8-10-2020

Applying Text Data Analytics Techniques to Wine Reviews

Gregoire J. Thebaud
University of North Carolina Wilmington, gjt3372@uncw.edu

Pierre Kraus
University of North Carolina Wilmington, pk1377@uncw.edu

Minoo Mondaresnezhad
University of North Carol?ina Wilmington, modaresm@uncw.edu

Applying Text Data Analytics Techniques to Wine Reviews

TREO Talk Paper

Gregoire J. Thebaud
University of North Carolina Wilmington
gjt3372@uncw.edu

Pierre Kraus
University of North Carolina Wilmington
pk1377@uncw.edu

Minoo Modaresnezhad
University of North Carolina Wilmington
modaresm@uncw.edu

Abstract

When purchasing a bottle of wine, we usually rely on the smell and flavor descriptions from experts' wine reviews. Just like restaurants, critics' reviews count as a marketing strategy in the wine industry. If your wine consistently impresses the tasters and makes their top picks, there is a good chance customer will follow and buy it. A study by Friberg and Gronqvist (2012) found that the effect of a favorable review peaks in the week after publication with an increase in demand of 6 percent, and the effect remains significant for more than 20 weeks (Friberg & Grönqvist, 2012). This project is part of a student paper for a Text Data Analytics course. The goal of this study is to evaluate the description of wines provided by the tasters. The data was scraped from Wine Enthusiast (https://www.winemag.com/), a world-known American multi-channel marketer during the week of June 15th, 2017. The dataset includes fields such as the type of grapes used to make the wine, the country that the wine is from, the number of points Wine Enthusiast rated the wine on a scale of 1-100, the cost for a bottle of the wine, taster full name, description of the wine by a sommelier, etc. We used R programming to analyze the dataset.

We will apply exploratory data analysis methods to find insights and interesting facts about this dataset. Specifically, we will count the number of reviews based on variety and flavors. Then, we will look at the total number of reviews by the taster. Tasters with at least 100,000 reviews can be categorized as master reviewers. We will use this knowledge to categorize reviews based on their expertise and experiences. We will also examine which variety of wine received the best score by reviewers and identify the reviewers' favorite types of wine. We will also calculate the point distribution based on reviewers.

Concerning the description column, the first step will be to pre-process and clean the text to deal with the missing values, remove the special characters, etc. We will apply descriptive text analytics techniques to understand the tasters' reviews better. We will calculate the most frequently used words in the descriptions, as well as the most important terms based on TF-IDF. We will also apply sentiment analysis and topic modeling techniques to categorize the tasters' reviews. Sentiment analysis enables us to extract insights from qualitative data, such as wine reviews. These insights can be used by the winemaker to respond to their clients' needs. By detecting positive, neutral, and negative opinions within the text, we can understand how tasters feel about a particular wine and make data-driven decisions. We will identify the polarity of reviews based on the taster's name, the varieties of wine, and the country. Polarity in sentiment analysis refers to identifying sentiment orientation (positive, neutral, and negative). Using it for our reviews could potentially yield exciting results. Finally, we will use classification methods to identify the common topics in the tasters' wine descriptions.

References

Friberg, R., & Grönqvist, E. (2012). "Do expert reviews affect the demand for wine?" *American Economic Journal: Applied Economics*, (*4*:1), pp.193–211.