Wine Analysis

Julia Gillas, Kiana Talavera, Luis Santamaria,
and Joey Sirko

Project Overview

01. **Goal of Analysis**

Explain what the project entails

02. **Logistic Regression Red Wine**

Logistic Regression based on Price & Ratings

03. **Logistic Regression White Wine**

Logistic Regression based on Price & Ratings

04. **Red Wine Logistic Regression with Age Column**

Looks at 4 different models based on Price

05. **White Wine Logistic Regression with Age Column**

Looks at 4 different models based on Price

06. **Conclusion**

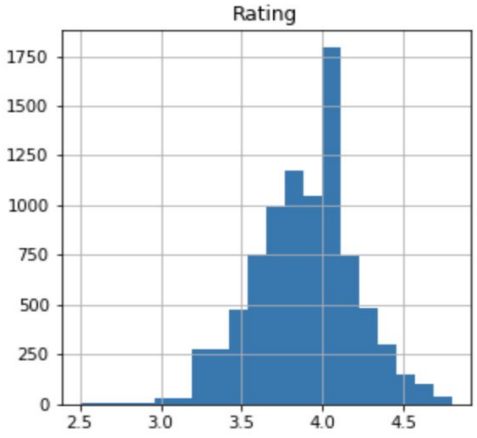
Final thoughts and evaluations on the different models

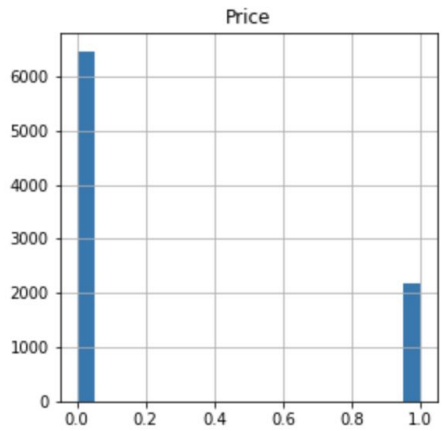
Goal of Analysis

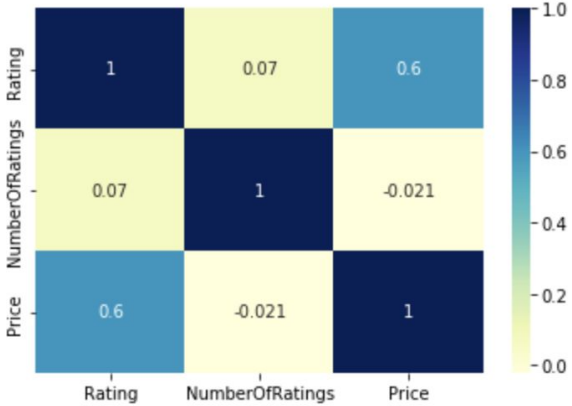
The goal of our analysis is to create different machine learning models to see which one is the best at predicting the best wines. We ran multiple models with the dependent variable varying between Price, Ratings, and Age.

Logistic Regression Red Wine

- Y variable = Price
- X variables = Rating, Number of Ratings, and Year
- Balanced Accuracy Score: 76.09%
- Cheaper Wines:
  - Precision: 87%
  - Recall: 93%
- Expensive Wines:
  - Precision: 74%
  - Recall: 59%
- Accuracy: 84%

