PRELIMINARY WINE ANALYSIS

TO: E. and J. Gallo Winery

FROM: Jake Moore

SUBJECT: Wine Data Preliminary Analysis

DATE :   March 26, 2025

Introduction

In this preliminary 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 investigate different relationships between wine prices and ratings and compare and contrast different regions such as South America and the Southern Hemisphere. Finally, the goal of this preliminary wine analysis was to provide different business insights and suggestions for Gallo Wine Company.

The analysis included descriptive statistics, comparisons of wine types with their average prices and ratings, hypothesis testing using t-tests, box-and-whisker plots, and visual data analytics using Tableau. We specifically examined how average price and rating differed between wine types and regions, such as South America and the Southern Hemisphere.

We found that Sparkling wines have both the highest average rating as well as the highest average price. In addition, we also found that Rosé has the lowest average rating and the lowest average price. We did find a statistically significant difference in average price between South America and the Southern Hemisphere.

We also recommend that Gallo Company purchase less wine from countries with lower average wine ratings- such as China, which had the lowest average rating at 3.20. More specifically, the purchasing of wines that already perform poorly in customer ratings, like Rosé and White wine, should be reduced greatly. We believe that this focus on quality control and country selection can help the Gallo company increase its overall ratings and help its brand reputation.

Data and Analysis

This dataset was provided by a distributor as a source of potential wines of Gallo stores. There are 13,050 rows, each including information about the wine's name, country, region, winery, type, price, and rating.

Section 1 – Discussion of Price and Rating

Wine Type	Count of Wines	Average Price	Average Rating
Red	8639	$38.66	3.89
Rosé	388	$12.64	3.74
Sparkling	272	$60.62	4.07
White	3751	$20.52	3.82
Overall	13050	$33.13	3.87

Table 1: Wine type counts, average prices, and average ratings.

The largest difference in average price was between Sparkling and Rosé wines. The smallest difference in average price was between Rosé and White wines. The largest difference in average rating was between Sparkling and Rosé wines. The smallest difference in average rating was between White and Rosé wines.

For all wines, skewness of price was 15.08, indicating that the shape is right-tailed. There is a price extreme outlier, Pomerol 2012. Skewness of rating was -0.002, indicating that the shape was roughly symmetric. There are no extreme outliers for rating.

For Red wine, skewness of price was 13.85, indicating that the shape is right-tailed. There is a price extreme outlier, Pomerol 2012. Skewness of rating was -0.057, indicating that the shape was roughly symmetric. There are no extreme outliers for rating.

For Rosé wine, skewness of price was 9.55, indicating that the shape is right-tailed. There is a price extreme outlier, Clos du Temple 2018. Skewness of rating was 0.092, indicating that the shape was roughly symmetric. There are no extreme outliers for rating.

For Sparkling wine, skewness of price was 2.73, indicating that the shape is right-tailed. There are no extreme outliers for pricing. Skewness of rating was 0.118, indicating that the shape was roughly symmetric. There are no extreme outliers for rating.

 For White wine, skewness of price was 9.92, indicating that the shape is right-tailed. There are no extreme outliers for pricing. Skewness of rating was -0.050, indicating that the shape was roughly symmetric. There are no extreme outliers for rating.

Section 2 – Comparing South America vs Southern Hemisphere Average Wine Rating and Price

To compare prices between South America and Southern Hemisphere countries, the null hypothesis is that there is no difference between average wine prices in South American countries compared to Southern Hemisphere countries. The alternative hypothesis is that there is a difference in average wine price in South American countries compared to Southern Hemisphere countries. The t-test statistic was t = -2.17 with a P-value of 0.030. The null hypothesis is rejected because the P-value is less than the 0.05 significance level. Based on the data, there is a statistically significant difference in average wine price between South American countries and Southern Hemisphere countries.

To compare ratings between South America and Southern Hemisphere countries, the null hypothesis is that there is no difference between average ratings in South American countries compared to Southern Hemisphere countries. The alternative hypothesis is that there is a difference in average ratings in South American countries compared to Southern Hemisphere countries. The t-test statistic was t = -4.02 with a P-value less than 0.001. The null hypothesis is rejected because the P-value is less than the 0.05 significance level. Based on the data, there is a statistically significant difference in average ratings between South American countries and Southern Hemisphere countries.

