# How Many Latours Is Too Many? Measuring Brand Name Congestion in Bordeaux Wine

Christopher Buccafusco [a], Jonathan S. Masur [b] and Ryan Whalen [c]

## Abstract

Firms rely on brand names to market goods to consumers, and consumers rely on brand names to locate goods that satisfy their preferences. If multiple firms are using the same or similar names, consumers may be confused about which product to buy, and firms may not obtain the benefits of their investments in quality. Recently, both firms and scholars in a number of industries have expressed concern about brand name congestion—too many firms clustering around too few terms. This paper applies computational linguistic analysis to chateau names in the Bordeaux wine region to study the degree of brand congestion within a mature, traditional, and high-value market. We find that Bordeaux producers have highly similar names to one another, far more than in comparable wine regions such as California and Alsace. More than a quarter of all Bordeaux producers have a name that is identical or nearly so to at least one other producer, and many terms are claimed by dozens of different producers. Interestingly, however, we find that the most famous and renowned producers have names that tend to be more distinctive than their less famous brethren. (JEL Classifications: C88, D83, L66, O34)

**Keywords:** distinctiveness, marketing, text analysis, trademarks, wine branding.

## I. Introduction

Firms rely on brand names to market goods to consumers, and consumers rely on brand names to locate goods that satisfy their preferences. An entire field of law—

[a] Benjamin N. Cardozo School of Law, Yeshiva University, 55 Fifth Ave., New York, NY 10003; e-mail christopher.buccafusco@yu.edu.
[b] University of Chicago School of Law, 1111 E. 60th Street, Chicago, IL 60637; e-mail: jmasur@uchicago.edu.
[c] University of Hong Kong Faculty of Law, 10/F, Cheng Yu Tung Tower, Centennial Campus, Pokfulam Road, Hong Kong; e-mail: whalen@hku.hk (corresponding author).

trademark law—exists to promote the efficient use of brand names as trustworthy signals between producers and consumers (Beebe and Fromer, 2018). Trademarks both reduce search costs for consumers looking to replicate previous positive experiences and protect firms' investments in creating products that consumers value.

Imagine a consumer who recently purchased and enjoyed a bottle of wine labeled Château Belair from Bordeaux. Now, she wants to order a case of Château Belair for an upcoming event. When she searches online, she discovers that multiple wines are being sold as Château Belair. Here, the consumer runs the risk of not being able to purchase the wine that she wants, and the original Château Belair risks losing sales to its competitors, some of whom may sell for lower prices because they do not maintain the same standards of quality. That risk, in turn, may cause Château Belair to reduce its investment in producing high-quality wines when it realizes that it cannot easily capture the benefits of doing so (McKenna, 2007).

For these reasons, both branding experts and trademark law encourage firms to choose names that are distinctive from their competitors. Distinctive, in this sense, means that the brand name is linguistically distant from the names other firms are using (Beebe, 2004). However, in some cases, firms have strong reasons to choose names that are not especially distinctive, including common terms or geographic signifiers. Especially in industries like wine, where the concept of *terroir* may attach a sense of quality to particular locations, geographical terms may have significant value (Cross, Plantinga, and Stavins, 2011). In addition, firms may have an incentive to cluster their names around better-known producers to trade-off their goodwill. If Château Belair has a reputation for quality, a firm might elect the name Chateau Bel Air and attempt to ride Chateau Belair's coattails.

In this paper, we undertake an empirical analysis of brand name similarity among wine producers. Here, we focus on wineries in Bordeaux, one of the largest and most economically valuable wine regions in the world. Bordeaux's status makes it a region of particular interest for brand name similarity analysis. The value that consumers place on tradition and prestige may subject Bordeaux producers to a different set of incentives than those in younger wine-producing regions or regions that place less emphasis on a producer's historical reputation. We thus build on earlier work involving linguistic analysis of wine descriptors (Yang et al., 2022; Capehart, 2021a, 2021b; McCannon, 2020), though here we focus on the names of wine producers rather than the words used to describe wines.

