Wine Scores

A report by Jorge Londono

Problem Statement

- With different wineries competing against each other to provide their best ingredients and mixtures, it can often be overwhelming for consumers to get a grasp of what to look for in a bottle of wine
- Luckily, there are expert reviewers and sommeliers who can assign point values based on the quality of a wine
- However, while we can look at specific reviews to figure out which wines perform well for these reviewers, this is a surface level into analyzing what makes wine great.
- What are the best predictors we can find to foresee the quality of a bottle of wine?

Benefactors of this Project

- Wine Producers
- Wine Consumers and enthusiasts
- Wine Industry Media

Data and Data Cleaning

Origin of Data

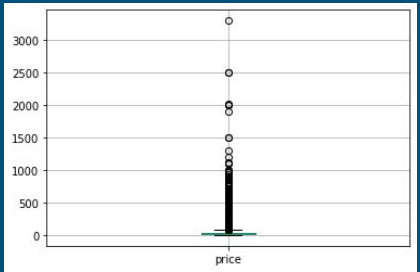
This data was provided by a kaggle user who scraped it from the Wine Enthusiast Magazine in 2017. It contains over 129,000 wine reviews

Missing Values

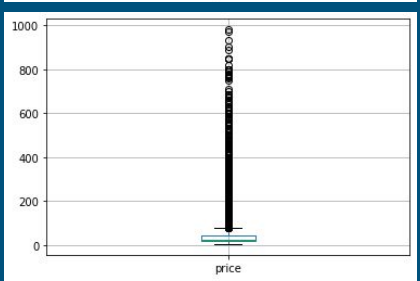
Using the ‘info’ method showed that there were around 9,000 observations where the price variable was missing.

In order to have an accurate predictive model, the observations with these null values were removed using the ‘dropna’ method on the data frame.

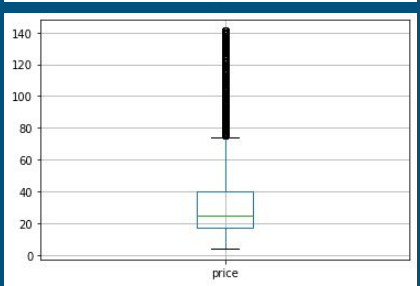
Dealing with Outliers



Initial data

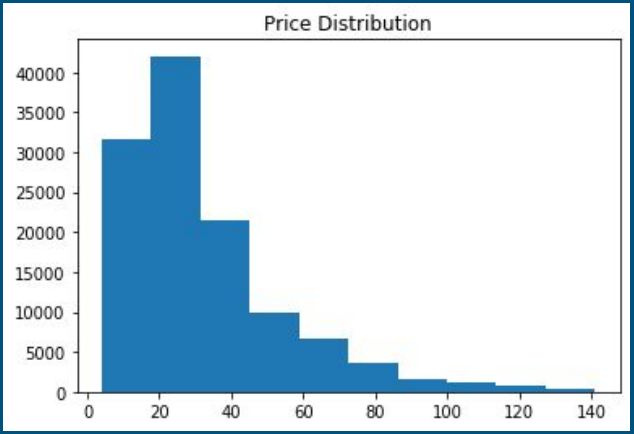


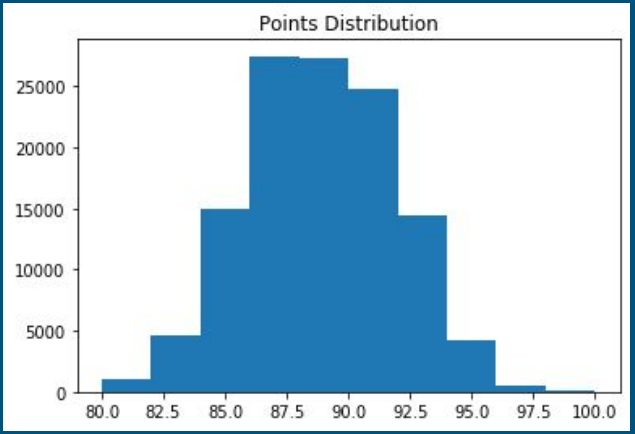
Data with wine priced over $1,000 removed



