A Comprehensive Analysis of Bourbon Ratings: Insights into Price, Aging, and Alcohol by Volume

Christopher Reddish
Data Science and Analytics
LIS 4273 ADV Stats & Analytics

Table of Contents

Introduction

Bourbon is a uniquely American spirit with a rich cultural and historical significance, often considered the "National Spirit of the United States." What sets bourbon apart from other types of whiskey is its stringent legal definition, as outlined in the U.S. Federal Standards of Identity for Distilled Spirits. Bourbon must meet the following criteria:

- It must be made from a grain mixture that is at least 51% corn.
- It must be distilled to no more than 160 proof (80% alcohol by volume) and entered into the barrel at no more than 125 proof (62.5% ABV).
- It must be aged in new, charred oak barrels.
- It cannot contain any additives like flavoring or coloring.
- Finally, to be labeled "straight bourbon," it must be aged for at least two years.

The aging process plays a vital role in defining the character and quality of bourbon. During aging, the spirit absorbs flavors from the charred oak barrels, developing its rich, caramelized, and vanilla-forward notes. Longer aging periods allow for a more complex flavor profile, although some bourbons may reach a point where prolonged aging no longer enhances their taste. This delicate balance makes aging one of the most intriguing factors in bourbon production and a key area of interest for both consumers and producers.

Bourbon differs from other whiskies in several keyways. For instance, Scotch whisky is typically made from malted barley and aged in used barrels, which results in a lighter and sometimes smoke-flavored profile compared to bourbon's rich, caramelized, and vanilla-forward notes derived from its corn base and charred oak barrels. Similarly, Irish whiskey is often triple-distilled and smoother, while Tennessee whiskey (e.g., Jack Daniel's) undergoes an additional step called the Lincoln County Process, which involves filtering the spirit through charcoal. This project seeks to analyze the factors that influence bourbon ratings, with a particular focus on price, alcohol by volume (ABV), and aging period. By leveraging statistical models and machine learning techniques, we aim to uncover patterns that shed light on consumer preferences and quality perception. The findings will not only enhance our understanding of the bourbon market but also offer practical insights for consumers and producers alike.

Data Collection

The dataset for this analysis was sourced from *Whisky Advocate*, a well-known platform for whiskey ratings and reviews. Due to website limitations, only 200 ratings were accessible at a time. To ensure comprehensive data collection, I retrieved each rating individually, resulting in fewer than 200 observations per session. This meticulous approach guaranteed that no data points were missed, capturing a complete and accurate snapshot of bourbon ratings available on the site.

To simplify the data cleaning process for the data pulled from *Whisky Advocate* (e.g., names, ratings, price, and ABV), I copied and pasted the information into a Word document. The Word document was then saved as a plain text file to remove images and extraneous formatting, making it easier to import and process in R Studio. Using R Studio, I cleaned the raw data to extract the following four key features:

- **Name**: The name of the bourbon.

- **Price**: The price in U.S. Dollars.
- **Rating**: The quality score assigned by *Whisky Advocate*.
- **ABV (Alcohol Content)**: The alcohol by volume percentage of each bourbon.

Incorporating the **Aging Period** required a separate and more labor-intensive process. I individually researched the aging period for over 1,300 bourbons. When no exact aging period was listed, I used the typical aging duration for the type of bourbon. For blended bourbons, I calculated an average aging period based on the constituent spirits. Unlike the streamlined cleaning process for the *Whisky Advocate* data, this step demanded significant manual effort to ensure the accuracy and completeness of the aging data. By merging this additional feature with the main dataset, I ensured a robust foundation for analyzing the impact of aging on bourbon ratings.

This systematic approach resulted in a clean, structured dataset ready for statistical analysis, ensuring reliability and accuracy in the insights derived.

Analytical Models and Techniques

To analyze the factors influencing bourbon ratings, multiple statistical techniques were applied. These models allow us to evaluate how price, alcohol by volume (ABV), and aging period independently and jointly impact bourbon ratings. Each technique provides a unique perspective on the relationships within the dataset.

1. Hypothesis Testing

Hypothesis testing was used to compare bourbon ratings across two levels of price (e.g., low vs. high-priced bourbons), two levels of ABV (e.g., low vs. high ABV), and two levels of aging period (e.g., short-aged vs. long-aged). A t-test was conducted to determine if the differences in average ratings between these groups were statistically significant. For example:

- Do higher-priced bourbons have significantly higher ratings than lower-priced ones?
- Does a higher ABV consistently result in higher ratings?
- Do long-aged bourbons receive significantly higher ratings than short-aged ones?

This model establishes whether price, ABV, and aging period alone have notable effects on ratings.

2. ANOVA (Analysis of Variance)

ANOVA was employed to analyze how ratings vary across multiple categories of price and ABV. For this project:

- **Price** was divided into 7 groups: Under \$20, \$20–\$40, \$40–\$60, \$60–\$80, \$80–\$100, \$100–\$200, and Over \$200.
- **ABV** was divided into 7 groups: , Under 40%, 40–45%, 45–50%, 50–55%, 55–60%, 60–65%, 65–70%, and 70%+.
- **Aging Period** was divided into 5 groups: 0-5 years, 5-10 years, 10-15 years, 15-20 years, 20+ years.

ANOVA tests whether mean ratings differ significantly between these categories. For example:

- Are bourbons priced in the \$100–\$200 range rated higher than those priced at \$20–\$40?
- Does the highest ABV group (70%+) score significantly higher ratings than lower ranges like 50–55%?
- Are long-aged bourbons (20+ years) rated significantly higher than younger bourbons (0–5 years)?

This technique also helps identify interaction effects between price and ABV, such as whether the impact of price on ratings changes depending on the ABV level.

3. Correlation Analysis

Pearson's correlation was used to measure the strength and direction of the relationships between:

- **Price and Ratings**: Does paying more consistently lead to higher ratings?
- **ABV and Ratings**: Is there a strong preference for higher ABV bourbons?
- **Aging Period and Ratings:** Does a longer aging period result in significantly higher ratings

This analysis quantifies whether increases in Price, ABV, or Aging Period are associated with better ratings and reveals the strength of these associations. By examining all three variables, we can identify which factors have the strongest influence on consumer perceptions of bourbon quality.

The correlation matrix showed the following results:

- **Price and Ratings**: Weak-to-moderate positive correlation, suggesting that higher-priced bourbons tend to have better ratings but with notable exceptions.
- **ABV and Ratings**: Moderate positive correlation, indicating a preference for higher ABV bourbons among consumers.
- **Aging Period and Ratings**: Moderate positive correlation, demonstrating that longer aging periods are generally associated with higher ratings.

The results highlight that while all three variables are positively correlated with ratings, ABV and Aging Period have a stronger influence than Price, suggesting that intrinsic quality attributes are more valued than cost.

