Wine Reviews

130k wine reviews with variety, location, winery, price, and description

Context

After watching [Somm](#) (a documentary on master sommeliers) I wondered how I could create a predictive model to identify wines through blind tasting like a master sommelier would. The first step in this journey was gathering some data to train a model. I plan to use deep learning to predict the wine variety using words in the description/review. The model still won't be able to taste the wine, but theoretically it could identify the wine based on a description that a sommelier could give. If anyone has any ideas on how to accomplish this, please post them!

Content

This dataset contains three files:

- **winemag-data-130k-v2.csv** contains 10 columns and 130k rows of wine reviews.
- **winemag-data_first150k.csv** contains 10 columns and 150k rows of wine reviews.
- **winemag-data-130k-v2.json** contains 6919 nodes of wine reviews.

Click on the data tab to see individual file descriptions, column-level metadata and summary statistics.

Acknowledgements

The data was scraped from [WineEnthusiast](#) during the week of June 15th, 2017. The code for the scraper can be found [here](#) if you have any more specific questions about data collection that I didn't address.
UPDATE 11/24/2017 After feedback from users of the dataset I scraped the reviews again on November 22nd, 2017. This time around I collected the title of each review, which you can parse the year out of, the tasters name, and the taster's Twitter handle. This should also fix the duplicate entry issue.

Columns

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

region_1 - The wine growing area in a province or state (ie Napa)

region_2 - Sometimes there are more specific regions specified within a wine growing area (ie Rutherford inside the Napa Valley), but this value can sometimes be blank

taster_name

taster_twitter_handle

title - The title of the wine review, which often contains the vintage if you're interested in extracting that feature

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

winery - The winery that made the wine