We find that Bordeaux producers have highly similar names to one another, far more than in comparable wine regions such as California and Alsace. Interestingly, however, we find that the most famous and renowned producers have names that tend to be more distinctive than their less famous brethren. Together, these data tell a story of vast numbers of wineries clustering around a set of common words and phrases, but with a small percentage managing to stand out from the pack both linguistically and qualitatively.

## II. Data and Methods

*Data.* Our Bordeaux data is sourced from *Lewin Guides' Vinidata World Wine Producer Database* (Vinidata, 2021), a database with extensive coverage of Bordeaux wineries. It includes both producer names as well as metadata on a total of 12,289 Bordeaux wineries. Our data include only winery names and not the names of individual wines. For example, our data do not include "second wines," which are, in effect, brand extensions and often have similar names to the *grand vin* (Lewin, 2009). To provide comparative analyses, we source global producer name data from the LWIN database provided by wine market data broker Liv-ex (2021).

*Methods.* Objective measurements of the similarity between names or marks are difficult to establish. We take a computational approach to this task. Our linguistic objective is to take two strings, compare them to one another, and produce a quantitative estimate of their similarity. We choose to use the Python difflib sequencematcher function, both because of its relative speed and because it produces results that tend to correspond to human intuition (Wołk and Marasek, 2014). This function takes two strings and computes their match ratio, which ranges from 0 for strings with no sets of characters that are the same to 1 for identical strings. The match ratio is defined as 2.0 * m/t where $t$ is the total number of characters, and $m$ represents the sum of the length of the matching subsequences across the input strings. Thus, minimal overlaps, such as sharing a single character, will produce low match ratio scores, while sharing large subsequences of characters will result in high match ratios. For example, *Domaine de Saint Marc* and *Château Saint Marc* have a match ratio of 1.0 after pre-processing that removes generic terms and articles. *Château Saint Marc* and *Le Lion Saint Marc* have a somewhat lower match ratio of 0.82 because of the "Le Lion" portion of the names, which do not match. *Le Lion Saint Marc* and *Château Latour* have a low match ratio of only 0.32 because they share only a few characters scattered across their names.

To accurately gauge the linguistic similarity between names, we remove common words such as "Le" ("the") and Château.[1] In selecting which words to remove, we were guided by the principle that few wine consumers would confuse two wines merely because they had the word "Château" in their names. We attempted to remove all such highly common words that would not likely lead to consumer confusion. If we were overzealous in doing so, this would bias our similarity scores downward.

---

[1] We clean each name by removing case characters and common words (cave, château, clos, cru, domaines, fils, haut, vignobles, and French articles) and subsequently removing spaces. Prior to removing articles, we combine them with the following word if the resulting word is in the set of words used by Bordeaux producers. So, if a winery's name contains the phrase "la tour," this would be reconciled to a single word "latour" prior to similarity comparisons.

## III. Results

### A. Total Similarity

Although there is no objective threshold at which name similarity becomes problematic, a very high similarity score suggests that there is at least one other producer with a high degree of name overlap. A firm with a competitor whose name is close to its own is, all else equal, less distinctive and subject to greater risk of consumer confusion than one whose competitors' names are more different. Accordingly, we compute the pairwise match ratio for each pair of producers and subsequently determine the maximum similarity for each producer. In effect, the maximum similarity represents the proximity of each producer's nearest neighbor in terms of name overlap.

We begin by simply counting the number of producers with a name that is highly similar to at least one other producer. Doing so reveals that 3,458 producers within our data have a nearest-name neighbor with a match ratio of at least 0.95, even after removing generic words and articles from a producer's name. This means that more than 28% of the 12,289 Bordeaux producers in our dataset share a *virtually identical name* with another Bordeaux producer.

