Wine Rating Predictor Report

Original features used: Country, price, province, region_1, taster_name, variety

Engineered features:

- **From title: year**
- **From description: description length**
- **From description: compound.**

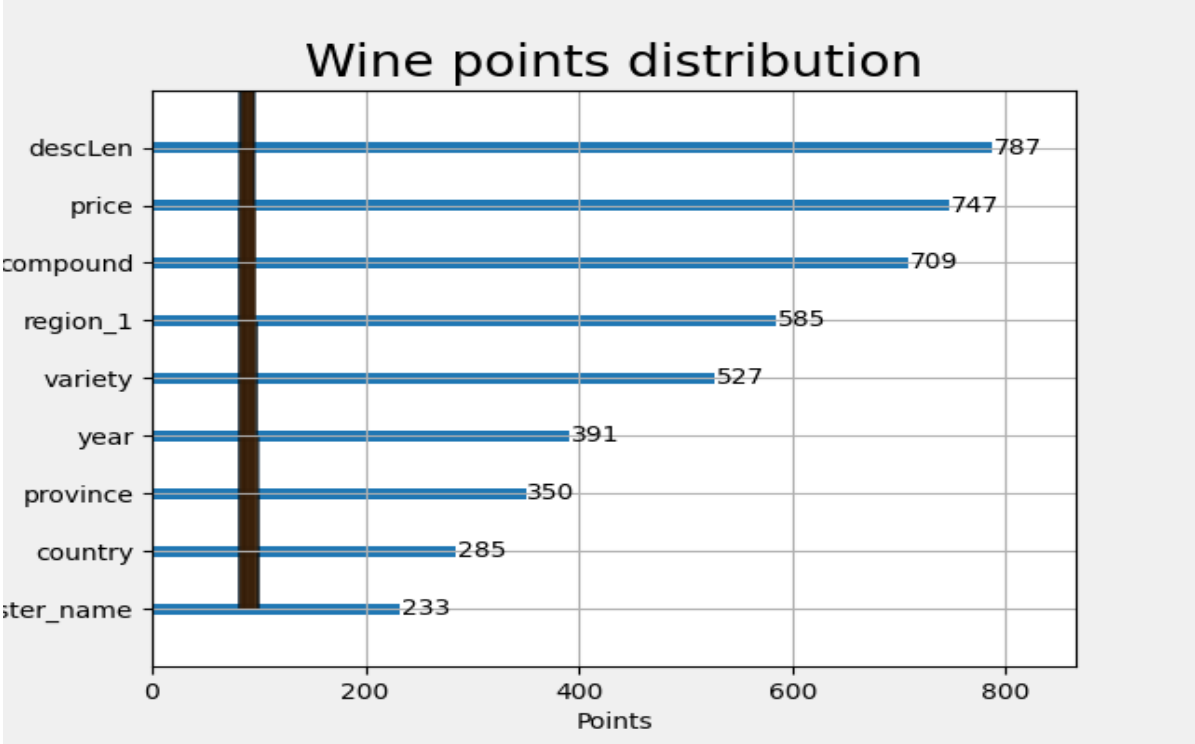
Compound is a score from -1 (extremely negative) to +1 (extremely positive) given by sentiment analysis tool VADER.

Baseline: Linear Regression
Mean Squared Error: 5.00

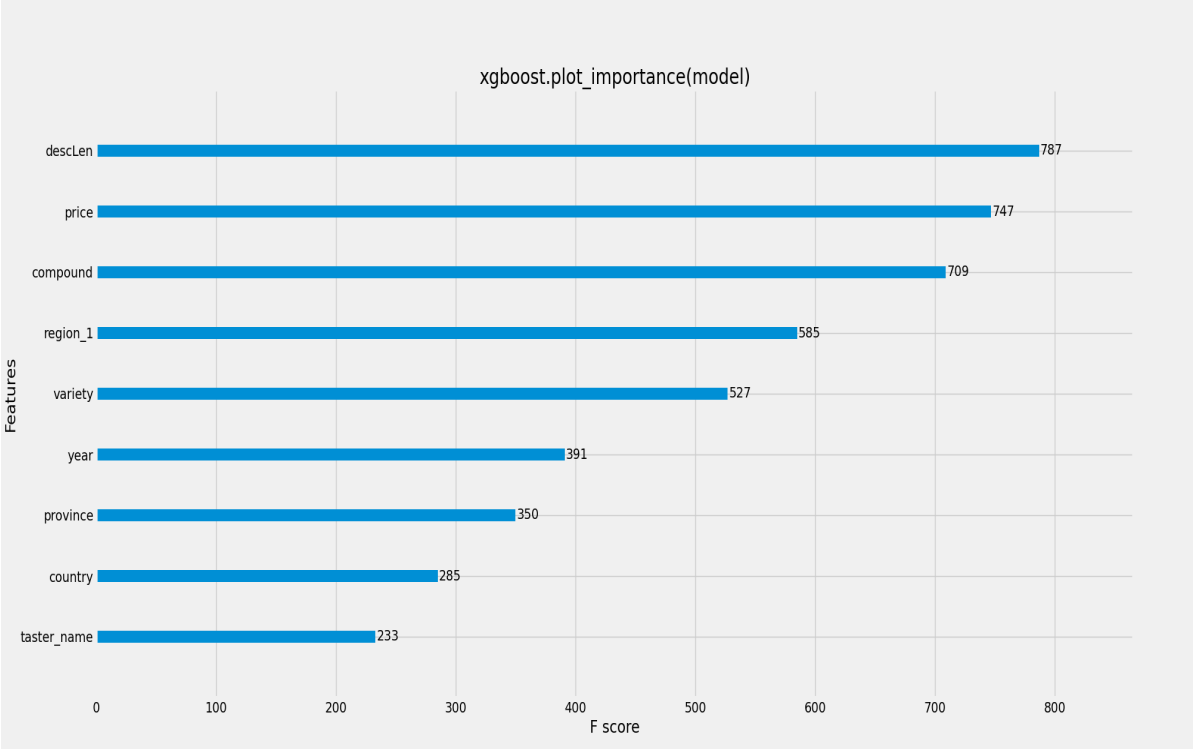
Baseline: Linear Regression
Mean Squared Error: 3.55

Visualizing predictions:
Actual rating of five first wines: [85 85 87 85 92]
Predicted rating of five first wines: [86.43445 85.77748 86.39695 86.17321 91.13946]

Actual x Predicted points distribution:



Most relevant features:



Conclusion:

Using a basic linear regression model as our baseline, we achieved a 5.0 MSE in our regression problem.
We were able to improve this performance by using a XGBoost model with default hyperparameters, which achieved a 3.55 MSE
We believe that this result can be further improved by tuning both the hyperparameter and the VADER sentiment analysis tool to our specific case. More experiments with different types of encoders can also help us achieve higher result.We recommend implementing a full production solution, not only for the promising predictor metrics, but also for its potential to provide important insight to our costumer about their products and comercial partners.
More technical details and further graphical visualization can be found on the attached notebook 'WineRatingPredictor.ipynb'.