

ORIE 4741 Project Proposal: Wine Pricing Prediction

Yuanzheng Cao(yc2575), Tao Ge(tg367), Zhoutong Li (zl683)

September 2019

1 Question

What factors affect the price of wine and based on those features and models established, how do we predict wine price?

2 Motivation

If you have ever been shopping for a bottle of wine in liquid shop, you would probably be overwhelmed by different types of wines and their price. When evaluating a bottle of wine, we often consider the producer, the variety, the winery and the time that it produced. However, we do not have any formulas to calculate the price of the wine and do not know whether the real value matches the price labeled or not. Our goals and models aim to predict the price of the wien given its producer, variety, winery, ratings and reviews, etc. Furthermore, the model could not only help consumers know the approximate price given by their preference, but also make producers aware that what kind of wine are more valuable, and gain more profits.

3 Data and Methodology

Our project would briefly reveal the relation between wine quality points and taster descriptions, as well as the relation of price and main attributes such as region, title, variety and winery.

For our project we are using the Wine Enthusiast reviews dataset found in Kaggle. This dataset contains more than 100k lines of wine reviews and information from a professional wine information sharing and ecommerce website, Wine Enthusiast, originated from Wine Enthusiast Magazine. Available information includes the price, rating, title, designation and the description of the wine, as well as variety (type of grapes used), winery, region, province, and country of the wine. We plan to first extract some information from the text columns (title, description), and then train our model with these features along with other wine information.