

Final Project: Wine - Review

Team: Group 3 Fireflies

Roles:

Data cleanup and exploration: Chunhui Zhu,

Data Modelling: YueChun Wong, Chunmei Zhu

Overview

We want to create a predictive model to identify wines using words in description/review. It could identify the wine based on a description that a sommelier could give.

<https://github.com/czhu505/Data620-Fireflies/tree/master/Final%20project>

The data was scraped from [WineEnthusiast](#) on November 22nd, 2017. The code for the scraper can be found <https://github.com/zackthoutt/wine-deep-learning>.

The author collected the title of each review, the tasters name, and the taster's Twitter handle.

Data

We have the 9 attributes in 37420 samples in the data set.

<https://www.kaggle.com/zynicide/wine-reviews>

Data Meta

Country - The country that the wine is from

Description

Designation - The vineyard within the winery where the grapes that made the wine are from

Points - The number of points WineEnthusiast rated the wine on a scale of 1-100 (though they say they only post reviews for wines that score ≥ 80)

Price - The cost for a bottle of the wine

Province - The province or state that the wine is from

Rutherford inside the Napa Valley), but this value can sometimes be blank

taster_name

taster_twitter_handle

Variety - The type of grapes used to make the wine (ie Pinot Noir)

Winery - The winery that made the wine

This dataset contains three files:

- winemag-data.csv contains 11 columns and 37.4k rows of wine reviews.

Questions: Can we find the following:

- Most popular wine
- Most expensive wine (cost for each wine)

- What kind of wine better pair with some kind of foods (e.g. fish or steaks) based on text mining
- What "adjustive" or property can be found for wine. Would wines are tasted similar although they are different type?
- Most used wine "adjustive" or property
- Wine producer location vs grapes from location
- Wine examiner name -> wine -> any fake review?
- Wine examiner's twitter impact/efficiency
- Unique words for specific wine
- Most reviewed Wine title?
- Wine / grapes best year
- Winery type for wine title and
- Wine winemaking process concern

Concerns:

There are many wine specific vocabularies that we may require us to look up. Non-english words could be troublesome in text mining and clean up. Also, pronouncing the wine type in presentation could be challenging but educating and interesting to us.