### B. Word Reuse

In some cases, a firm's name might have diminished distinctiveness not because its entire name is similar to another firm's name but rather because another firm uses a significant segment of the firm's name. In Bordeaux, this sort of similarity could happen for a number of reasons, including prior similar ownership (e.g., the Léoville estates), use of geographic terms (e.g., Brane Cantenac and Cantenac Brown), or choices of common terms (e.g., references to towers, windmills, or good air). From the perspective of many consumers, especially non-Francophone consumers, similarities like these could confuse a wine's source.

Accordingly, we also analyze word reuse across producers in addition to the match ratio of entire name pairs. After cleaning generic terms and articles, there are 7,869 words used in Bordeaux producer names. This reveals that there must be a significant overlap in word usage by the 12,289 producers in our dataset. Figure 1 provides insight into common word usage across the producer set. We find that 584 words are appearing in at least 5 distinct producer names, 226 words that feature in at least 10 producer names, 58 words common to at least 25 producers, and 22 terms that appear in at least 50 different producers' names.

The phrase "la croix" provides an example of both name identicality and word reuse in producer names. There are more than 100 Bordeaux producers using some version of "la croix" or "lacroix" in their names. These include three Château la Croix, two Château de la Croix, a Château Lacroix, Domaine la Croix, Cru de La Croix, Château Haut la Croix, and Vignobles Lacroix.
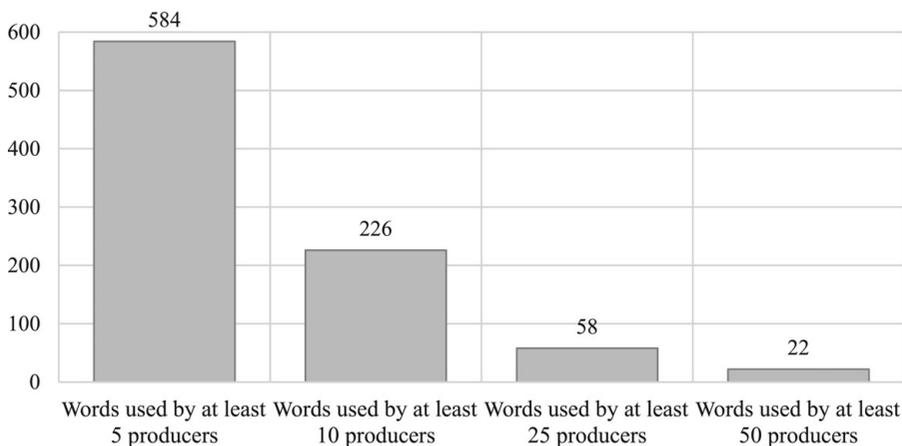
*Figure 1*
**Frequently Reused Word Occurrences**

Producers using some variation of "tour" or "latour" also demonstrate the widespread potential for confusion. Thirty-three producers use either "latour" or "la tour" in their names—including Château Latour, Domaine Château Latour, and Clos de la Tour—and more than 200 producers reference a tour of some sort—including many highly similar or exact matches such as three Château Tour Canons, three Château Tour Blanches, and two Château Tour Bel Airs. Indeed, names with non-unique words such as *lacroix* or *latour* are more common than not. Only about 28 percent of producers have names that use no words appearing in another producer's name, even after removing generic terms and articles.

## C. Comparative Analysis

Absent an objective threshold for problematic similarity, a comparative analysis can provide insight into whether a region has unusual levels of producer name similarity. By way of comparison to Bordeaux's prestigious and tradition-bound wine market, we compare the nearest neighbor similarity distributions of Alsace and California producers (Figure 2).[2] Alsace offers a domestic comparison, showing the extent to

---

[2] We pre-process Alsatian names in the same manner as Bordeaux names and California names. For Californian producers, in addition to removing case characters and whitespace, we remove English articles and common terms (cellars, clos, estate(s), family, vineyard(s), wine, and winery).

