Springboard–DSC Program // Capstone Project 2 Proposal

Red Wine Flavor Profiles and their Relation to Cost and Quality

By Steve Walters

May, 2023

Problem Statement (Hypothesis): Is it possible to reverse engineer the most popular or profitable flavor profiles in red wines so that a blend can be characterized that will rate a score above 92 and command a price that is at least 15% higher than the average price for similar blends?

Our client has access to a number of red wine varietals, but is uncertain which blend of those varietals would produce the highest rated wines and the potential for above average selling prices for the wines produced. They believe that, given a deeper understanding of the connections between flavor profile and price/quality rating, it is possible to create a red wine blend that will be above average in both rated quality and retail selling price. Using data from Wine Enthusiast we will attempt to determine what flavor profiles score highest and have the highest price. The client will be able to use that information to produce a superior wine with a higher than average selling price.

The aim of this project would be to use the combination of rating, prices, and textual reviews to try and answer the question: How does a winery know what flavor profiles to aim for in a given vintage? This could be useful in determining what flavor profiles are popular, and what flavor profiles wineries and winemakers should aim to create if they wish to satisfy current popular trends in red wine flavor profiles. This could not only inform winemakers about which varietals to combine in blends, but could also help inform wineries which varietals to plant, given that they have the proper climate for such varietals (which will not be part of this project).

The criteria for success in this case is defining a minimum of one flavor profile that is shown to increase either the rating of the wine by >=2 points, or increase the price by a minimum of 10%. The scope of this analysis is only red wine varietals, and only those that have received a rating score and a textual review, and have a listed price. In this way we can determine if there is a correlation between the flavor profile characterized by professional wine tasters in the textual review and the rating/price.

While the data does contain 130,000 rows, we might find ourselves constrained by N/A data in the rating, price, or review. It is also possible, though not likely, that the vast majority of the

reviews are focused on varietals that don't fit in our study i.e. white wines or dessert wines. If this is the case we could switch the focus of the study to white wine varietals, but given that red wine made up 46% of wine sales in 2020 and Cabernet Sauvignon is the most popular varietal in the U.S., this shouldn't be necessary.

The stakeholders could be the owner of a winery, a marketing manager at a winery, the winemaker, the person responsible for sourcing grapes (if they aren't being grown on-site), and potentially the retail manager if there is also a retail sales area at the winery.

Data sources: This project will use the Kaggle dataset from Wine Enthusiast (https://www.kaggle.com/datasets/zynicide/wine-reviews) includes over 130,000 rows of data and the following features; country, description, designation, #points, price, province/state, region_1, region_2, and the name of the reviewer. This dataset appears to have all the information we will need to conduct this study i.e. description, #points, price. That said, it does not contain a column for the varietal, but that can be extracted from the description through the terms that are used to describe red wines, such as 'red', 'cabernet', 'tannin' and the known flavor characteristics of red wines i.e. 'blackberry', 'plum', 'cedar', etc.

Deliverables:
- A GitHub repo containing the work completed for each step of the project, including:
  - A slide deck
  - A project report