Winery Price Prediction

Luccas Bracci

Business Problem

The business problem we are trying to solve is to predict the price of a wine based on its features, such as the grape variety, region, winery, etc.

Data Source

- The data used for this project was obtained from the Kaggle dataset "Spanish Wine Quality Dataset"

- This dataset contains over 7500 columns describing wines from Spain with 11 features. It describes wines from various locations across Spanish wineries.

Analytical Insights

Average Wine Price by Region

Analytical Insights cont

Wine Price vs Wine Type with Rating distributions
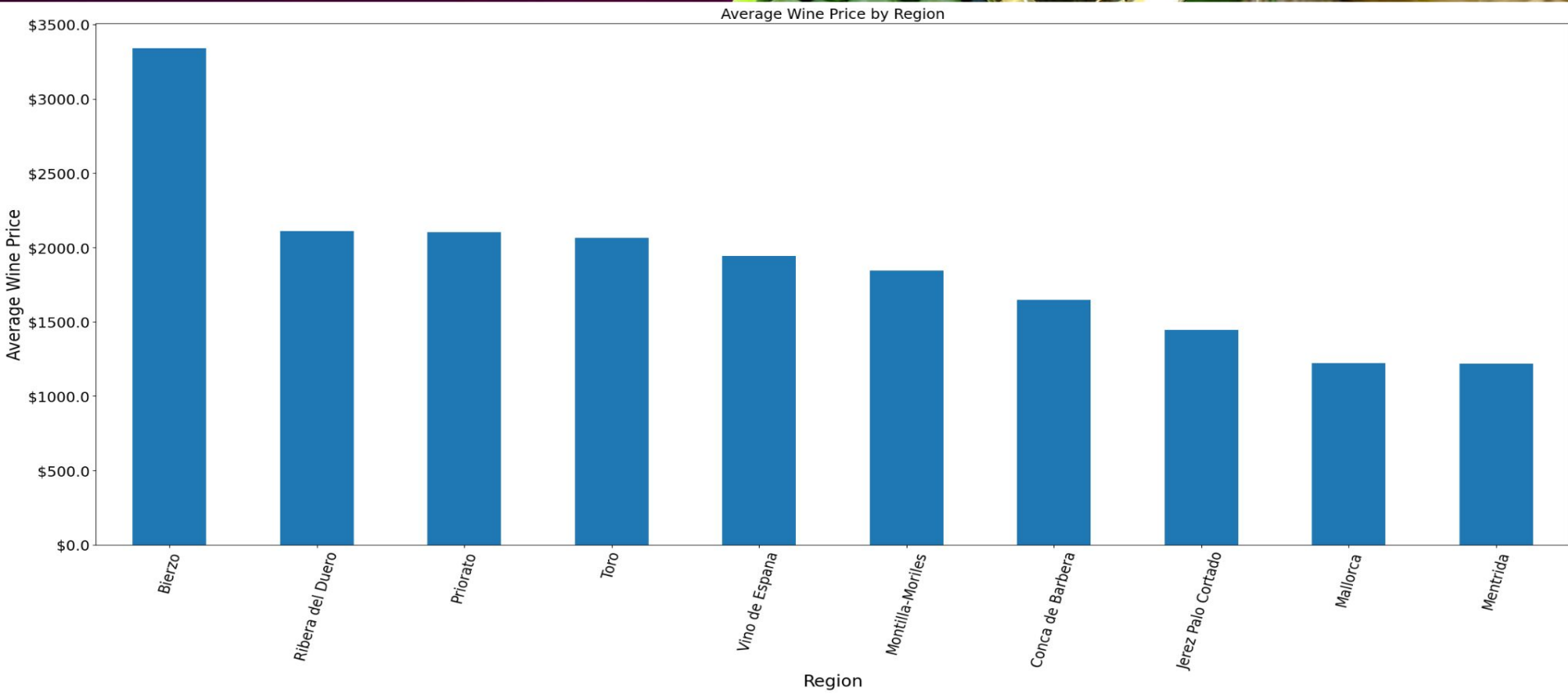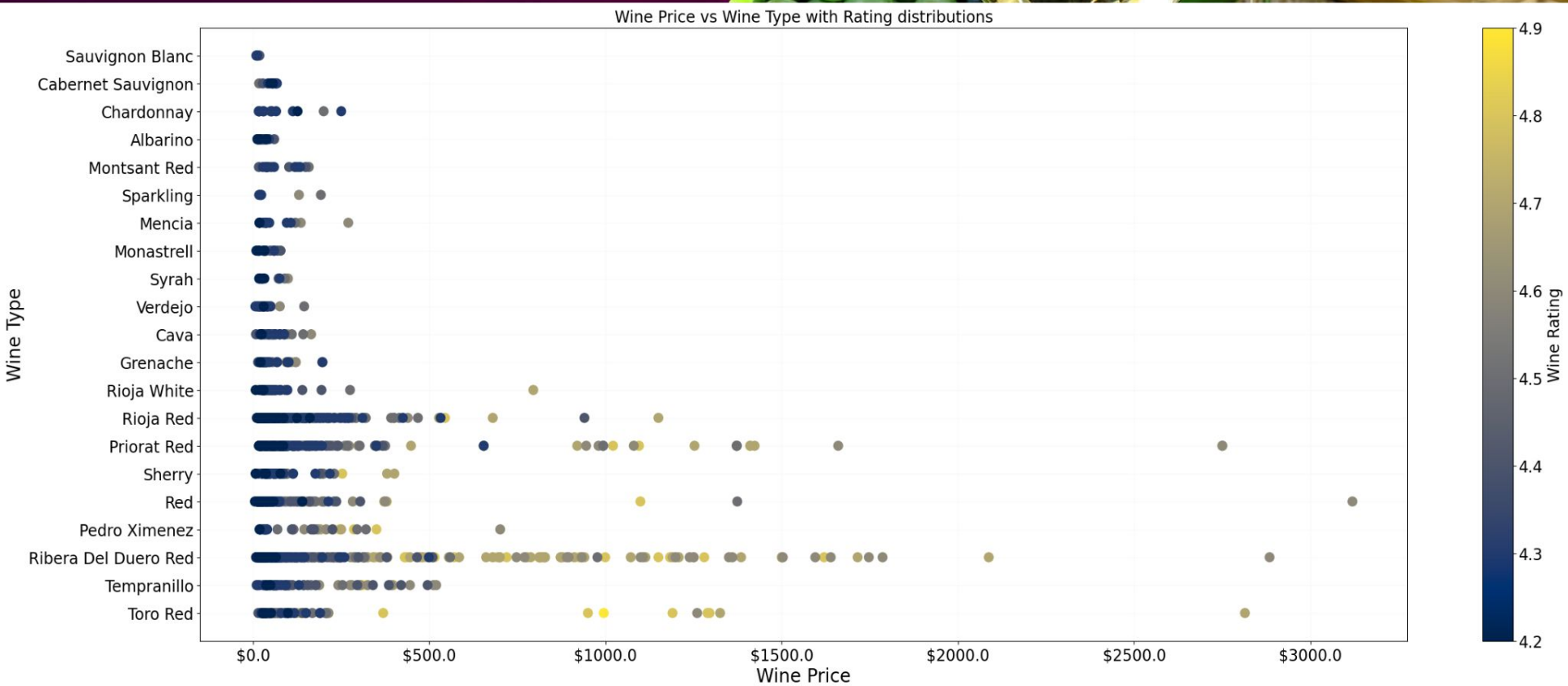
Best Model Metrics

Metrics	Score
Mean Squared Error	19787.45
R^2 Score	0.77
Explained Variance Score	0.77

Our model is able to predict the price of a wine with an MSE of $140. This means that, on average, our model's predictions are off by 140 dollars from the actual wine prices. This level of accuracy can be useful for wine producers and retailers to price their products competitively and maximize their revenue when investing in higher selling wine and wineries.

Model Performance

For our best model, we will use mean squared error (MSE) as the evaluation metric. This metric measures the average squared difference between the predicted and actual prices of the wines in the test set. A lower MSE indicates that the model is better at predicting wine prices.

Recommendations

It is recommended this model be used to predict prices of wine that are known to have a higher price range. The model assumes linear relationship between the features and target data and cannot generalize unseen data. For modeling with new datasets it is recommended to recursively train the model using new data with higher unique values.

Conclusion

- Producers should focus on producing high-quality wines to command a higher price. Our analysis showed that the price of wines is strongly correlated with the number of points the wine received in its review.
- Retailers should focus on promoting the most popular wine varieties, such as Ribera Del Duero Red, Priorat Red, Rioja Red, and Toro Red. These four varieties account for almost half of the wines in our dataset and are likely to be popular among customers.