis techniques, it is now possible to analyze large amounts of wine reviews in order to gain insights into the flavor profiles of different wines. High dimensional text analysis allows for the sensory profiling to be extended beyond subjective characterization of wine and may prove to be a potential alternative for producing granular lexicons relating to sensory attributes.

In this paper, I explore the nexus of traditional sensory analysis and computational linguistics by analyzing a text corpus of over 500,000 wine reviews. The text itself is cleaned extensively as to only consider olfactory and gustatory factors. The purpose of the study is to analyse the difference in flavours of wine styles not only within wine regions, but across countries using structural topic modeling. I find that different wine regions were associated
with distinct perceived flavor profiles and topics. For example the black fruit, oaky and earthy characteristics identified within wine reviews differs significantly from South Africa, France, Australia and California. By identifying the most commonly discussed topics and flavor profiles associated with different styles and regions, winemakers and marketers can better understand how to position their products to appeal to different consumer preferences. The insights gained from these analyses could have profound implications not only for the field of flavour profiling for wine, but for the future of the wine industry.

# Text Analysis in Sensory Science

The traditional method of flavor profiling involves a structured process of sensory evaluation where trained professionals use a standard set of ‘descriptors’ to evaluate the flavor, aroma, mouthfeel, taste, appearance, or other sensory attributes of a wine [@lawless2013developing]. The Wine Aroma Wheel, first conceptualized by @noble1987modification, is a well-known example of such a system. This wheel provides a framework of aroma and flavor categories, such as fruity, spicy, and earthy, and allows tasters to identify and describe specific attributes within each category. The language used in this method is often technical and specific, and requires significant training to use effectively.

In recent years, with the increase of computing power and availability of machine-readable text, the burden of sifting through large amounts of text, among others to understand trends in the wine industry, has shifted to machines in what has evolved into a field called computational linguistics. The application of these techniques have become especially prominent within the analysis of the relationship between price and the online reviews (see e.g. @oczkowski1994hedonic, @lecocq2006determines, @krumme2009nose, @kotonya2018wines, @mccannon2020wine). Although the relationship between online reviews and prices are of interest, I posit that online reviews provide a rich set of information that can inform on flavour characteristics across regions and time. I analyse a data set collected from the popular wine review website, Vivino, using structural topic modeling, a well known natural language processing technique within computational linguistics. The key characteristic of this technique is that it captures latent clustering of tokens, or ```topics``` using an unsupervised machine learning algorithm and a corpus of compiled text. The main argument for using topic models is to aggregate core ideas from a large corpus of text, which would be too labour intensive for researchers to code, in order to better understand changing preferences or comparative narration across wine regions and time.

File Upload (PDF only)
- abstract.pdf

Consent
- ✔️ I agree to the privacy policy.

You find the link to our privacy policy at the bottom of the page. In the privacy policy you find a link for exporting and/or erasing your personal data stored in our database.
Exploring wine region sensory characteristics through text analysis
Insights from online reviews

Hanjo Odendaal*+

*University of Stellenbosch, Economics, Stellenbosch,

Abstract

In this paper, I explore the nexus of traditional sensory analysis and computational linguistics by analyzing a text corpus of over 500,000 wine reviews. The text itself is cleaned extensively as to only consider olfactory and gustatory factors. The purpose of the study is to analyse the difference in flavours of wine styles not only within wine regions, but across countries using structural topic modeling. I find that different wine regions were associated with distinct perceived flavor profiles and topics. For example the black fruit, oaky and earthy characteristics identified within wine reviews differ significantly from South Africa, France, Australia and California. By identifying the most commonly discussed topics and flavor profiles associated with different styles and regions, winemakers and marketers can better understand how to position their products to appeal to different consumer preferences. The insights gained from these analyses could have profound implications not only for the field of flavour profiling for wine, but for the future of the wine industry.

Keywords: Text Analysis, Structural Topic modelling, Sensory Science, Big Data

1. Introduction

Wine tasting is a complex sensory experience, with numerous factors such as soil, climate, varietal, vinification processes, and the winemakers’ expertise having an influence on the taster’s perception of flavors and aromas. Traditionally, wine tasters use a set of standard flavor descriptors to characterize wines through what is known as sensory analysis or Sensory Science (SenSci). This process is time-consuming, expensive, can results in highly variable outcomes and may suffer from small sample bias (Francis and Williamson, 2015). With the advent of text analysis techniques, it is now possible to analyze large amounts of wine reviews in order to gain insights into the flavor profiles of different wines. High dimensional text analysis allows for the sensory profiling to be extended beyond subjective characterization of wine and may prove to be a potential alternative for producing granular lexicons relating to sensory attributes.

In this paper, I explore the nexus of traditional sensory analysis and computational linguistics by analyzing a text corpus of over 500,000 wine reviews. The text itself is cleaned extensively as to only consider olfactory and gustatory factors. The purpose of the study is to analyse the difference in flavours of wine styles not only within wine regions, but across countries using structural topic modeling. I find that different wine regions were associated with distinct perceived flavor profiles and topics. For example the black fruit, oaky and earthy characteristics identified within wine reviews differs significantly from South Africa, France, Australia and California. By identifying the most commonly discussed topics and flavor profiles associated

*Corresponding author
Email address: hanjo.oden@gmail.com (Hanjo Odendaal)
with different styles and regions, winemakers and marketers can better understand how to position their products to appeal to different consumer preferences. The insights gained from these analyses could have profound implications not only for the field of flavour profiling for wine, but for the future of the wine industry.