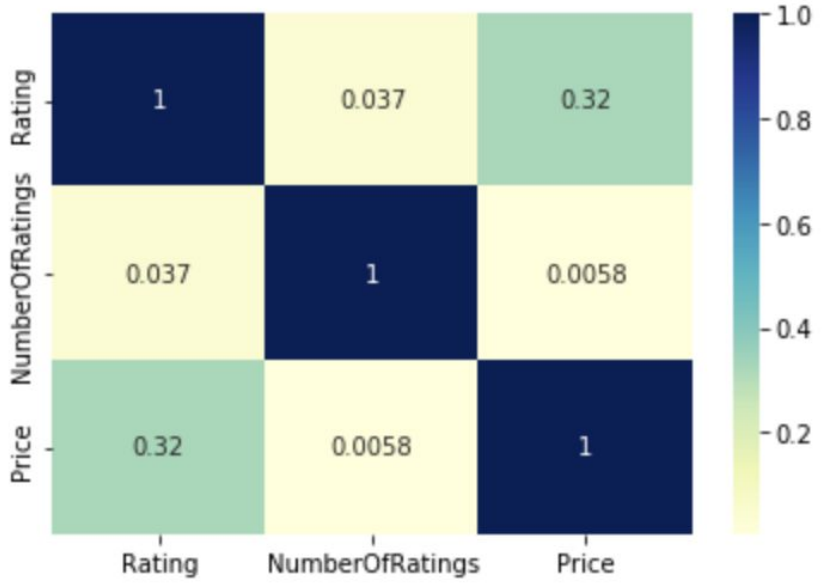






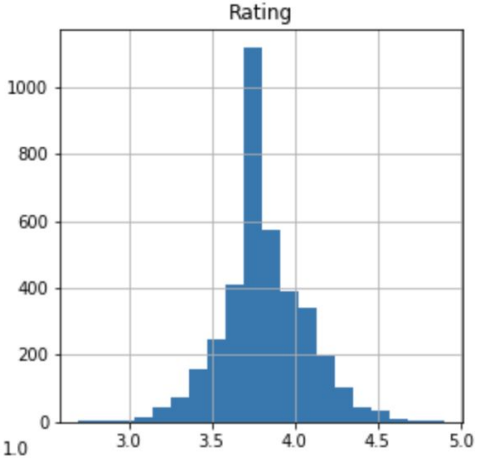
Logistic Regression Red Wine Cont.

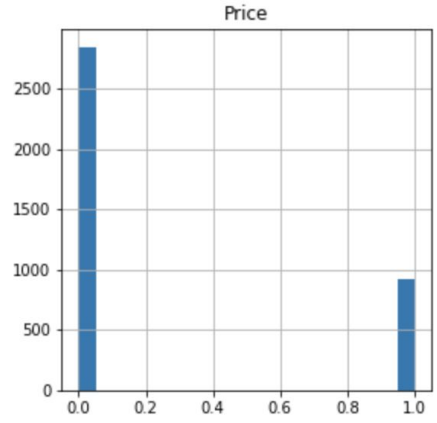
- Y variable = Rating
- X variables = Number of Ratings, Price and Year
- Balanced Accuracy Score: 78.03%
- Higher Rated Wines:
  - Precision: 79%
  - Recall: 90%
- Lower Rated Wines:
  - Precision: 83%
  - Recall: 66%
- Accuracy: 80%

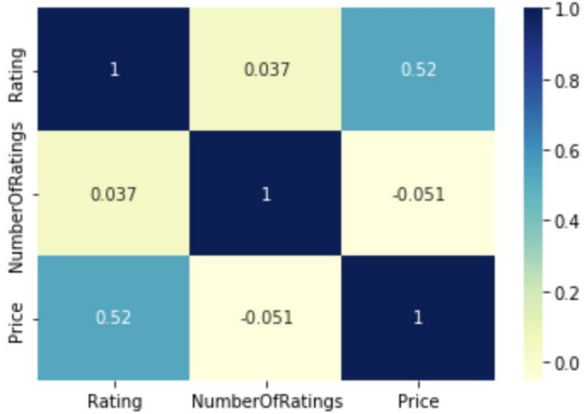


Logistic Regression White Wine

- Y variable = Price
- X variables = Rating, Number of Ratings, and Year
- Balanced Accuracy Score: 74.68%
- Cheaper Wines:
  - Precision: 87%
  - Recall: 95%
- Expensive Wines:
  - Precision: 78%
  - Recall: 54%
- Accuracy: 85%

