

Time Impact to Sentiment Analysis

Leon King

August 2025

1 Introduction

We aim to investigate how the ratings of wine reviews vary by year and sentiment scores.

2 Background

After we run the AMIC model on the wine review dataset and get the predicted sentiment scores for each wine review, we want to use these sentiment scores to study the changes in rating scores by year. AMIC is a model that integrates self-attention language models with traditional logistics regression.

3 Dataset

I used the part of the new dataset that contains a valid review year. The total number of reviews is 13,696.

	index	true_rating	predicted_rating	sentiment	rating	Clean_desc	review_year
0	13343	88.0	86.771576	0.937251	88	Light and soft, with cherry and raspberry flav...	NV (reviewed 2019)
1	4793	92.0	92.506035	1.662355	92	This creamy Champagne has a round juiciness to...	NV (reviewed 2023)
2	13811	91.0	90.806600	1.413998	91	Warm fruitcake, brown bread and mulled raisin ...	NV (reviewed 2017)
3	11029	90.0	90.160310	1.328816	90	A spiced rosé, with a Lambic-like hint of mace...	NV (reviewed 2014)
4	7355	93.0	89.342350	1.226706	93	Flattering and stylish, with baked peach, quin...	NV (reviewed 2021)

Figure 1: Wine Review Data Snap

The rating score distribution is roughly 60-100, and the sentiment range is from 0 to 2.5.

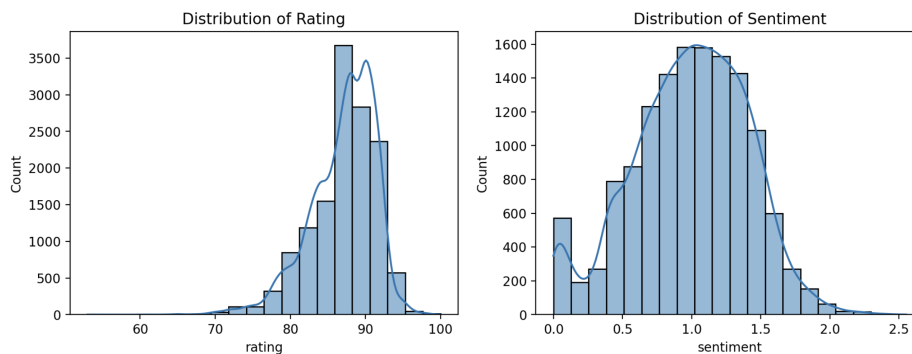


Figure 2: Distribution of Rating and Sentiment

The review year is from 1985 to 2025. The most reviews happened in 2015.

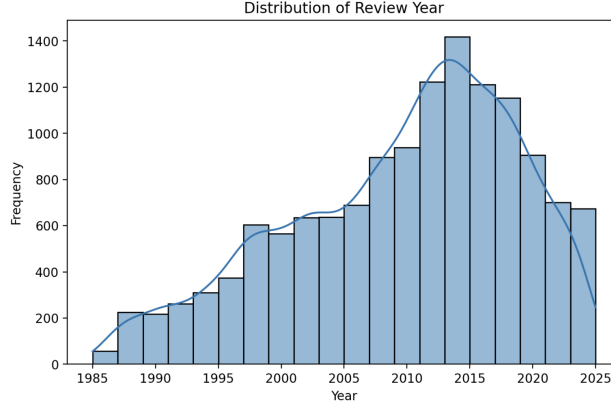


Figure 3: Enter Caption

4 Simple Linear Regression

The first step I took was to do a simple linear regression on rating by the variables: year, sentiment. I consider the year as a continuous variable. The summary of the regression is attached as follows. The adjusted R-squared is 0.796. 79.6% of rating variability is explained by year and sentiment. The remaining 20% rating variability is explained by factors that don't appear in the model. However, 79.6% still shows the relationship between rating and the two variables is strong. The p-values of two variables are statistically significant, especially the variable sentiment, which makes sense that sentiments are AMIC model outcomes.

$$\text{rating} = \beta_0 + \beta_1 \cdot \text{year} + \beta_2 \cdot \text{sentiment} + \varepsilon$$

$$E(\text{rating}) = 27.3 + 0.02 \cdot \text{year} + 9.03 \cdot \text{sentiment}$$

