

What qualities of a wine will boost its review score?

By: Yuqiu Dong, Grace Zheng, Daniel Muñoz, Jimmy Li

Date: 25 April 2020

Key Question:

Determine which features of wines are predictors of getting great reviews to help businesses improve their reputation by identifying higher quality wine, and in turn increase wine sales.

Abstract:

There are so many different types of wine, but what helps differentiate a great wine from a good wine? We picked this question based on the assumption that a great wine score (given by a reviewer) will drive higher sales. We want to help wine businesses at many points in the supply chain, and we are interested in discovering which qualities of a wine lead to higher scores. This information will help businesses decide which types of wines to invest in. To investigate this, our team looked into a dataset that contained data on more than 120,000 wines. It contained information such as the description, score, price, country, province, region, taster, year, and variety. Through our analysis, we came up with several strategies to improve wine review scores.

Analysis Summary:

To answer the question of what contributes to a great score, our team used several models to see what features were able to better help us predict the score. For clarity, we defined that great wines were in the upper quartile of the wine scores in our dataset and will use this definition of great wines for the rest of the report. From here we built 2 models to predict the scores and 2 other models to better understand the textual information of the wines in relation to the score.

Results:

- By building models that predict great wines significantly better than random chance, we conclude that there are inherent qualities of a wine that make them score higher than others.
- Although one might presume that the age of a wine correlates to a better score, we found that there was no significant correlation between the year of the wine and its score.
- There is a moderate correlation between price and review score, where more expensive wines are more likely to receive higher scores. In particular, if the price of the wine is above \$42, it tends to be a great wine.
- Pinot Noir from California and Chardonnay from France are more likely to be great wines.
- There are descriptive words that are prevalent in both good and great wines, but some words such as “blackberry”, “dark”, and “rich” only appear in the descriptions for great

wines. After a more rigorous analysis, we discovered that the existence of the words “black”, “rich”, and “tannins” and the absence of the words “dry”, “oak”, and “acidity” in the descriptions were high indicators of a great wine.

- A significant portion (20%) of our reviews were completed by one prime taster, Roger Voss. Initially, we were concerned that Voss would bias the results of the dataset through his personal tastes. However, by building multiple models, we discovered that his scoring was representative of the entire dataset with more than 18 reviewers, so we concluded that the scoring of a wine is not strictly subjective.

Recommendations:

Using the results of our analysis, there are several initiatives we would like to propose for businesses:

- 1) Idea: Increase promotions for great wines based on certain taste characteristics.
Details: We note that great wines are often described with the following characteristics: “blackberry”, “dark”, “rich”, “black”, and “tannins”. Thus, we recommend tapping into this by giving samples to introduce consumers to great wines in order to increase awareness, reputation and ultimately sales.
- 2) Idea: Increase investments in great wines based on certain taste characteristics.
Details: After promoting these specific types of wines, we believe that wine producers and stores should invest more into wines having “blackberry”, “dark”, “rich”, “tannins”, and “black” characteristics. Promotion will increase sales, so stores should increase supply and wineries should increase production.
- 3) Idea: Invest in Wineries for Chardonnay in France and for Pinot Noir in California
Details: From our business knowledge, we assume that certain grapes grow better in particular climates. Our data shows that Pinot Noir grapes grown in California and Chardonnay grapes grown in France often lead to wines with great scores. Therefore, we think that it would be a good investment for wineries to open locations in California for producing Pinot Noir and in France for producing Chardonnay.
- 4) Idea: Implement spotlight advertising for “signature wines” - Pinot Noir and Chardonnay
Details: Our data indicates that these 2 wine varieties in the upper quartile for price have more great reviews. Prioritizing advertisements for these 2 wine varieties by placing them at the entrances of stores will draw attention to them, so people will be more inclined to try them and leave great reviews. Therefore, this will lead to more future purchases.

Conclusion:

Based on the analysis, we have identified several predictors of great wine reviews, so our business objective was satisfied. We believe implementing the above recommendations will lead to more great reviews for wines and in turn benefit businesses who work with us by increasing financial gains.

Technical Notes:

Above & Beyond Note: Our group had a lot of fun working on this project and wanted to use it as an opportunity to try some advanced data analysis techniques. We decided to try pre-processing by creating a custom function using regular expressions to extract the year out of a string of text, word cloud visualizations, using data analysis tools with faster processing power such as Google Collab and a GPU accelerator for text mining association rules, and delved into a personalized analysis for one prime reviewer (Roger Voss) for more insights.

Future Improvements for Analysis: Something worth looking into would be the use of clustering on the different wine locations. This would be beneficial because it would help us differentiate which specific types of wine in a location are better. Moreover, in order to validate our assumption that a great review would lead to an increase of sales, we would need to combine this dataset with sales data. With this additional information, we could also create customer segmentations to better target the sale of wines.