

**Patricia Degner, Andrew Dively, Ashley Moss.**

Our GitHub repository is called Project2\_Degner\_Dively\_Moss.

**The main dataset** we would like to analyze is the [“Wine Reviews”](#) dataset from Kaggle.

Because, who doesn't like wine? This file is too big to upload to GitHub, but can be found on the Kaggle link.

**Supplemental dataset:** list of wine notes.

[https://en.wikipedia.org/wiki/Wine\\_tasting\\_descriptors](https://en.wikipedia.org/wiki/Wine_tasting_descriptors)

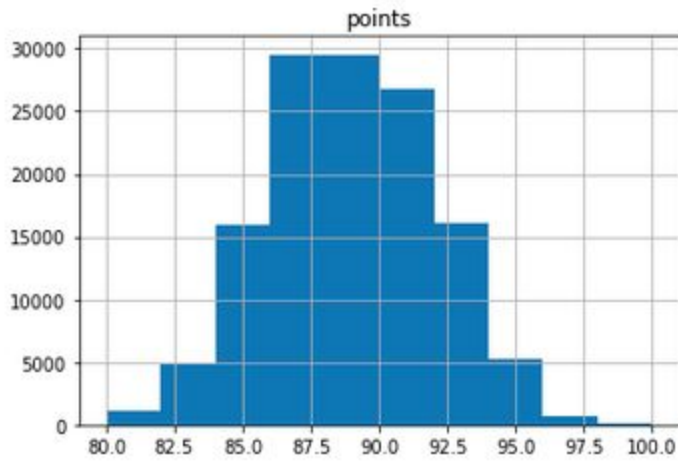
Our data is 129971 observations of 14 variables describing:

- Index
- Country: the country 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)
- 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
- 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

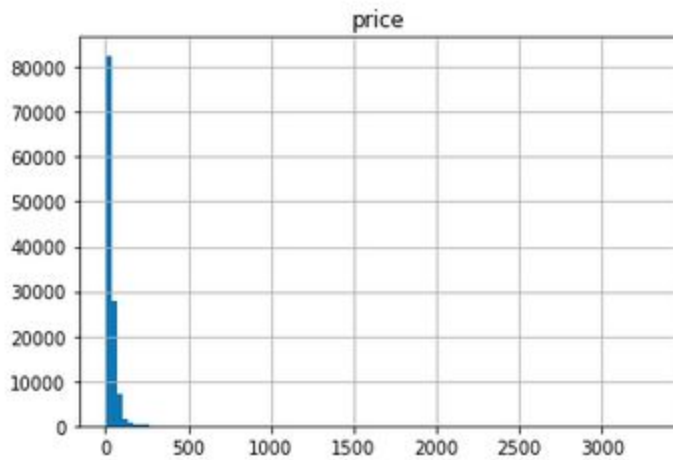
Notes:

- There is missing data in nearly every column
- Possible data points to analyze:
  - Price
  - Points
  - Country
  - Variety

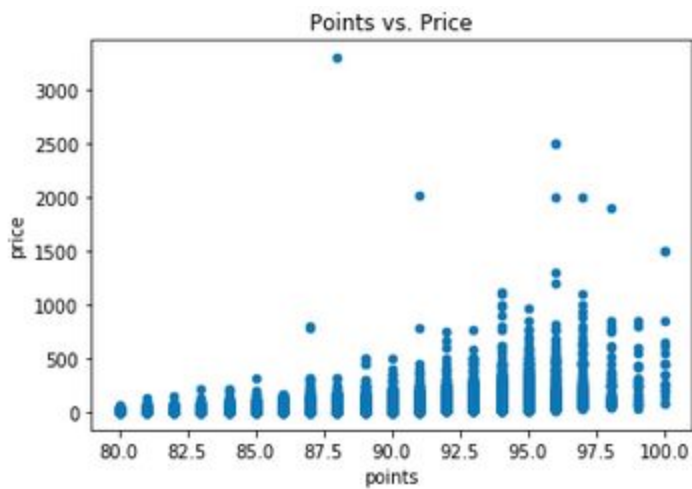
The “points” seem to follow a normal distribution.



Price appears to be logarithmic:



Here is a comparison of price and points (not log-adjusted)





We think there are some interesting comparisons to make with this dataset.

### Potential questions:

#### Variations by Region

- Which countries/regions produced the highest scoring wines?
- What is the distribution of prices by countries/regions?
- Which countries/regions produced particular varieties of wines?
- Which countries/regions had the best value in terms of points/dollar?

#### Price

- How does rating distribution change along price?
- Male vs Female preference for certain wines?

#### Tasting Notes

- What descriptors correspond to each country?
- Are there descriptors that affect price?

How are values (points per dollar) distributed across the data set?

Which wine notes lead to higher points?

In our final report, we will provide answers to these questions along with supporting graphs. We will draw conclusions about which factors seem to affect quality and price the most.