ANALYTICS REPORT

TO: E. AND J. GALLO WINERY

FROM: JAKE MOORE

SUBJECT: WINE DATA FOLLOW-UP ANALYSIS

DATE: MAY 2, 2025

Introduction

In this follow-up wine analysis, we analyzed a dataset provided by the University of Arizona, with information on wines from around the world, with 13,050 different entries detailing wine name, country, region, winery, type, price, and rating.

The main goal of this analysis was to explore how well wine rating, year produced, and number of ratings could be used to predict price. We analyzed whether these relationships were strong enough to be used as an accurate predictor for business use.

To determine the relationships, we used correlation analysis, simple and multiple regression models, and created a Tableau dashboard to visualize the patterns.

We found that all three predictor variables have a statistically significant impact on price, with rating having the strongest positive effect, especially with sparkling wines. However, the overall model only explained 26.5% of the variation in wine prices.

One important recommendation from this analysis is that Gallo should not rely on this regression model for pricing decisions. Our recommendation is that they explore additional variables like in person sales and online sales to improve pricing accuracy.

Data Analysis

Correlations and Scatterplots

The following table describes the correlation and relationship shape for all wines overall and the four wine types. All correlations between price and rating are positive, ranging from moderate to strong, depending on the different wine types. The relationship shape for each wine type appears to be non-linear and curved upward. This indicates that higher ratings are generally associated with disproportionately higher prices, and this data seems to be more exponential.

Wine Type	Correlation	Interpretation	Relationship Shape
Red	0.457	Positive and moderate	Non-linear, curved upward
Rosé	0.434	Positive and moderate	Non-linear, curved upward
Sparkling	0.730	Positive and strong	Non-linear, curved upward
White	0.464	Positive and moderate	Non-linear, curved upward
Overall	0.455	Positive and moderate	Non-linear, curved upward

Table 1: Correlation, interpretation, and scatterplot shape for all wines and wine types.

Regression Results

Single-predictor Models
For all wines, as rating increases by 1 star, price increases by $107.15 on average.
For red wines, as rating increases by 1 star, price increases by $121.72 on average.
For rosé wines, as rating increases by 1 star, price increases by $25.57 on average.
For sparkling wines, as rating increases by 1 star, price increases by $190.44 on average.
For white wines, as rating increases by 1 star, price increases by $53.66 on average.

There appears to be a relationship between correlation strength and size of the correlation coefficient on rating across wine types. Sparkling wine which has the strongest correlation between price and rating, also appears to show the largest increase in price per rating star, while rosé has the weakest correlation, and the smallest rating effect.

Multiple Regression Equation

$$\widehat{Price} = 10948.86 - 0.0022 * NumberofRatings - 5.58 * Year + 86.92 * Rating$$

R^2 Interpretation
26.5% of variance in price is explained by the model with number of ratings, year, and rating.

Rating
H_0: Rating does not significantly impact price.
H_A: Rating significantly impacts price.

Reject the null hypothesis, because the P-value of less than 0.001 is less than the 0.05 significance level. Rating significantly impacts wine price.
As rating increases by 1 star, price increases by $86.92, on average and all else constant.

Number of Ratings
H_0: Number of ratings does not significantly impact price.
H_A: Number of ratings significantly impacts price.

Reject the null hypothesis, because the P-value of less than 0.001 is less than the 0.05 significance level. Number of ratings significantly impacts wine prices.
As number of ratings increases by 1, price decreases by $0.0022, on average and all else constant.

Year Produced
H_0: Year does not significantly impact price.
H_A: Year significantly impacts price.

Reject the null hypothesis, because the P-value of less than 0.001 is less than the 0.05 significance level. Year significantly impacts wine price.
As year increases by 1, price decreases by $5.58, on average and all else constant.

Analysis Takeaways and Recommendations

The Gallo company should not trust price predictions from this model, because the model only explains 26.5% of the variation in price, and the estimate of uncertainty is over $60 per wine. Gallo company could consider adding the age of the wine, such as the difference between the current year and the production year. This would be helpful as older wine is typically more valuable. In order to provide a set of diversely priced wines, Gallo should choose some inexpensive and expensive wines. Inexpensive wines have relatively low ratings, recent years and high numbers of ratings. Expensive wines have higher ratings, older production years, and fewer ratings.

Tableau Dashboard

Click here to see a visualization in Tableau. On the dashboard, average ratings and prices are displayed by wine type. You can use the filters and actions to identify specific countries, years, and wine types.

One visible insight in the dashboard is that sparkling wines tend to have both the highest average rating and price, and this supports our original regression models where we found that price is especially sensitive to ratings for sparkling wines.

A recommendation for further analysis would be to explore deeper into how the wine is sold, either through online sales or in-person sales. It is possible that a consumer might be willing to pay more for a certain wine in a specialty store than an online buyer. Understanding these differences could help Gallo company understand where to place their products to maximize their profits.

Conclusion

This analysis confirmed that wine rating, year, and number of ratings all have a statistically significant impact on price. However, the multiple regression model explained only 26.5% of the variation in wine prices, indicating that other important factors could be missing from this dataset to optimize the model.