Data with wine priced with a Z-score over 3

Exploratory Data Analysis



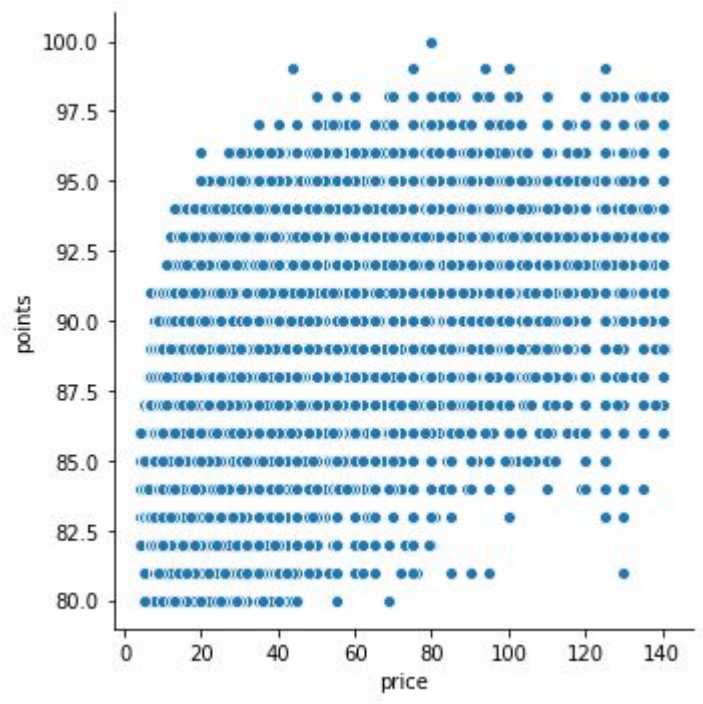


From the graphs above, we can see that the points column of our data frame appears to be normally distributed. However, the price column of our data set seems to be right skewed and might need to be transformed via a logarithmic function.

Comparing Price Groups

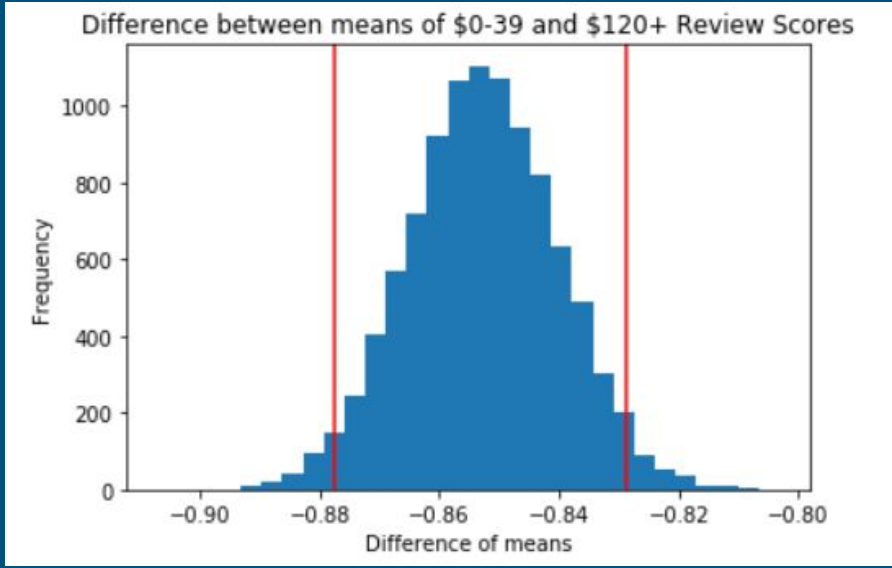
Separating the wines by prices shows us that there is a difference between the most expensive and least expensive wines.

Correlation analysis



- Pearson correlation coefficient: 0.55 between price and review points.
- Moderately positive correlation

Bootstrapping



- Null Hypothesis: wine under priced $40 has the same mean score as wine priced $120 and over
- Alternate Hypothesis: he mean score of wine priced under $40 is not the same as the mean score of wine priced at $120 and over.
- Because this interval does not include 0, at 95% we can reject the null hypothesis that the wine priced under $40 has a mean review score equal to that of wine priced at $120 and over.