* Problem Statement
  + For novice wine drinkers, it can often to be tough to determine what type of wine is associated with different reviews. Furthermore, novice wine drinkers are often unable to determine where the best rated wine is coming from, and also what words may be associated with better quality wine. While this will not have some kind of revolutionary impact, it could be an interesting start to determining how experts review wine and also how novice drinkers can purchase the highest quality based on text reviews.
* Data Set
  + I will be using a data set sourced from Kaggle. The data was scraped from *Wine Enthusiast* using a scraper sourced from GitHub. The following data is available in the data set:
    - *Rating*
    - *Variety*
    - *Description*
    - *Country*
    - *Province*
    - *Region 1*
    - *Region 2*
    - *Winery*
    - *Designation*
    - *Price*
  + I think it would be more interesting to have a source with a number of prices, which could then be compared to find the proper price at which a variety of wine with a certain rating should be purchased.
* Hypothesis
  + There are certain words and phrases that are associated with particular kinds of wine. When wine experts are describing wine, they use specific words that characterize the type of wine they are discussing. Perhaps, they even use certain words to describe wine from certain regions.
* Methods and Models
  + The model will be built using text classification to assign a value for words used for specific varieties of wine. There is also a possibility to use a latent variable model to discover patterns pertaining to how a varsity of wine is described, or how the highest rated wines are described by the experts. I would also like to explore the use of Word2Vec to replace overlapping contexts and also find words that are similar to each other to train the model to predict the correct variety of wine.
* Assumptions/Risks
  + There is an underlying risk, like with all machine learning projects, that there are simply words that are used often across all wine varieties. This would result in the model not being able to predict the proper wine variety with a high accuracy based on the inaccurate assumption that experts use specific words for types of wine, or wine with a higher rating. There is also a risk of duplicate varieties having different reviews, which may cause an issue within the model; duplicates may have to be removed which will decrease the about of observations available.
* Goal
  + My goal is to create a model that predicts the variety of a wine with a high accuracy. If possible, I would also like to explore if expert wine reviewers award a higher rating to certain wines in a singular region, or if there are words associated with higher rated wines. Referencing a Kernel from Kaggle, my goal would be to outperform a score of 53% accuracy on the top 5 reoccurring wine varieties.