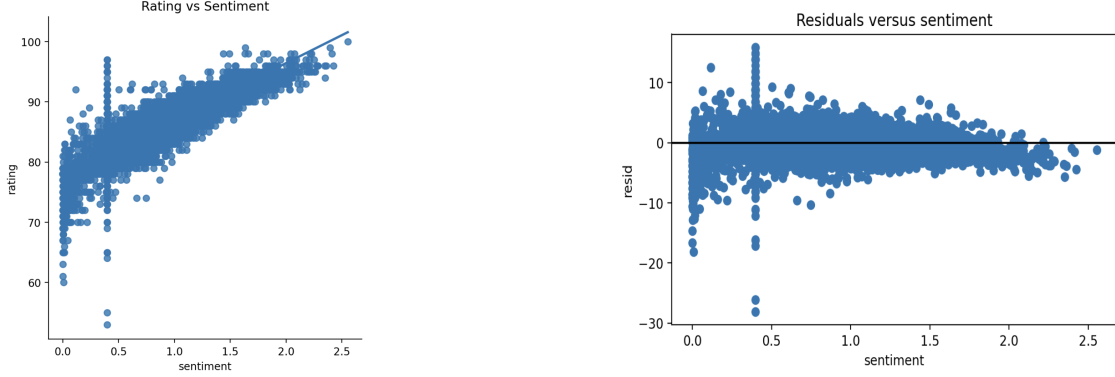
OLS Regression Results						
Dep. Variable:	rating	R-squared:	0.796			
Model:	OLS	Adj. R-squared:	0.796			
Method:	Least Squares	F-statistic:	2.668e+04			
Date:	Mon, 11 Aug 2025	Prob (F-statistic):	0.00			
Time:	18:33:16	Log-Likelihood:	-28748.			
No. Observations:	13696	AIC:	5.750e+04			
Df Residuals:	13693	BIC:	5.752e+04			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	27.3381	4.133	6.615	0.000	19.237	35.439
sentiment	9.0352	0.045	199.546	0.000	8.946	9.124
clean_year	0.0253	0.002	12.223	0.000	0.021	0.029
Omnibus:	3261.335	Durbin-Watson:	2.004			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	95177.840			
Skew:	-0.509	Prob(JB):	0.00			
Kurtosis:	15.874	Cond. No.	4.92e+05			

Notes:
 [1] Standard Errors assume that the covariance matrix of the errors is correctly
 [2] The condition number is large, 4.92e+05. This might indicate that there are strong multicollinearity or other numerical problems.

Figure 4: Summary of Linear Regression

The following plot shows the relationship between sentiment on the x-axis and rating on the y-axis. It shows a robust positive relationship, reinforcing our earlier simple linear regression. As we can observe, the variation of rating becomes less when the sentiment score is higher.

The following graph is the residuals versus sentiment from this linear regression. As we can observe, the residuals are evenly scattered across the solid zero line and majority of residuals fall within the absolute value 10. The residuals are much smaller when the sentiment scores are higher, which confirms our previous conclusion about the relationship between rating and sentiment. Finally, it is a QQ normal plot to diagnose



(a) Linear Plots between sentiment and rating

(b) Residuals versus sentiment

Figure 5: Comparison of model fit and residuals for the regression analysis.

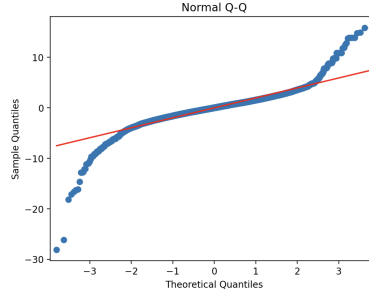


Figure 6: QQ Normal Plot

the linear regression. As we can observe, the beginning curve is below the diagonal line and the tail is above the diagonal line. This heavier tail suggests more extreme values than expected. There are more outliers in rating scores. The linear regression assumption that residuals are normal distribution is slightly compromised but the point estimations are still unbiased. The statistical tests may be inaccurate. One solution to this is to apply robust linear regression to reduce the influence from the outliers.