2. Text Analysis in Sensory Science

The traditional method of flavor profiling involves a structured process of sensory evaluation where trained professionals use a standard set of ‘descriptors’ to evaluate the flavor, aroma, mouthfeel, taste, appearance, or other sensory attributes of a wine (Lawless and Civille, 2013). The Wine Aroma Wheel, first conceptualized by Noble et al. (1987), is a well-known example of such a system. This wheel provides a framework of aroma and flavor categories, such as fruity, spicy, and earthy, and allows tasters to identify and describe specific attributes within each category. The language used in this method is often technical and specific, and requires significant training to use effectively.

In recent years, with the increase of computing power and availability of machine-readable text, the burden of sifting through large amounts of text, among others to understand trends in the wine industry, has shifted to machines in what has evolved into a field called computational linguistics. The application of these techniques have become especially prominent within the analysis of the relationship between price and the online reviews (see e.g. Oczkowski (1994), Lecocq and Visser (2006), Krumme (2009), Kotonya et al. (2018), McCannon (2020)). Although the relationship between online reviews and prices are of interest, I posit that online reviews provide a rich set of information that can inform on flavour characteristics across regions and time. I analyse a data set collected from the popular wine review website, Vivino, using structural topic modeling, a well known natural language processing technique within computational linguistics. The key characteristic of this technique is that it captures latent clustering of tokens, or “topics” using an unsupervised machine learning algorithm and a corpus of compiled text. The main argument for using topic models is to aggregate core ideas from a large corpus of text, which would be too labour intensive for researchers to code, in order to better understand changing preferences or comparative narration across wine regions and time.

2.1. Structural topic modeling

Mathematically we can define a collection of documents as $D$, indexed by $d$, with each of the documents containing $w_d$ words. The unit of analysis within this paper is regarded to be all reviews for a wine in a given year with a particular vintage.

A topic $k$ can then be defined by the proportion of distinct words (or tokens) $V$ (equation 1), which is known as the topic prevalence matrix ($\theta$).

Each of these topics are associated with a $V$-dimensional probability mass function, $\beta_k$ (topic content matrix), that controls the frequency according to which tokens are generated from that topic (equation 2).

$$\theta = \begin{bmatrix}
\text{Doc1} & 0.20 & 0.01 & \ldots & 0.01 \\
\text{Doc2} & 0.20 & 0.03 & \ldots & 0.02 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\text{DocD} & 0.02 & 0.20 & \ldots & 0.04
\end{bmatrix}$$

(1)

$$\beta^T = \begin{bmatrix}
\text{text} & 0.020 & 0.010 & \ldots & 0.001 \\
\text{data} & 0.220 & 0.030 & \ldots & 0.002 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\text{analysis} & 0.020 & 0.204 & \ldots & 0.004
\end{bmatrix}$$

(2)
where the $D \times K$ matrix in equation (1) defines the topic prevalence, and $V \times K$ matrix in equation (2) defines topical content.

The difference between Blei and Lafferty (2007) and Roberts et al. (2014), is that in the latter, the prevalence and (topical content) are no longer global parameters shared by all documents, but are optional inputs specified as document-level covariates.

In this paper we employ both prevalence and topic content within the estimation of the parameters.

3. Implications for the Future of the Wine Industry

By analyzing large amounts of unstructured data, we can identify trends and patterns that may not be apparent through traditional flavor profiling methods. This information have implications in how the wine industry structures marketing strategies through highlighting specific flavor profiles, targeting certain demographic groups or even showcasing similarity between regions. In addition, text analysis can be used to identify emerging trends in the industry, such as the increasing popularity of organic or biodynamic wines, change in taste preference or even up and coming wine regions.

Although computational linguistics and machine learning shows promise, there are also challenges associated with the use of text analysis for SenSci in the wine industry. The first challenge is the lack of standardized language used to describe wine characteristics within wine reviews. Another challenge is the potential for bias in the reviews, as certain types of reviewers may be more likely to use certain language, focus on certain aspects of the wine or purely just review more wines than other reviewers. These challenges highlight the need for careful decision-making and explicit documenting of the data preprocessing steps that were applied before the analysis was conducted. Following established preprocessing protocol in text analysis will improve the validity and reliability of the results, as well as make them reproducible for other researchers.

In conclusion, the combination of traditional SenSci and text analysis techniques has the potential to provide valuable insights into not only the flavor profiles of different wines, but allows researchers to integrate large sets of meta information into analyses. By exploiting meta information related to large amounts of online reviews, researchers stand a better chance to understand not only the what, but the why behind their results. Large unstructured data sets such as online wine reviews provide a possible new source that could facilitate in a deeper understanding of the factors that influence the perception of sensory attributes and broadens our understanding of the relationship between (and within) regions.

References