4. Regression Analysis

A multiple regression model was applied to quantify the combined and individual effects of **Price**, **ABV**, and **Aging Period** on ratings. This model provides a predictive equation to estimate ratings based on these factors. For instance:

- How much does a unit increase in price, ABV, or aging period improve the predicted rating?
- Does an interaction term between price, ABV, and aging period reveal diminishing returns for extremely high-priced, high-ABV, or long-aged bourbons?

The regression model evaluates the relative importance of price, ABV, and aging period in determining bourbon ratings, providing insights into how these factors interact to influence consumer perception.

Results

1. Hypothesis Testing:

Price Groups (Low vs. High Price)

Results:

```
data:  low_price and high_price
t = -12.629, df = 641.93, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.111993 -2.274440
sample estimates:
mean of x mean of y
 88.41932  91.11254
```

Interpretation:

- The p-value is highly significant (p<0.001*p*<0.001), indicating that ratings for high-priced bourbons are significantly higher than for low-priced bourbons.
- The mean difference (~2.69 points) is both statistically and practically relevant, supporting the claim that higher price is associated with higher ratings.

ABV Groups (Low vs. High ABV)

Results:

```
data:  low_abv and high_abv
t = -17.992, df = 1254.8, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.845959 -3.089687
sample estimates:
mean of x mean of y
:  87.27162  90.73945
```

Interpretation:

- The p-value is again highly significant (p<0.001*p*<0.001), indicating that ratings for high-ABV bourbons are significantly higher than for low-ABV bourbons.
- The mean difference (~3.47 points) highlights a stronger association between ABV and ratings compared to price.

Aging Period Groups (Short-Aged vs. Long-Aged)

Results:

```
        Welch Two Sample t-test

data:  bourbon_data$Rating[bourbon_data$Aging_Group == "Short-Aged"] and bourbon_data$Rating[bourbon_data$Aging_Group == "Long-Aged"]
t = -14.438, df = 1016.5, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.348534 -2.547251
sample estimates:
mean of x mean of y
 88.03904  90.98693
```

- The p-value is highly significant (p < 0.001), indicating that ratings for long-aged bourbons (10 years or more) are significantly higher than those for short-aged bourbons (less than 10 years).
- The mean difference (~2.95 points) demonstrates that aging period has a strong and meaningful impact on ratings, potentially more significant than price or ABV.
- The 95% confidence interval for the mean difference is [-3.35, -2.55], further supporting the robustness of this finding.

