Spanish Wine Quality

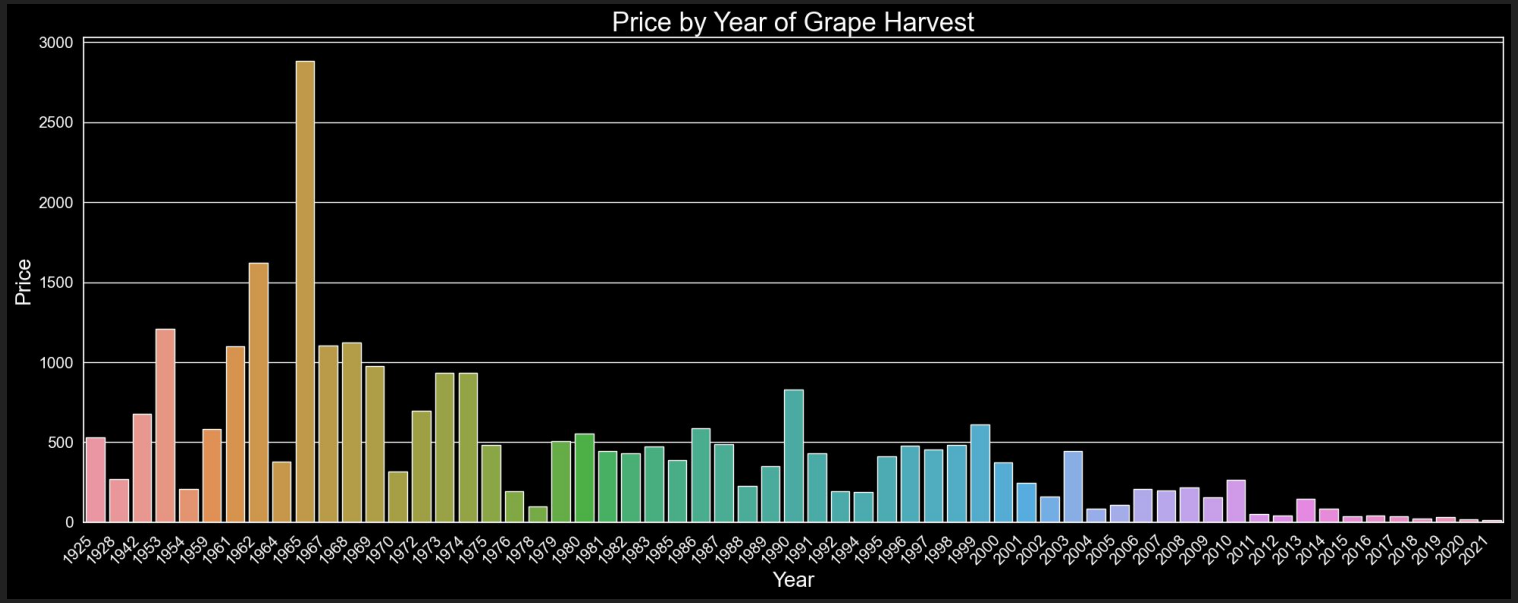
Stakeholders + Data Description

Stakeholders

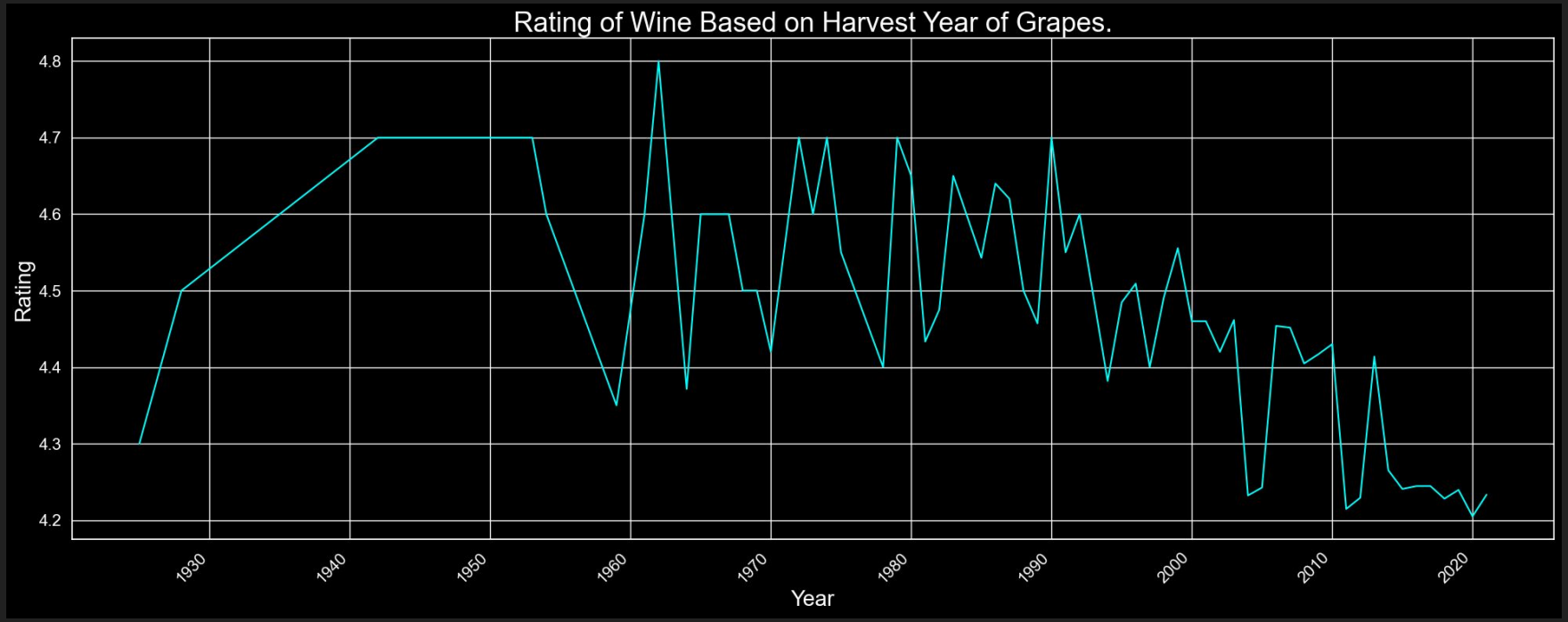
- Wineries
- Wine Customers
- Wine Reviewers + Wine Critics
- Wine Distributors
- Wine Enthusiast Communities

Description

- A list of 7000+ wines that describes their price, rating, and even flavor description.
- This data has many useful features and can be used for a wide variety of applications.



This bar plot illustrates that after the year 1969, there is a significant decline in the sale of wines priced over $1000. This information is valuable as it suggests a potential issue with grape quality. Additionally, the bar plot demonstrates that the average price of wine is undergoing significant changes.



This line graph shows a very strong decline after 1990. This further indicates the possible decline in grape quality.

Strengths and Limitations

Strengths

- High R2 Score: A high R2 score indicates that the model can explain a significant amount of the variance in the data. This also signifies its ability to interpret a substantial number of relationships between features and targets.
- The model also has low mean absolute error, mean squared error, and root mean squared error. These lower numbers indicate the model has good accuracy in predicting the target.

Limitations

- Lack of interpretability: Random Forest models, are not easy to interpret. They can give accurate predictions. Understanding the reasons behind them can be hard.

Final Recommendations.

XGBoost, as an alternative to all the other models this dataset was run on, led to improved performance and better handling of the dataset.

REGRESSION METRICS FOR: Test Data XGBoost Tuned

- R2 Score: 0.9467
- Mean Absolute Error: 1.868367
- Mean Squared Error: 17.491149
- Root Mean Squared Error: 4.182242
- Explained Variance Score: 0.954208
- Max Error: 81.715500

REGRESSION METRICS FOR: Training Data XGBoost Tuend

- R2 Score: 0.9857
- Mean Absolute Error: 1.638309
- Mean Squared Error: 5.274667
- Root Mean Squared Error: 2.296664
- Explained Variance Score: 0.992968
- Max Error: 60.851318