

Project Proposal

ajd242, eer48, lfl42

September 20th, 2017

1 Question

How good is this bottle of wine?

2 Motivation

To quantify personal preferences is often a hard task to approach, however, if you've ever been shopping for a bottle of wine you have probably done so. When valuing a bottle of wine we often consider the producer, it's rating, the region which it is from and any reviews we have available to us. Unfortunately, we don't have a formula for determining if this wine is a good buy at its price point, but what if we could? Our goal is to determine if we can predict the selling price of a bottle of wine given its ratings, reviews, region, etc. This model can help producers understand the value of the wine they are making, and consumers to make smarter purchases. Ultimately, we hope our model for predicting the price of a bottle of wine will create a more efficient market for buying and selling wine.

3 Data/Approach

In order to predict wine quality, we will examine data obtained from <https://www.kaggle.com/zynicide/wine-reviews/data>. The data contains the following features: rating, description, variety, country, province, region 1, region 2, winery, designation, and price.

Our assumptions include that all ratings are by experts and are an accurate description of the bottle of wine. Our model also includes inputs like the grape varietal, region, producer, designation, and country to output the potential price of the bottle of wine. Additionally, we wish to gather data on the weather conditions for each vintage and use those to predict the future price and rating of each vintage. We will use this as a model to predict wine futures and as a method for producers to maximize the potential profit from each vintage.

We plan to analyze the description of a bottle of wine through a natural language processor (NLP) to predict the region, grape varietals, and producer. Similar to a sommelier, we look at the characteristics listed and use them to provide insight into the history of the bottle. Our current plan is to attempt to do so using word embeddings.