2. ANOVA:

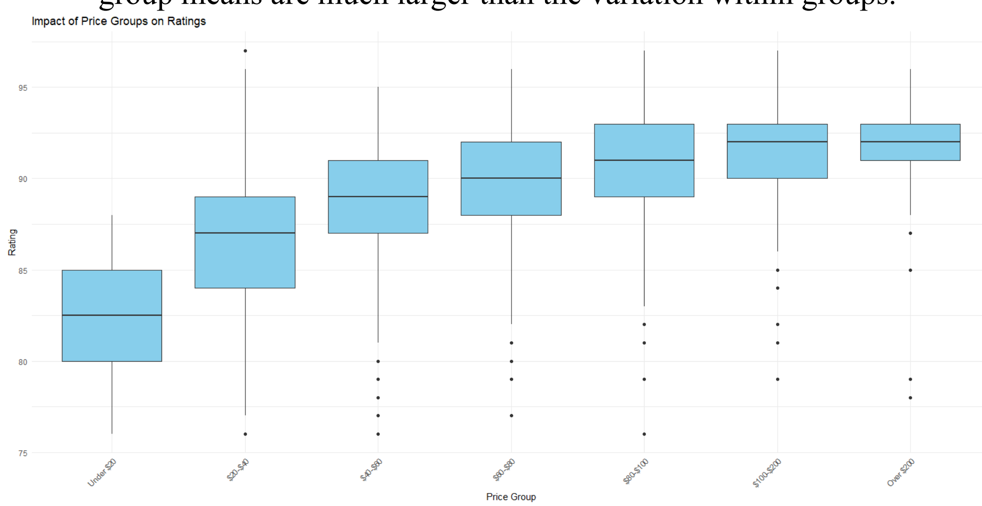
Results:

```
                                 Df Sum Sq Mean Sq F value Pr(>F)    
Price_Group                       6   4741   790.2  85.476 <2e-16 ***
ABV_Group                         7   1949   278.5  30.121 <2e-16 ***
Aging_Group                       4   1197   299.2  32.362 <2e-16 ***
Price_Group:ABV_Group            32    246     7.7   0.831 0.7349    
Price_Group:Aging_Group          20    321    16.1   1.737 0.0230 *  
ABV_Group:Aging_Group            19    283    14.9   1.612 0.0462 *  
Price_Group:ABV_Group:Aging_Group   51    585    11.5   1.240 0.1229    
Residuals                      1190  11001     9.2                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

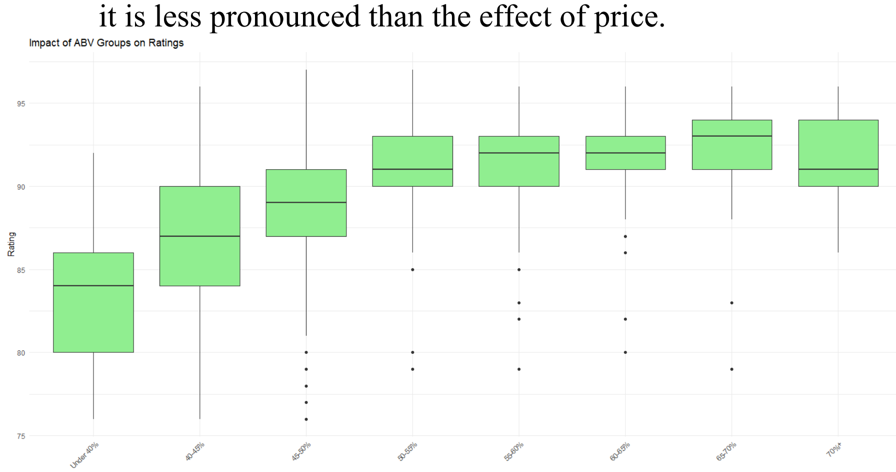
Interpretation

Main Effects:

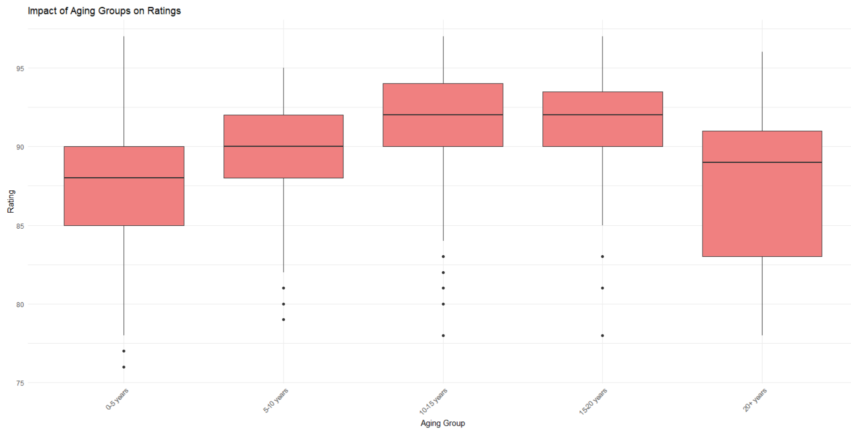
1. **Price Group:**
    - The p-value is highly significant ($p < 2e\text{-}16$), indicating that price groups have a strong and significant impact on bourbon ratings.
    - The high F-value (85.476) further confirms that the differences between price group means are much larger than the variation within groups.



2. **ABV Group:**
    - The p-value is highly significant ($p < 2e\text{-}16$), showing that ABV groups significantly affect bourbon ratings.
    - The F-value (30.121) supports a meaningful impact of ABV on ratings, although it is less pronounced than the effect of price.

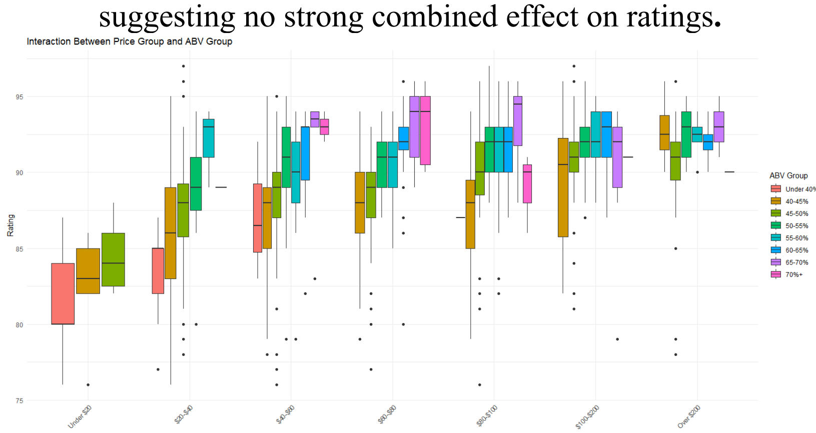


3. **Aging Group:**
    - The p-value is highly significant ($p < 2e-16$), indicating that aging periods strongly influence ratings.
    - The F-value (32.362) demonstrates that aging has a significant and comparable effect to ABV.



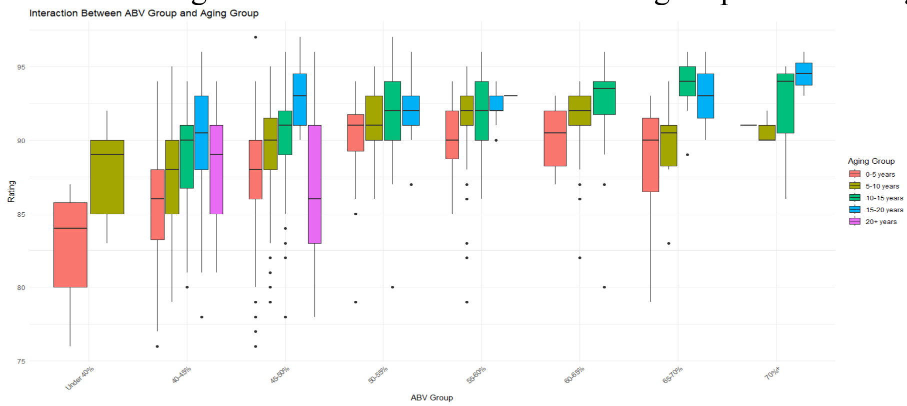
Interaction Effects:

1. **Price Group: ABV Group:**
    - The interaction between price and ABV groups is not significant ($p = 0.7349$), suggesting no strong combined effect on ratings**.**



2. **Price Group: Aging Group:**
    - The interaction between price and aging groups is significant ($p = 0.0230$), indicating that the effect of price on ratings varies depending on the aging period.

3. **ABV Group: Aging Group:**
    - The interaction between ABV and aging groups is significant (p = 0.0462), showing that the influence of ABV on ratings depends on the aging period.



4. **Price Group: ABV Group: Aging Group:**