**Kernel Density Estimates Plots of Producer Name Similarity Distributions**

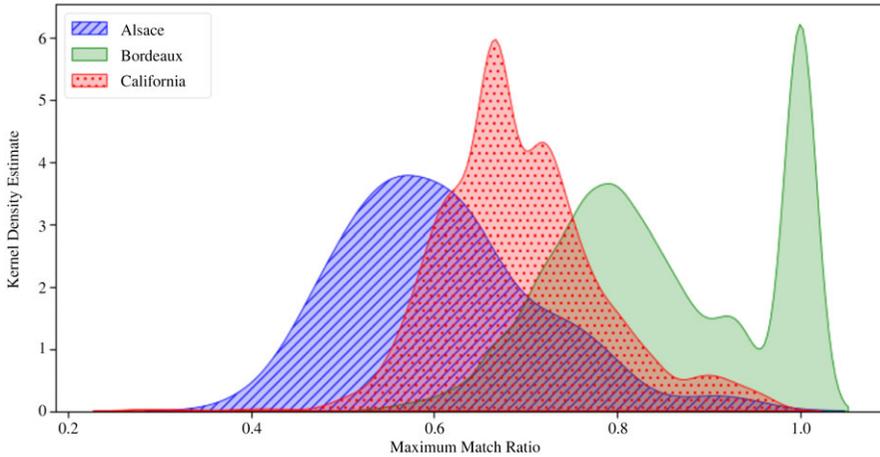which Bordeaux naming tendencies might differ from those of other French wine regions. California offers an international comparison via a region that is often compared to Bordeaux by wine critics and drinkers.

We see that Bordeaux producers are much more likely than producers in Alsace or California to have a highly similar nearest neighbor. In Bordeaux, chateaux tend to have more similar names overall, and in addition, a startlingly high percentage of chateaux have a competitor with a nearly identical name. This suggests a greater potential for consumer confusion among Bordeaux consumers than Alsace or California consumers. Robustness checks using both the average similarity and the mean similarity among the ten nearest neighbor names show similar results, with Bordeaux wineries demonstrating consistently more similar naming tendencies.

## D. Geography and Quality

What might account for the comparatively high degree of name similarity among Bordeaux producers? To provide more insight into this question, we turn to potential correlates of geographic proximity and producer prestige.

Some wine producers include references to geographic or topographic features in their names. For instance, some producer names will reference their regions, such as Château Grand Médoc or Château Blancaillou Médoc. Accordingly, one hypothesis is that producer name similarity should correlate with geographic proximity, with producers who are closer to one another more likely to share common names. To explore this possibility, we use producers' addresses to identify producer latitude

and longitude coordinates,[3] and subsequently, calculate pairwise geographic distances between producers. Doing so reveals that there is a significant, if substantively relatively minor, relationship between geographic distance and producer name similarity (Figure 3) ($r = -0.01$, $p < 0.0001$). This suggests a relatively minor tendency for nearby producers to cluster around similar names.

Quality and fame are also possible correlates of name similarity. A producer who wishes to enter the market might be more inclined to choose a name that is similar to a famous producer's and thus potentially benefits from consumer confusion. On the other hand, famous producers might be more aggressive in policing their brand names and deterring other producers from adopting similar names. It is also possible that causality runs in the other direction. It might be harder at the margins for a producer to acquire a reputation for high quality if there are other similarly named producers.

To explore this question empirically, we compare the name similarity of Bordeaux producers that have been included in official wine classification ranks (e.g., the 1855 Medoc classification, the various Cru Bourgeois classifications, or the St. Emilion or Graves specific classifications) with those of unclassified producers.[4] We manually identify 634 Bordeaux producers that have ever been classified in one of the classifications.

Examining the nearest neighbor similarity distributions across both classified and unclassified producers (Figure 4) shows somewhat similar distributions but with a markedly higher tendency for unclassified producers to have very similar nearest neighbors. Comparing means across these distributions shows that, while the differences are minor in absolute terms, they are statistically significant ($T = 2.90$, $p < 0.01$). This suggests that classified producers are likely to have more unique names than their unclassified peers.