Following the realization that there is a significant difference in average wine prices and ratings between South America and the Southern Hemisphere countries, we suggested zooming in the scope to two individual countries, being Argentina and South Africa. This zoomed-in approach helps Gallo company understand why there is a difference between the two parts of the

world by comparing and contrasting. We recommend that they continue to test and compare individual countries to understand why this statistical difference exists.

To compare prices between Argentina and South Africa, the null hypothesis is that there is no difference between average prices in Argentina compared to South Africa. The alternative hypothesis is that there is a difference in average prices in Argentina compared to South Africa. The test statistics were t = 1.65 with a P-value of 0.10. The null hypothesis is not rejected because the P-value is more than the 0.05 significance level. Based on the data, there is not a statistically significant difference in average prices between Argentina and South Africa.

Section 3 – Tableau Average Rating by Type and Map of Average Rating by Country

For a Tableau visualization, click here. There are two graphics – average rating by wine type and average rating by country. On the graph of countries, you can use the filter to select countries to view. The country with the highest average rating was Moldova at 4.17, while the country with the lowest average rating was China at 3.20.

Given that the highest average rating comes from Moldova, we suggest that Gallo Company purchases more wine from Moldova- specifically wines that tend to be lower rated overall, like Rosé and White. We believe that this could help elevate the general opinion of the consumer on these wine types by providing higher-quality examples from a top-rated country.

We also recommend that Gallo Company purchase less wine from countries with lower average wine ratings- such as China, which had the lowest average rating at 3.20. More specifically, the purchasing of wines that already perform poorly in customer ratings, like Rosé and White wine, should be reduced greatly. We believe that this focus on quality control and country selection can help the Gallo company increase its overall ratings and help its brand reputation.

Conclusion

In this preliminary wine analysis, we examined a dataset of 13,050 wine entries to uncover relationships between wine type, price, and rating. We used descriptive statistics, t-tests, and visualizations in Tableau to identify key trends and differences across different wine types and regions around the world. Our analysis showed that Sparkling wines have the highest average price and rating, while Rosé wines have the lowest average price and rating. Ratings were generally consistent among different wine types, while prices were much more skewed, with a few outliers that brought up the average price

Hypothesis tests showed that there was a significant difference in average price between South America and Southern Hemisphere countries. The Tableau visualization supported these findings by showing different variations in average ratings and prices across countries, including Moldova as the highest rated and China as the lowest rated. We then compared Argentina and South Africa to further explore country-level insights, given that there was a significant difference in average wine prices and ratings between South America and Southern Hemisphere countries. However, we did not find any statistically significant difference in their average prices. This comparison leads to our recommendation that the Gallo Company should dig deeper into individual country level performance and look outside solely regional trends. We also recommend that Gallo Company continues comparing individual countries within key regions to better understand where high-quality wines come from and why those differences exist. Given

Moldova's strong average rating, we concluded that it would be best to continue purchasing more Moldovan wines, especially in the categories that tend to score lower- like Rosé and White wines to possibly raise consumer perception of these wine types. On the other hand, we recommend purchasing fewer wines from countries with consistently lower ratings- like China. We also recommend staying away from purchasing low-performing categories like Rosé and White wine from those regions as well. These recommendations can help Gallo improve its overall wine ratings and strengthen its brand reputation through better selection and quality control.

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

Thank you,
Jake Moore

Appendix

Price		Rating	
Mean	33.1270881	**Mean**	3.8685517
Standard Error	0.61551089	**Standard Error**	0.0026151
Median	15.95	**Median**	3.9
Mode	9.9	**Mode**	3.8
Standard Deviation	70.3138691	**Standard Deviation**	0.298741
Sample Variance	4944.04018	**Sample Variance**	0.0892462
Kurtosis	487.517911	**Kurtosis**	0.0981507
Skewness	15.084846	**Skewness**	-0.0021981
Range	3407.24	**Range**	2.4
Minimum	3.55	**Minimum**	2.5
Maximum	3410.79	**Maximum**	4.9
Sum	432308.5	**Sum**	50484.6
Count	13050	**Count**	13050
Jake Moore			

Figure 1: Summary statistics of all wines.