Based on these findings, we recommend that Gallo does not rely on this model alone for predicting prices for business decisions as the level of uncertainty is too high. We recommend that Gallo company continue to collect additional data such as in person sales and online sales. We also recommend that Gallo company collects additional data on the wine age or years since production as older wine typically sells for more.

Although the model is not highly trustworthy for prediction, the significant coefficients indicate that rating, year, and number of ratings have a statistically significant impact on the price.

If you have any questions, don't hesitate to get in touch with Jake Moore at jakemoore@arizona.edu

Thank you,

Jake Moore

Appendix

Regression Statistics	
Multiple R	0.455227019
R Square	0.207231639
Adjusted R Square	0.207170881
Standard Error	62.60813861
Observations	13050

ANOVA

	df	SS	MS	F	Significance F
Regression	1	13369503.66	13369503.66	3410.779943	0
Residual	13048	51145276.66	3919.779021		
Total	13049	64514780.33			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-381.371041	7.11847088	-53.57485441	0	-395.3242819	-367.4178002	-395.3242819	-367.4178002
Rating	107.145557	1.834625039	58.40188304	0	103.5494244	110.7416896	103.5494244	110.7416896

Jake Moore

Figure 1: Single-predictor regression results for all wines.

Regression Statistics	
Multiple R	0.457128985
R Square	0.208966909
Adjusted R Square	0.208875322
Standard Error	72.98419849
Observations	8639

ANOVA

	df	SS	MS	F	Significance F
Regression	1	12153558.97	12153558.97	2281.632984	0
Residual	8637	46006649.42	5326.693229		
Total	8638	58160208.39			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-434.8071822	9.943130756	-43.72940404	0	-454.2980917	-415.3162726	-454.2980917	-415.3162726
Rating	121.7180178	2.548191002	47.76644203	0	116.7229552	126.7130804	116.7229552	126.7130804

Jake Moore

Figure 2: Single-predictor regression results for red wines.

Regression Statistics	
Multiple R	0.433751468
R Square	0.188140336
Adjusted R Square	0.186037073
Standard Error	14.61138853
Observations	388

ANOVA

	df	SS	MS	F	Significance F
Regression	1	19097.26763	19097.26763	89.45162943	3.13507E-19
Residual	386	82408.17246	213.4926748		
Total	387	101505.4401			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-83.05275845	10.144915	-8.186639163	3.95908E-15	-102.9989675	-63.10654937	-102.9989675	-63.10654937
Rating	25.57	2.703643779	9.457887155	3.13507E-19	20.255046	30.88646954	20.255046	30.88646954

Jake Moore

Figure 3: Single-predictor regression results for rosé wines.

Regression Statistics	
Multiple R	0.729726079
R Square	0.53250015
Adjusted R Square	0.530768669
Standard Error	51.07203894
Observations	272

ANOVA

	df	SS	MS	F	Significance F
Regression	1	802173.6925	802173.6925	307.5402918	1.74504E-46
Residual	270	704255.3536	2608.353162		
Total	271	1506429.046			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-713.5355141	44.25299454	-16.1240052	1.98407E-41	-800.660324	-626.4107043	-800.660324	-626.4107043
Rating	190.44	10.85946226	17.53682673	1.74504E-46	169.060518	211.8204982	169.060518	211.8204982

Jake Moore

Figure 4: Single-predictor regression results for sparkling wines.

Regression Statistics	
Multiple R	0.463813415
R Square	0.215122884
Adjusted R Square	0.214913527
Standard Error	27.13645853
Observations	3751

ANOVA

	df	SS	MS	F	Significance F
Regression	1	756670.3605	756670.3605	1027.543898	1.7944E-199
Residual	3749	2760716.294	736.3873816		
Total	3750	3517386.654			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-184.3580761	6.406593065	-28.776305	1.0418E-164	-196.918823	-171.7973292	-196.918823	-171.7973292
Rating	53.66	1.673841204	32.05532558	1.7944E-199	50.37379678	56.93725273	50.37379678	56.93725273

Jake Moore

Figure 5: Single-predictor regression results for white wines.

Regression Statistics	
Multiple R	0.515092895
R Square	0.265320691
Adjusted R Square	0.265151747
Standard Error	60.2753622
Observations	13050

ANOVA

	df	SS	MS	F	Significance F
Regression	3	17117106.1	5705702.032	1570.469225	0
Residual	13046	47397674.23	3633.119288		
Total	13049	64514780.33			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	10948.86087	353.2100861	30.99815466	1.2542E-203	10256.51759	11641.20416	10256.51759	11641.20416
NumberOfRatings	-0.002225118	0.00072086	-3.086757103	0.00202777	-0.003638108	-0.000812129	-0.003638108	-0.000812129
Year	-5.582241519	0.173980727	-32.08540178	1.7639E-217	-5.923269116	-5.241213921	-5.923269116	-5.241213921
Rating	86.91654793	1.884163897	46.13003575	0	83.22331191	90.60978396	83.22331191	90.60978396

Jake Moore

Figure 6: Multiple-predictor regression results.