## IV. Discussion

This initial analysis explores the naming similarity patterns of Bordeaux wineries. We show that Bordeaux producers are more likely to have a highly proximate nearest linguistic neighbor than producers in other regions such as California or Alsace. We next demonstrate that part, but not all, of this name, similarity might be explained by geographic proximity. Finally, we show that, on average, classified

---

[3]We initially queried the Nominatim API with the producer address. If unsuccessful, we subsequently queried the Base Adresse Nationale France API. This allowed us to identify latitude and longitude for 12,249 producers. For some producers, we have full addresses—allowing for precise coordinates—while others have limited information. Where limited information is available, we assign the producer a location in the geographic center of the identified region (e.g., postal code area).

[4]Pomerol, a high-status region, has never been classified, but its small number of chateaux would not change the results.

*Figure 3*

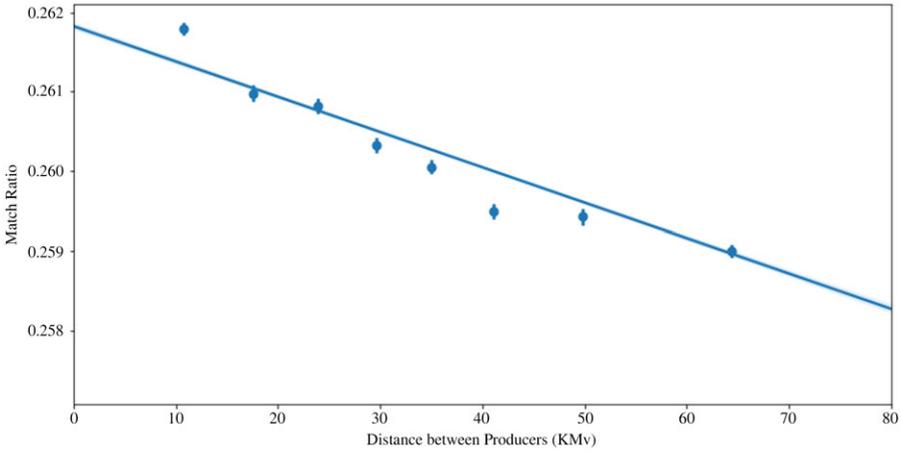**The Relationship between Distance and Name Similarity**

*Figure 4*

**Classified and Unclassified Producer Maximum Match Ratio Distributions**
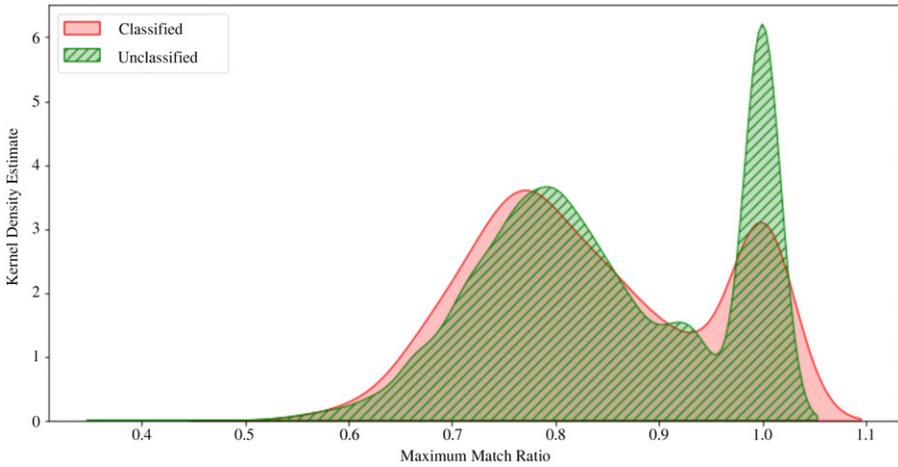
producers have somewhat more unique names than their unclassified peers, despite the incentives of lesser producers to choose names proximate to those of the prestigious classified producers.