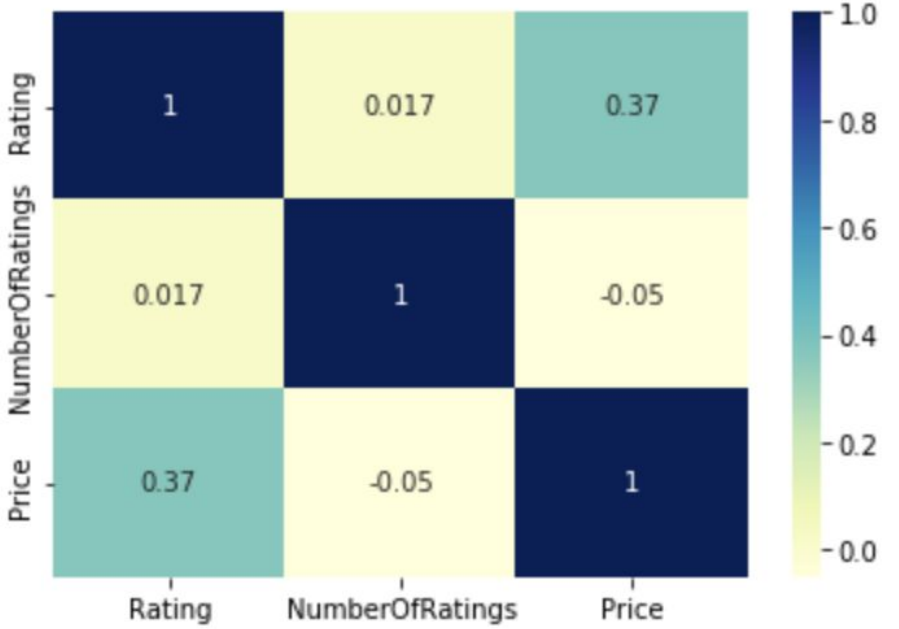






Logistic Regression White Wine Cont.

- Y variable = Rating
- X variables = Number of Ratings, Price and Year
- Balanced Accuracy Score: 71.61%
- Higher Rated Wines:
  - Precision: 81%
  - Recall: 95%
- Lower Rated Wines:
  - Precision: 80%
  - Recall: 48%
- Accuracy: 81%



Red Wine Models w/ Age Column

- Logistic Regression Model
  - Balanced Accuracy: 77.90%
  - Cheaper Wines:
    - Precision: 88%
    - Recall: 95%
  - Expensive Wines:
    - Precision: 79%
    - Recall: 61%
  - Accuracy: 86%

- RandomOverSampler Model
  - Balance Accuracy: 84.03%
  - Cheaper Wines:
    - Precision: 94%
    - Recall: 84%
  - Expensive Wines:
    - Precision: 63%
    - Recall: 85%
  - Accuracy: 84%

- KNN Model
  - Accuracy Score: 77.51%

- SVC Model
  - Accuracy Score: 75.24%

White Wine Models w/ Age Column

- Logistic Regression Model
  - Balanced Accuracy: 80.45%
  - Cheaper Wines:
    - Precision: 90%
    - Recall: 94%
  - Expensive Wines:
    - Precision: 78%
    - Recall: 67%
  - Accuracy: 87%

- RandomOverSampler Model
  - Balanced Accuracy: 85.36%
  - Cheaper Wines:
    - Precision: 95%
    - Recall: 84%
  - Expensive Wines:
    - Precision: 64%
    - Recall: 86%
  - Accuracy: 85%

- KNN Model
  - Accuracy Score: 77.55%

- SVC Model
  - Accuracy Score: 75.64%

Conclusion

In conclusion the models calculated with the age column are the best because creating the age column allows for the machine to have another concrete item to take into account. We have determined that logistic regression model was the best when predicting the price of white and red wines.

The logistic regression model did a much better job at predicting the price over predicting the rating.

THANKS

Do you have any questions?


The
End.