The next two plots show the trend of average ratings and average sentiment by year. As we can observe, it is obvious that the average ratings have been increasing over the year from 80 to 90. The similar trend occurs with the average trend of sentiment, increasing from 0.4 to 1.4. This shows that wine reviewers have been rating wines at a higher score over years.

5 Another Linear Regression

In this section, I used another type of linear regression to measure the impact of year on rating score. I used the year indicator, which is 0 when year is less than 2010 and 1 when year is more than 2010. The variables are one-year indicator, sentiment, and interaction term between year indicator and sentiment.

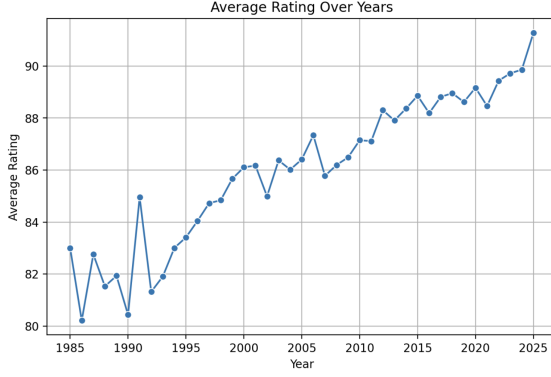
$$\text{rating} = \beta_0 + \beta_1 \cdot \text{year_indicator} + \beta_2 \cdot \text{sentiment} + \beta_3 \cdot \text{year_indicator:sentiment} + \varepsilon$$

$$E(\text{rating}) = 77.1 + 2.1 \cdot \text{year_indicator} + 10.0 \cdot \text{sentiment} - 1.8 \cdot \text{year_indicator:sentiment}$$

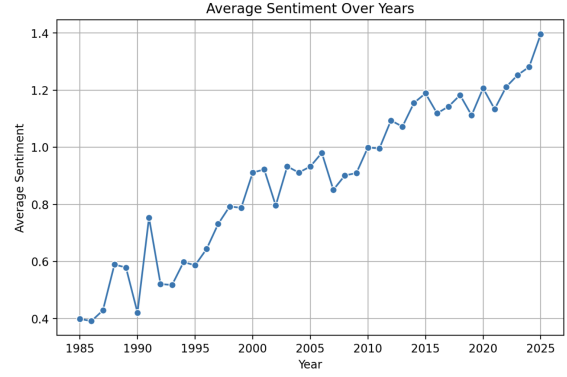
The adjusted R squared for this regression is 80.1% which means this model explains the variations more than the previous simple regression. All three variables are statistically significant.

The following two plots are residuals and the QQ normal diagnosis plot. Residuals are evenly scattered across the zero line and the heavy tail. shows in the QQ normal plot.

The year indicator coefficient is 2.1 in this regression model, which shows a positive relationship between year indicator and rating. In the first case where the review year is less than 2010, the regression equation



(a) Average Rating by year

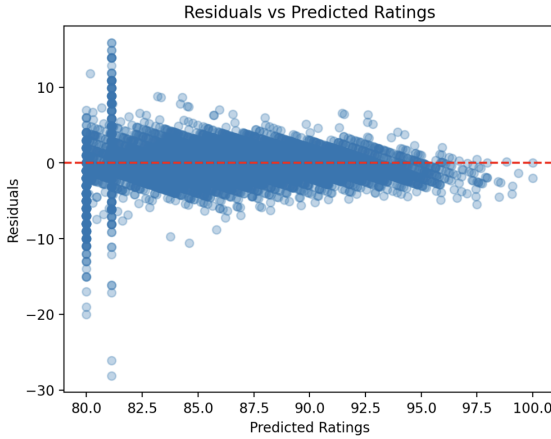


(b) Average Sentiment by year

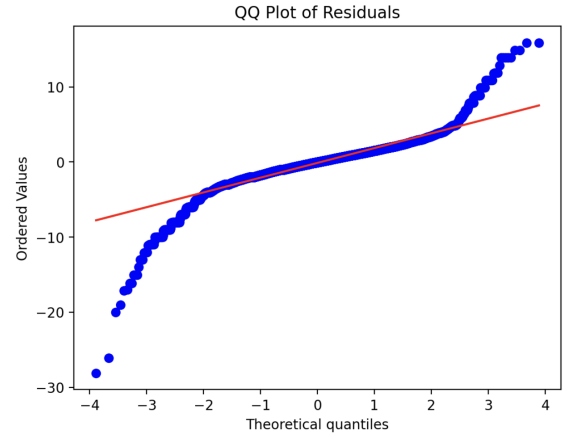
Figure 7: Comparison of average rating and sentiment over time.