Price		Rating	
Mean	38.65745	Mean	3.8898
Standard Error	0.882825	Standard Error	0.0033
Median	18.14	Median	3.9
Mode	9.9	Mode	3.8
Standard Deviation	82.05525	Standard Deviation	0.3082
Sample Variance	6733.064	Sample Variance	0.095
Kurtosis	391.3636	Kurtosis	0.0036
Skewness	13.85057	Skewness	-0.0571
Range	3407.24	Range	2.3
Minimum	3.55	Minimum	2.5
Maximum	3410.79	Maximum	4.8
Sum	333961.7	Sum	33604
Count	8639	Count	8639
Jake Moore			

Figure 2: Summary statistics for Red Wine.

Price		Rating	
Mean	12.64	Mean	3.7423
Standard Error	0.8222	Standard Error	0.0139
Median	8.945	Median	3.75
Mode	7.9	Mode	3.8
Standard Deviation	16.195	Standard Deviation	0.2747
Sample Variance	262.29	Sample Variance	0.0755
Kurtosis	123.16	Kurtosis	0.6894
Skewness	9.5503	Skewness	0.0918
Range	245.3	Range	2.1
Minimum	3.7	Minimum	2.7
Maximum	249	Maximum	4.8
Sum	4904.3	Sum	1452
Count	388	Count	388
Jake Moore			

Figure 3: Summary statistics for Rosé Wine.

Price		Rating	
Mean	60.619	Mean	4.065074
Standard Error	4.5207	Standard Error	0.017322
Median	35.135	Median	4.1
Mode	14.9	Mode	3.9
Standard Deviation	74.557	Standard Deviation	0.285687
Sample Variance	5558.8	Sample Variance	0.081617
Kurtosis	9.0016	Kurtosis	-0.58074
Skewness	2.727	Skewness	0.117996
Range	488.5	Range	1.4
Minimum	6.5	Minimum	3.3
Maximum	495	Maximum	4.7
Sum	16488	Sum	1105.7
Count	272	Count	272
Jake Moore			

Figure 4: Summary statistics for Sparkling Wine.

Price		Rating	
Mean	20.5156	Mean	3.8183
Standard Error	0.50006	Standard Error	0.0043
Median	13.15	Median	3.8
Mode	9.9	Mode	3.8
Standard Deviation	30.6263	Standard Deviation	0.2647
Sample Variance	937.97	Sample Variance	0.0701
Kurtosis	148.088	Kurtosis	0.3296
Skewness	9.92313	Skewness	-0.0498
Range	677.63	Range	2.2
Minimum	3.74	Minimum	2.7
Maximum	681.37	Maximum	4.9
Sum	76954.1	Sum	14323
Count	3751	Count	3751
Jake Moore			

Figure 5: Summary statistics for White Wine.

Hypothesis Test for Prices		
t-Test: Two-Sample Assuming Unequal Variances		

	S. America Price	*S. Hemisphere Price*
Mean	19.7352986	22.91655199
Variance	599.6050053	1814.947154
Observations	787	1308
Hypothesized Mean Difference	0	
df	2089	
t Stat	-2.16986871	
P(T<=t) one-tail	0.015064586	
t Critical one-tail	1.645583378	
P(T<=t) two-tail	0.030129172	
t Critical two-tail	1.961100233	
Jake Moore		

Figure 6: Hypothesis test comparing prices.

Hypothesis Test for Ratings		
t-Test: Two-Sample Assuming Unequal Variances		

	S. America Rating	*S. Hemisphere Rating*
Mean	3.7463787	3.805122324
Variance	0.1137623	0.090876573
Observations	787.0000000	1308
Hypothesized Mean Difference	0.0000000	
df	1513.0000000	
t Stat	-4.0153592	
P(T<=t) one-tail	0.0000311	
t Critical one-tail	1.6458614	
P(T<=t) two-tail	0.0000623	
t Critical two-tail	1.9615331	
Jake Moore		

Figure 7: Hypothesis test comparing ratings.

Hypothesis Test for Prices		
t-Test: Two-Sample Assuming Unequal Variances		

	Argentina Price	South Africa Price
Mean	20.96078176	18.28871394
Variance	623.6620654	487.4613619
Observations	307	832
Hypothesized Mean Difference	0	
df	493	
t Stat	1.651640213	
P(T<=t) one-tail	0.04962234	
t Critical one-tail	1.647950288	
P(T<=t) two-tail	0.09924468	
t Critical two-tail	1.964787528	
Jake Moore		

Figure 8: Hypothesis test comparing prices between Argentina and South Africa.