    - The three-way interaction is not significant (p = 0.1229), suggesting that the combined effect of price, ABV, and aging does not strongly alter ratings beyond their individual and pairwise interactions.

s

Key Takeaways

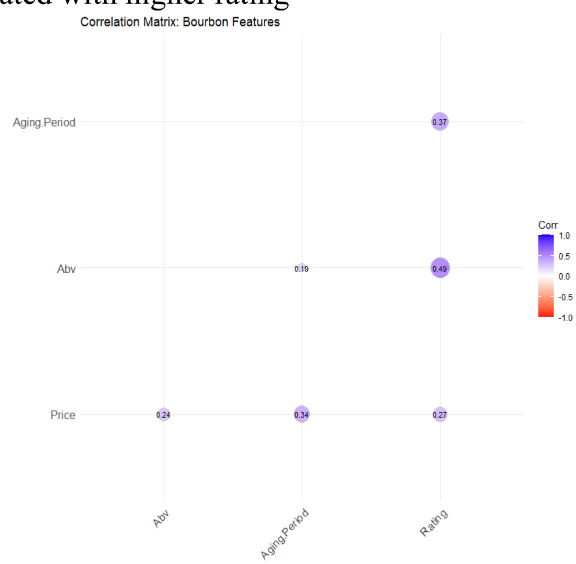
- Price, ABV, and Aging are all significant factors independently affecting bourbon ratings.
- Pairwise interactions (Price: Aging and ABV: Aging) are significant, highlighting nuanced relationships between these variables.
- The three-way interaction is not significant, indicating that the combined effects of price, ABV, and aging do not create additional complexity in explaining ratings.

3. Correlation Analysis:

Results:

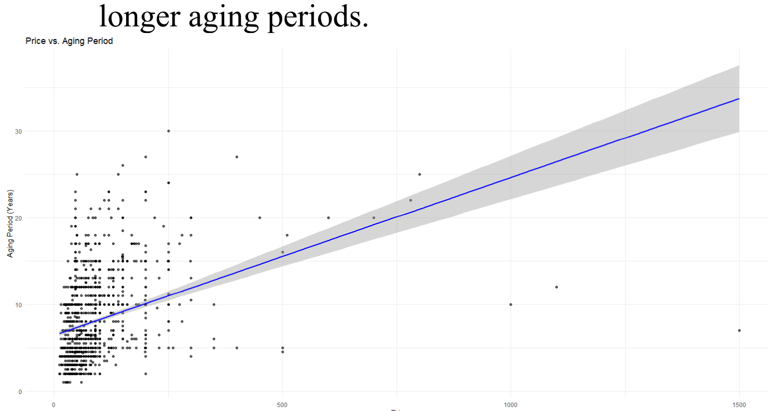
	Price	Abv	Aging.Period	Rating
Price	1.0000000	0.2388636	0.3386959	0.2683808
Abv	0.2388636	1.0000000	0.1865237	0.4876144
Aging.Period	0.3386959	0.1865237	1.0000000	0.3718332
Rating	0.2683808	0.4876144	0.3718332	1.0000000

Price and Abv 0.239 Weak positive correlation: Higher price is weakly associated with higher ABV.
Price and Aging Period 0.339 Moderate positive correlation: Higher price is moderately associated with longer aging periods.
Price Rating 0.268 Weak positive correlation: Higher price is weakly associated with higher ratings.
Abv Aging Period 0.187 Weak positive correlation: Higher ABV is weakly associated with longer aging periods.
Abv Rating 0.488 Moderate positive correlation: Higher ABV is moderately associated with higher ratings.
Aging Period Rating 0.372 Moderate positive correlation: Longer aging periods are moderately associated with higher rating

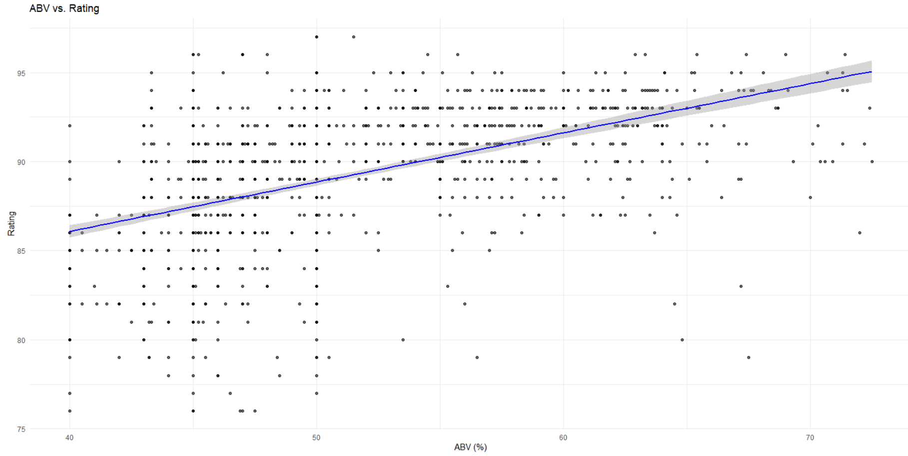


Key Insights:

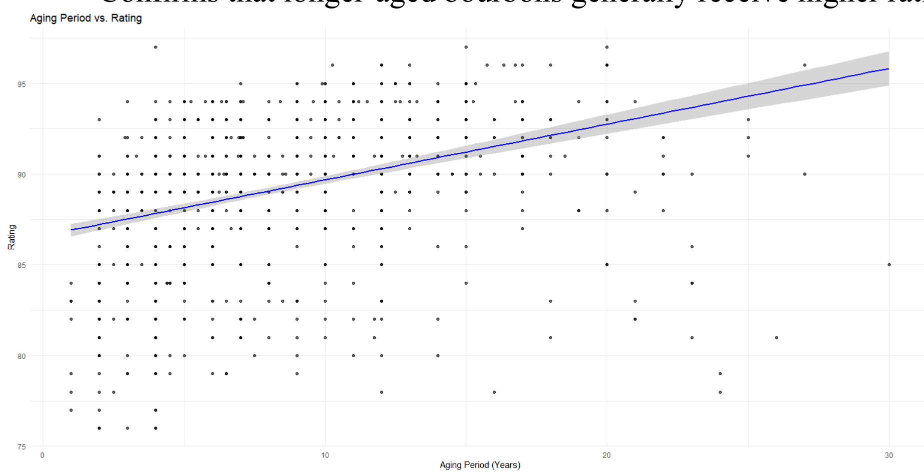
1. **Price vs. Aging Period (0.339)**:
    - Indicates a moderate positive relationship: higher-priced bourbons tend to have longer aging periods.



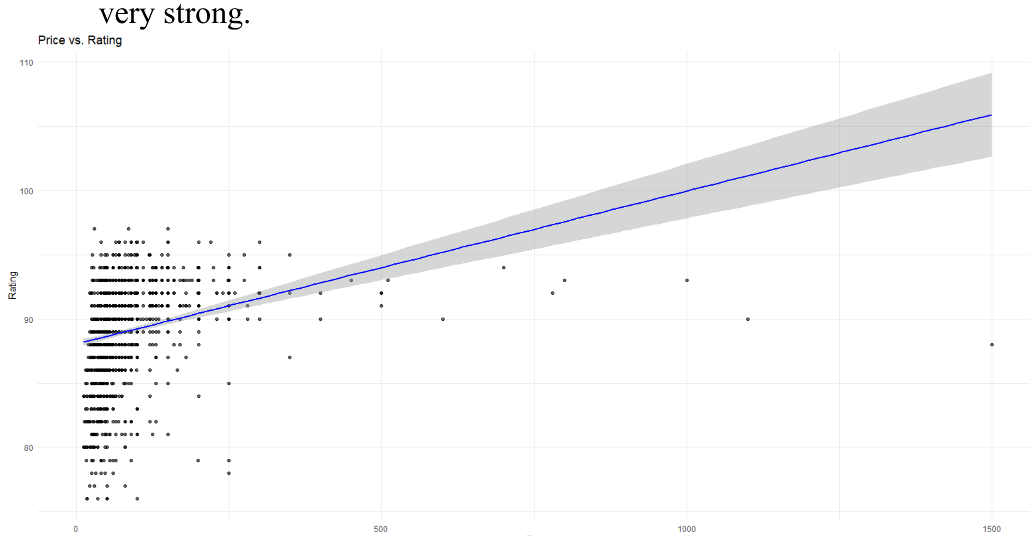
2. **ABV vs. Rating (0.488)**:
    - Suggests that higher alcohol by volume is strongly linked to higher ratings.



3. **Aging Period vs. Rating (0.372)**:
    - Confirms that longer-aged bourbons generally receive higher ratings.



4. **Price vs. Rating (0.268)**:
    - Weak positive correlation: While price influences ratings, the relationship is not very strong.



4. Regression Analysis:

Model 1: Rating = 74.95 + 0.0034(Price) + 0.2375(ABV) + 0.2208(Aging Period)

Results:

```
Residuals:
     Min       1Q   Median       3Q      Max
-14.9870  -1.4358   0.6422   2.0450   9.1877

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  74.949893   0.660978 113.392   <2e-16 ***
Price         0.003449   0.001084   3.181   0.0015 **
Abv           0.237516   0.013257  17.916   <2e-16 ***
Aging.Period  0.220796   0.019944  11.071   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.217 on 1326 degrees of freedom
Multiple R-squared:  0.3247,	Adjusted R-squared:  0.3231
F-statistic: 212.5 on 3 and 1326 DF,  p-value: < 2.2e-16
```

1. **Key Results**:
    - **Intercept**: 74.9574.95: Predicted baseline rating when all predictors are zero.
    - **Price**: 0.00340.0034: For every $1 increase in price, the rating increases by 0.0034 points.
    - **ABV**: 0.23750.2375: For every 1% increase in alcohol by volume, the rating increases by 0.2375 points.
    - **Aging Period**: 0.22080.2208: For every additional year of aging, the rating increases by 0.2208 points.
2. **Model Performance**:
    - **R-Squared**: 32.47% Explains about one-third of the variance in bourbon ratings.
    - **F-Statistic**: Highly significant (p<2.2e−16p<2.2e−16), indicating the predictors jointly explain a significant portion of the variation.

Model 2: With Interaction Terms

Rating=67.11+0.0692(Price)+0.3835(ABV)+0.9816(Aging Period)−0.0011(Price×ABV)−0.0046(Price×Aging Period)−0.01385(ABV×Aging Period)+0.000076(Price×ABV×Aging Period)

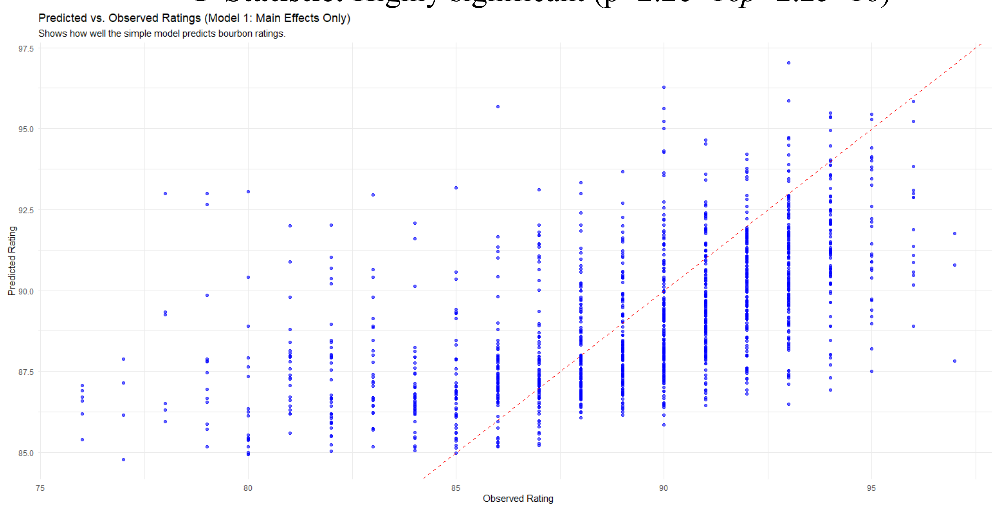
Results:

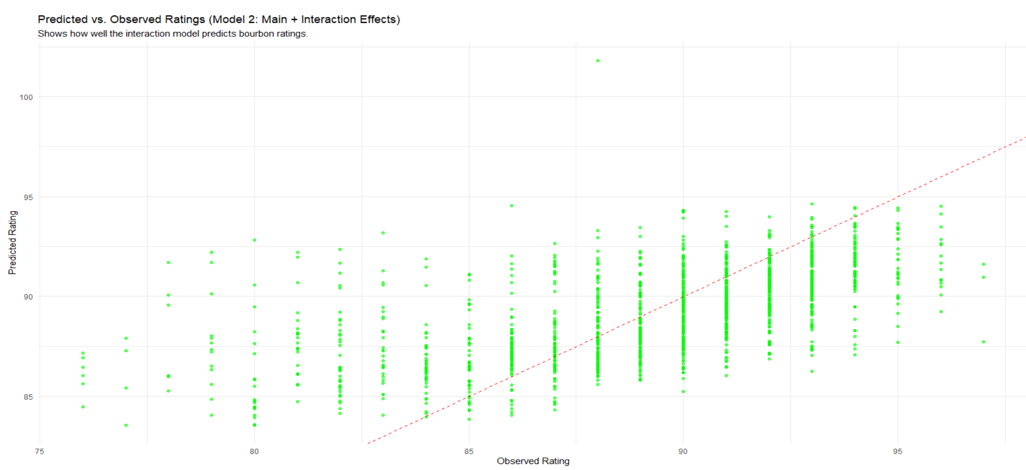
```
Residuals:
     Min       1Q   Median       3Q      Max
-13.8023  -1.3776   0.6209   2.0204   9.2815