OLS Regression Results						
Dep. Variable:	rating	R-squared:	0.801			
Model:	OLS	Adj. R-squared:	0.801			
Method:	Least Squares	F-statistic:	1.838e+04			
Date:	Tue, 12 Aug 2025	Prob (F-statistic):	0.00			
Time:	13:40:25	Log-Likelihood:	-28568.			
No. Observations:	13696	AIC:	5.714e+04			
Df Residuals:	13692	BIC:	5.717e+04			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	77.1079	0.054	1414.867	0.000	77.001	77.215
year_indicator	2.1017	0.092	22.752	0.000	1.921	2.283
sentiment	10.0366	0.060	168.129	0.000	9.920	10.154
year_indicator:sentiment	-1.8389	0.086	-21.325	0.000	-2.008	-1.670
Omnibus:	3158.404	Durbin-Watson:	2.004			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	98084.631			
Skew:	-0.439	Prob(JB):	0.00			
Kurtosis:	16.081	Cond. No.	14.1			

Figure 8: Summary of this interaction regression



(a) Residuals vs. interaction term



(b) QQ plot for interaction model residuals

Figure 9: Residual and QQ plots for interaction term model

becomes into

$$E(\text{rating}) = 77.1 + 10.0 \cdot \text{sentiment}$$

. This equation shows that every 1 point increase in sentiment will increase the rating by 10. In the second

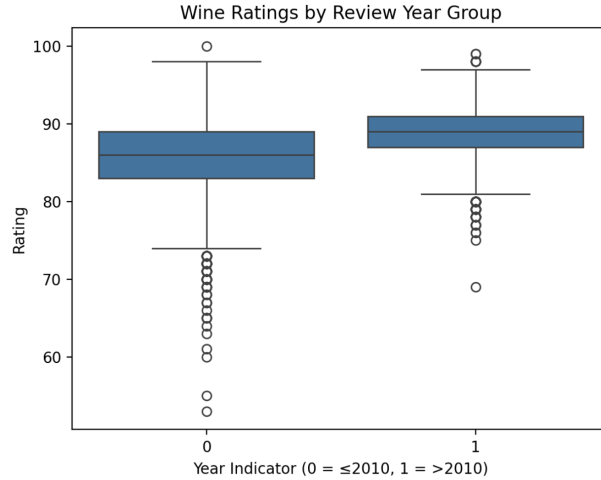


Figure 10: Boxplot for two case

sentiment_bin	
0	1768
1	4913
2	5670
3	1276
4	69

Figure 11: Counts of each bin

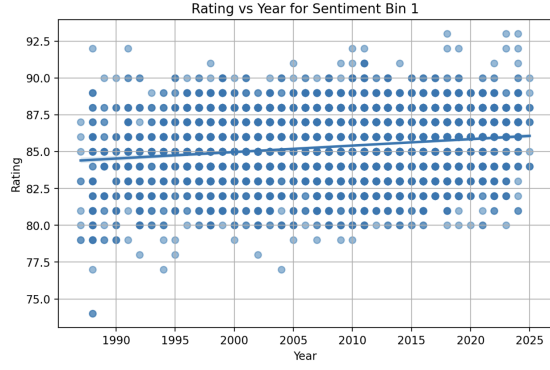
case where the review year is more than 2010, then the regression equation becomes into