Our data suggest that there may be substantial value for producers in some regions to cluster around particular traditional terms even when other names might help them stand out. Although Bordeaux producers could adopt more distinctive names—such as whimsical ones associated with animals—most seem to believe that the marketplace rewards those who follow the patterns laid down by other wineries. Consumers may tolerate or even desire creatively named wines from California or Australia, but they prefer their Bordeaux to resonate with traditional phrases and icons. Alternatively, Bordeaux producers may follow these naming conventions simply because of traditional norms within the industry. The use of whimsical names might lead to reputational sanctions if producers are seen as foregoing a tradition in favor of chasing market position (Sherman and Tuten, 2011). These sanctions could conceivably exceed whatever sanction might attend the clustering of names around certain words or phrases.

This analysis has consequences for both producers and consumers of Bordeaux wines. First, the likelihood of confusion is high for consumers of Bordeaux wines, higher than for comparable wines from California or Alsace. The likelihood of confusion is highest for the lower status Bordeaux wines, those that have never been classified. Consumers who are buying lower-priced Bordeaux wines are at risk of purchasing a wine they did not want, not least of all because they are probably less sophisticated as wine consumers.

For producers, name congestion makes clear how difficult it is for any wine to stand out from the crowd and how many wines must be drafting (intentionally or unintentionally) off of the reputations of their linguistic neighbors. This connotes a mature wine market, one that has existed for centuries and where new entrance will be especially difficult. This is in contrast to a younger wine market such as California's, where the linguistic space is not nearly so crowded, and new entrants will find it easier to distinguish themselves (Beebe and Fromer, 2018).

With more complete longitudinal data on pricing or founding years, we hope to further explore both the causes for name crowding as well as the extent to which name distinctiveness enables wineries to benefit from investments in quality.

## References

Beebe, B. (2004). The semiotic analysis of trademark law. *UCLA Law Review*, 51(3), 621–704.

Beebe, B., and Fromer, J. C. (2018). Are we running out of trademarks? An empirical study of trademark depletion and congestion. *Harvard Law Review*, 131(4), 945–1045.

Capehart, K. W. (2021a). Willingness to pay for wine bullshit: Some new estimates. *Journal of Wine Economics*, 16(3), 260–282.

Capehart, K. W. (2021b). Expensive and cheap wine words revisited. *Journal of Wine Economics*, 16(4), 000–000.

Cross, R., Plantinga, A. J., and Stavins, R. N. (2011). The value of terroir: Hedonic estimation of vineyard sale prices. *Journal of Wine Economics*, 6(1), 1–14.

Lewin, B. (2009). *What Price Bordeaux?* Dover, DE: Vendange Press.

Liv-Ex (2021). *Liv-Ex*. Available at http://liv-ex.com.

McCannon, B. C. (2020). Wine descriptions provide information: A text analysis. *Journal of Wine Economics*, 15(1), 71–94.

McKenna, M. P. (2007). The normative foundations of trademark law. *Notre Dame Law Review*, 82(5), 1839–1916.

Sherman, S., and Tuten, T. (2011). Message on a bottle: The wine label's influence. *International Journal of Wine Business Research*, 23(3), 221–234.

Vinidata (2021). *Vinidata World Wine Producer Database*. Available at http://www.vinidata.com/GuideView.

Wołk, K., and Marasek, K. (2014). A sentence meaning based alignment method for parallel text corpora preparation. In Á. Rocha, A. M. Correia, F. B. Tan, and K. A. Stroetmann (eds.), *New Perspectives in Information Systems and Technologies, Volume 1*, 229–237. Basel: Springer International Publishing.

Yang, C., Barth, J., Katumullage, D., and Cao, J. (2022). Wine review descriptors as quality predictors. Evidence from language processing techniques. *Journal of Wine Economics*, 17(1), forthcoming.