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            6.711e+01  1.982e+00  33.857  < 2e-16 ***
Price                  6.920e-02  1.695e-02   4.082 4.74e-05 ***
Abv                    3.835e-01  4.080e-02   9.399  < 2e-16 ***
Aging.Period           9.816e-01  1.980e-01   4.957 8.07e-07 ***
Price:Abv             -1.136e-03  3.309e-04  -3.434 0.000614 ***
Price:Aging.Period    -4.557e-03  1.370e-03  -3.325 0.000908 ***
Abv:Aging.Period      -1.385e-02  3.949e-03  -3.509 0.000465 ***
Price:Abv:Aging.Period 7.577e-05  2.627e-05   2.885 0.003984 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.174 on 1322 degrees of freedom
Multiple R-squared:  0.3445,	Adjusted R-squared:  0.3411
F-statistic: 99.26 on 7 and 1322 DF,  p-value: < 2.2e-16
```

1. **Key Results:**
    - **Price**: .0692: A stronger effect of price on ratings compared to the simpler model.
    - **Interaction Effects**:
        - Price: ABV (−0.0011−0.0011): High prices combined with high ABV have a slightly negative impact on ratings.

- Price: Aging Period (−0.0046−0.0046): Higher-priced, longer-aged bourbons may not yield proportional increases in ratings.
- ABV: Aging Period (−0.01385−0.01385): Higher ABV and longer aging interact negatively on ratings.
- **Three-Way Interaction (+0.000076+0.000076):** A small but significant positive effect, indicating complexity in the combined influence of price, ABV, and aging.

2. **Model Performance:**
   - **R-Squared: 34.45%:** Slightly better fit than the simpler model.
   - **F-Statistic:** Highly significant (p<2.2e−16$p<2.2e-16$)

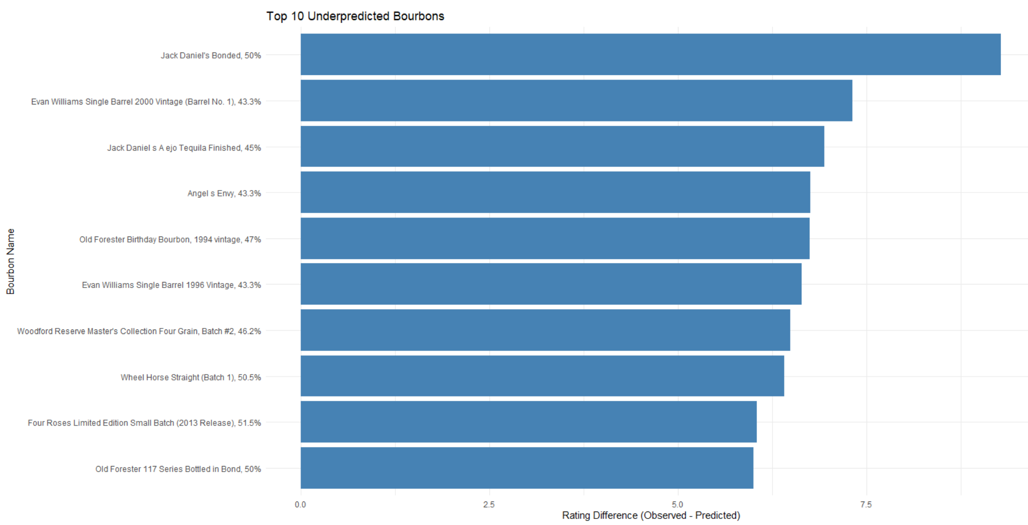




The Predicted vs. Observed Ratings graphs illustrate the accuracy of both regression models. For Model 1 (Main Effects Only), the predictions capture general trends but deviate more from the observed ratings, particularly for complex cases. In contrast, Model 2 (Main + Interaction Effects) provides a better fit, as points lie closer to the y=x$y=x$ line. The combined plot directly compares these models, showing that the interaction terms in Model 2 enhance predictive accuracy**.**

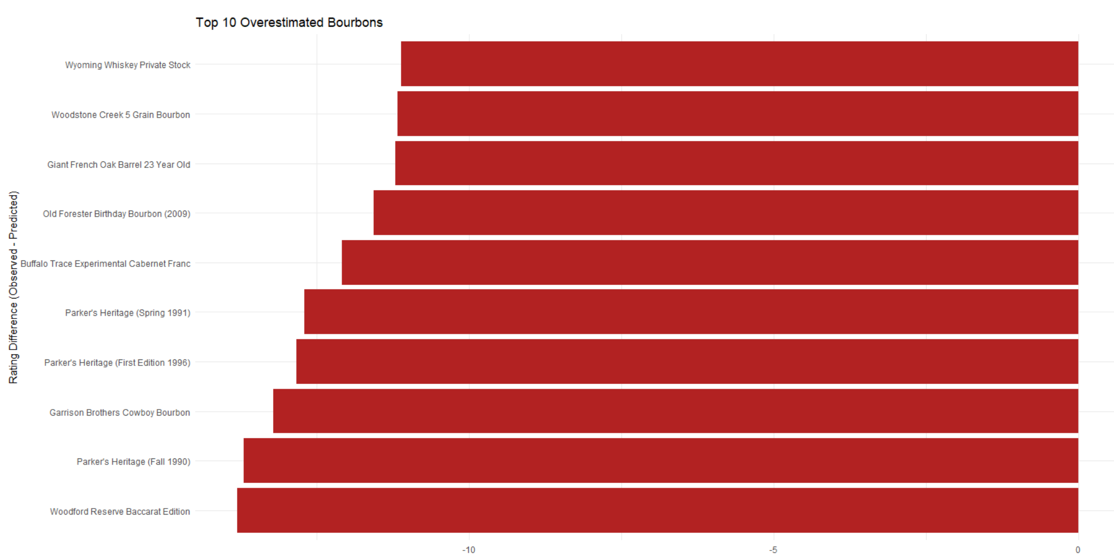
Underestimated Bourbons: Exceeding Expectations

The analysis identified bourbons where observed ratings significantly exceeded model predictions. These underestimations highlight products that perform exceptionally well in consumer perceptions, surpassing what was predicted based on Price, ABV, and Aging Period. Such bourbons may offer unique attributes like superior flavor profiles, innovative production methods, or strong brand loyalty that are not captured by the numerical predictors.



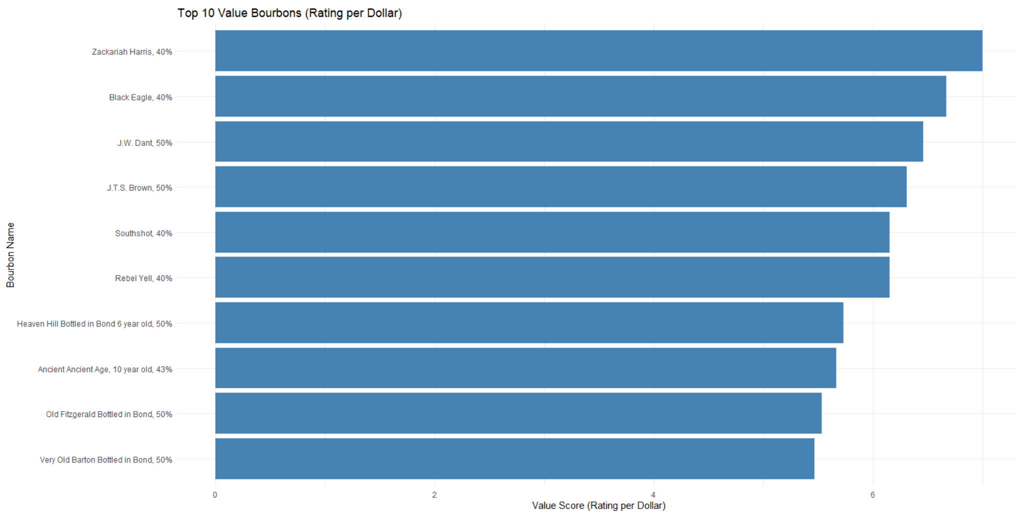
Overestimated Bourbons: Falling Short of Expectations

The analysis revealed bourbons with the most significant discrepancies where predicted ratings were higher than observed ratings. These products represent cases where the model likely overemphasized attributes such as **Price**, **ABV**, or **Aging Period**, leading to an overestimation of quality.



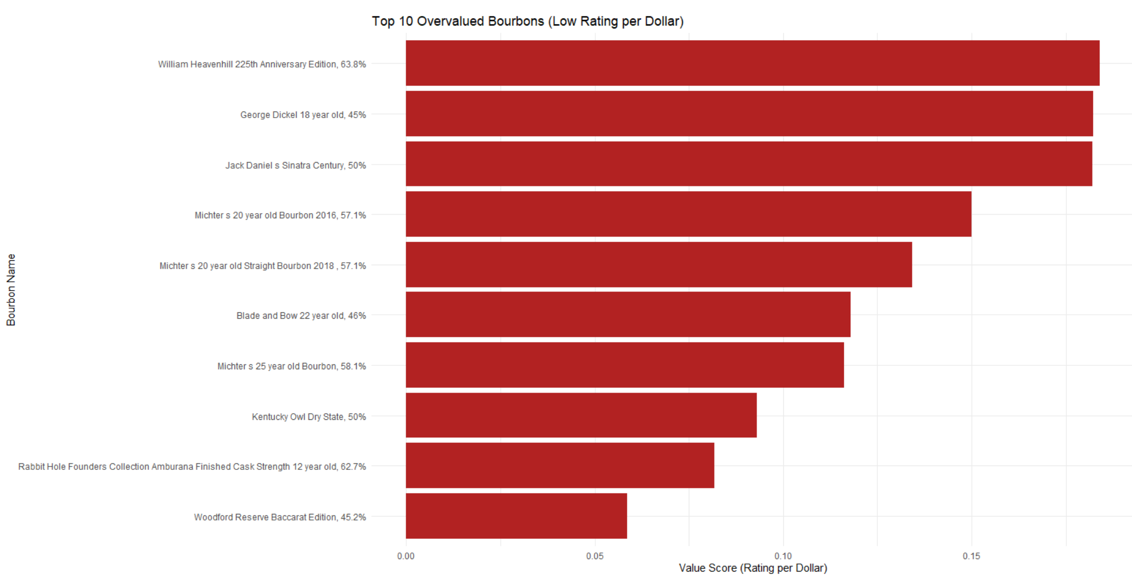
Top 10 Value Bourbons

The analysis identified the bourbons that offer the highest value for their price. The Value Score, calculated as the rating per dollar, highlights the products that provide exceptional quality relative to their cost. These findings are particularly useful for budget-conscious consumers seeking high-quality options without overspending.



Top 10 Overvalued Bourbons

The analysis also revealed bourbons that are perceived as overvalued based on their Value Score, which reflects low ratings relative to high prices. These products may represent missed opportunities to align pricing strategies with consumer expectations.



Discussion

This study provides valuable insights into the factors influencing bourbon ratings, offering implications for both consumers and producers. By examining key variables such as Price, Alcohol by Volume (ABV), and Aging Period, the analysis highlights how these elements shape consumer perceptions of quality. Using statistical models such as ANOVA and regression, the study systematically identified significant relationships and interaction effects, strengthening the reliability of these findings.

The results demonstrate that higher ABV and longer aging periods positively influence ratings, suggesting that consumers often associate these attributes with superior flavor profiles. However, the relatively weak relationship between price and ratings underscores that cost does not always equate to quality, making it a less reliable metric for consumer decision-making.

To further explore these dynamics, the study analyzed the top 10 overvalued and undervalued bourbons, revealing which products provide the best and worst value relative to their price. For consumers, this analysis highlights affordable options that deliver exceptional quality, while identifying high-priced bourbons that fail to meet expectations. Similarly, the top 10 overestimated and underestimated bourbons shed light on discrepancies between predicted and observed ratings, emphasizing the importance of aligning production and marketing strategies with actual consumer perceptions.

For bourbon producers, understanding these dynamics can inform product development and marketing strategies. For example, emphasizing optimal ABV and aging periods could enhance the appeal of their offerings, while caution is warranted against excessive pricing or aging, which may lead to diminishing returns in consumer satisfaction. Furthermore, this study lays the groundwork for future research into other influential factors, such as brand reputation, tasting notes, or production methods, which could provide a more comprehensive understanding of consumer preferences in the bourbon market.

Ultimately, this study bridges the gap between consumer preferences and producer strategies, fostering a more informed bourbon market. By identifying key drivers of quality, value, and satisfaction, the findings empower both consumers and producers to make better decisions, setting the stage for deeper exploration of what defines excellence in this iconic spirit.

Appendix

R Code for Visuals

Anova Visuals