$$E(\text{rating}) = 79.2 + 8.2 \cdot \text{sentiment}$$

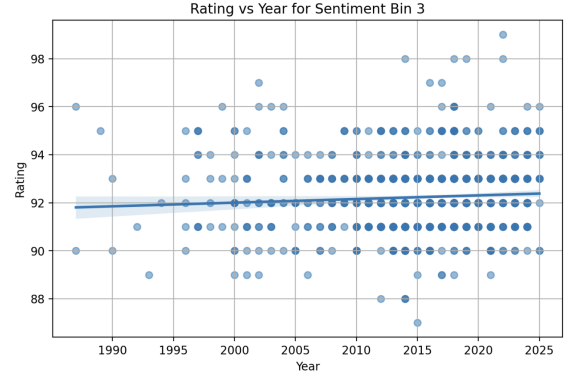
The intercept for this regression model is 2.1 higher than the baseline case. This makes sense when the later years would inflate the rating scores. However, the sentiment effect is weaker because every 1 point increase would only increase rating by 8.2 on average. The following plot shows the boxplot for two categories. The left box is for cases when review year is less and equal to 2010, and the right box is for cases when review year is larger than 2010. As we can observe, the median value of rating when review year is greater than 2010 is almost same as the upper quartile of rating when review year is less than 2010.

6 Bin Sentiment

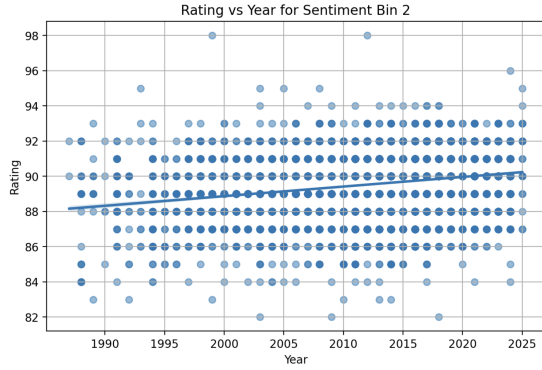
Now I want to visualize the relationship between 'rating' and 'clean_year' for similar scores of 'sentiment' using scatter plots. I created discrete bins for the 'sentiment' column based on specified intervals. I used the min value of sentiment, 0.5, 1.0, 1.5, 2.0 and max value of sentiment as the intervals to place all sentiment scores into each bin. For each sentiment bin, I created a scatter plot showing the relationship between 'rating' (on the y-axis) and 'clean_year' (on the x-axis). As we can observe, except for bin 4, in each bin there is a clear up slope trend of rating by year. Bin 3 and Bin 0 have slightly flat trend while Bin 1 and Bin 2 have more clear up slope trend. Bin 4 has down hill trend but we have much less data points, only 69 observations in this bin, so the accuracy of trend is uncertainty.



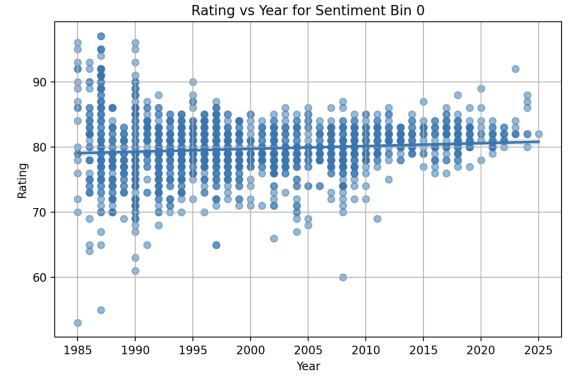
(a) Bin 1



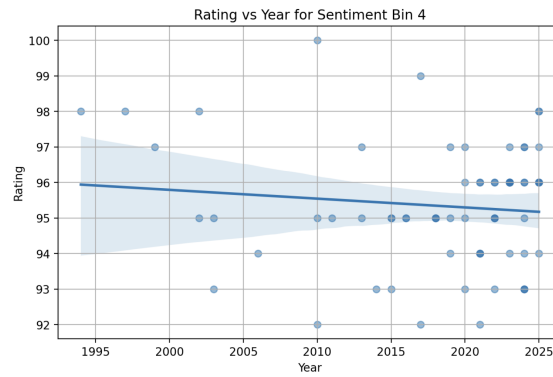
(b) Bin 3



(c) Bin 2



(d) Bin 0



(e) Bin 4

Figure 12: Comparison of ratings and sentiments across different bins