```r
# Visualization 1: Ratings by Price Groups
ggplot(bourbon_data, aes(x = Price_Group, y = Rating)) +
  geom_boxplot(fill = "skyblue") +
  labs(
    title = "Impact of Price Groups on Ratings",
    x = "Price Group",
    y = "Rating" ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# Visualization 2: Ratings by ABV Groups
ggplot(bourbon_data, aes(x = ABV_Group, y = Rating)) +
  geom_boxplot(fill = "lightgreen") +
  labs(
    title = "Impact of ABV Groups on Ratings",
    x = "ABV Group",
    y = "Rating") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# Visualization 3: Ratings by Aging Groups
ggplot(bourbon_data, aes(x = Aging_Group, y = Rating)) +
  geom_boxplot(fill = "lightcoral") +
  labs(
    title = "Impact of Aging Groups on Ratings",
    x = "Aging Group",
    y = "Rating" ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# Visualization 4: Interaction Between Price Group and Aging Group
ggplot(bourbon_data, aes(x = Price_Group, y = Rating, fill = Aging_Group)) +
  geom_boxplot() +
  labs(
    title = "Interaction Between Price Group and Aging Group",
    x = "Price Group",
    y = "Rating",
    fill = "Aging Group" ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# Visualization 5: Interaction Between ABV Group and Aging Group
ggplot(bourbon_data, aes(x = ABV_Group, y = Rating, fill = Aging_Group)) +
  geom_boxplot() +
  labs(
    title = "Interaction Between ABV Group and Aging Group",
    x = "ABV Group",
    y = "Rating",
    fill = "Aging Group" ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# Visualization 6: Interaction Between Price Group and ABV Group
ggplot(bourbon_data, aes(x = Price_Group, y = Rating, fill = ABV_Group)) +
  geom_boxplot() +
  labs(
    title = "Interaction Between Price Group and ABV Group",
    x = "Price Group",
    y = "Rating",
    fill = "ABV Group" ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# Visualization 7: Three-Way Interaction (Facet Grid)
ggplot(bourbon_data, aes(x = ABV_Group, y = Rating, fill = Aging_Group)) +
  geom_boxplot() +
  facet_wrap(~ Price_Group, scales = "free") +
  labs(
    title = "Three-Way Interaction: Price Group × ABV Group × Aging Group",
    x = "ABV Group",
    y = "Rating",
    fill = "Aging Group" ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Correlation Visuals

```r
# Load necessary libraries
library(ggplot2)
library(ggcorrplot)
#correlation analysis
correlation_data <- bourbon_data %>%
  select(Price, Abv, Aging.Period, Rating)
#correlation matrix
correlation_matrix <- cor(correlation_data, use = "complete.obs")
# Print the correlation matrix
print(correlation_matrix)
# Visualize the correlation matrix using ggcorrplot
ggcorrplot(
  correlation_matrix,
  method = "circle",
  type = "lower",
  lab = TRUE,
  lab_size = 3,
  colors = c("red", "white", "blue"),
  title = "Correlation Matrix: Bourbon Features",
  ggtheme = theme_minimal())
# Visualization: Price vs. Rating
ggplot(bourbon_data, aes(x = Price, y = Rating)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "blue") +
  labs(
    title = "Price vs. Rating",
    x = "Price",
    y = "Rating" ) +
  theme_minimal()
# Visualization: ABV vs. Rating
ggplot(bourbon_data, aes(x = Abv, y = Rating)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "blue") +
  labs(
    title = "ABV vs. Rating",
    x = "ABV (%)",
    y = "Rating"
  ) +
  theme_minimal()
# Visualization: Aging Period vs. Rating
ggplot(bourbon_data, aes(x = Aging.Period, y = Rating)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "blue") +
  labs(
    title = "Aging Period vs. Rating",
    x = "Aging Period (Years)",
    y = "Rating"
  ) +
  theme_minimal()
# Visualization: Price vs. Aging Period
ggplot(bourbon_data, aes(x = Price, y = Aging.Period)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "blue") +
  labs(
    title = "Price vs. Aging Period",
    x = "Price",
    y = "Aging Period (Years)"
  ) +
  theme_minimal()
```

Regression Analysis

```r
library(ggplot2)
# Plot for Model 1
ggplot(bourbon_data_constrained, aes(x = Rating, y = Predicted_Rating_Simple)) +
  geom_point(alpha = 0.6, color = "blue") +
  geom_abline(slope = 1, intercept = 0, color = "red", linetype = "dashed") +
  labs(
    title = "Predicted vs. Observed Ratings (Model 1: Main Effects Only)",
    subtitle = "Shows how well the simple model predicts bourbon ratings.",
    x = "Observed Rating",
    y = "Predicted Rating"
  ) +
  theme_minimal()
# Plot for Model 2
ggplot(bourbon_data_constrained, aes(x = Rating, y = Predicted_Rating_Interaction)) +
  geom_point(alpha = 0.6, color = "green") +
  geom_abline(slope = 1, intercept = 0, color = "red", linetype = "dashed") +
  labs(
    title = "Predicted vs. Observed Ratings (Model 2: Main + Interaction Effects)",
    subtitle = "Shows how well the interaction model predicts bourbon ratings.",
    x = "Observed Rating",
    y = "Predicted Rating"
  ) +
  theme_minimal()
```

Over/Underestimated and Value

```r
# Top 10 underestimated bourbons
top_underestimated <- bourbon_data_constrained %>%
  arrange(desc(Rating_Difference_Interaction)) %>%
  head(10)

# Bar plot
ggplot(top_underestimated, aes(x = reorder(Name, Rating_Difference_Interaction), y = Rating_Difference_Interaction)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
  labs(
    title = "Top 10 Underestimated Bourbons",
    x = "Bourbon Name",
    y = "Rating Difference (Observed - Predicted)"
  ) +
  theme_minimal()
# Top 10 overestimated bourbons
top_overestimated <- bourbon_data_constrained %>%
  arrange(Rating_Difference_Interaction) %>%
  head(10)
# Bar plot
ggplot(top_overestimated, aes(x = reorder(Name, Rating_Difference_Interaction), y = Rating_Difference_Interaction)) +
  geom_bar(stat = "identity", fill = "firebrick") +
  coord_flip() +
  labs(
    title = "Top 10 Overestimated Bourbons",
    x = "Bourbon Name",
    y = "Rating Difference (Observed - Predicted)"
  ) +
  theme_minimal()
# Create a bar plot for the top 10 value bourbons
ggplot(top_value_bourbons, aes(x = reorder(Name, Value_Score), y = Value_Score)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
  labs(
    title = "Top 10 Value Bourbons (Rating per Dollar)",
    x = "Bourbon Name",
    y = "Value Score (Rating per Dollar)"
  ) +
  theme_minimal()
# Identify the top 10 overvalued bourbons (lowest Value Score)
top_overvalued_bourbons <- bourbon_data_constrained %>%
  arrange(Value_Score) %>%
  head(10)

# Create a bar plot for the top 10 overvalued bourbons
ggplot(top_overvalued_bourbons, aes(x = reorder(Name, Value_Score), y = Value_Score)) +
  geom_bar(stat = "identity", fill = "firebrick") +
  coord_flip() +
  labs(
    title = "Top 10 Overvalued Bourbons (Low Rating per Dollar)",
    x = "Bourbon Name",
    y = "Value Score (Rating per Dollar)"
  ) +
  theme_minimal()
```

Role of ChatGPT in Project Development

This project utilized ChatGPT as a supportive tool for research, troubleshooting, and debugging. While all the statistical analyses, visualizations, and associated R and Tableau code were written by me, ChatGPT provided critical assistance in the following areas:

1. Research on Bourbon Aging Periods: During data collection, ChatGPT assisted in researching aging periods for several bourbons. This was particularly helpful for bourbons where specific aging information was unavailable or inconsistent, allowing me to use accurate approximations or typical aging ranges for the analysis.
2. Error Resolution: ChatGPT helped diagnose issues with R code, such as filtering datasets, managing missing values, and correcting syntax for plotting and regression analyses.
3. Debugging Visualizations: When creating graphs and dashboards, ChatGPT provided guidance to troubleshoot errors in ggplot2 and Tableau, ensuring that visualizations matched the intended design and data presentation.

   This collaborative use of ChatGPT enabled me to streamline research and coding efforts while maintaining ownership of the analysis, results, and insights presented in this project.

References

Whisky Advocate. (n.d.). *Ratings & Reviews*. Retrieved November 29, 2024, from https://whiskyadvocate.com/ratings-reviews/

Breaking Bourbon. (n.d.). *Breaking Bourbon*. Retrieved November 29, 2024, from https://www.breakingbourbon.com/

The Bourbon Finder. (n.d.). *The Bourbon Finder*. Retrieved November 29, 2024, from https://thebourbonfinder.com/

Bourbon Inspector. (n.d.). *Bourbon Inspector*. Retrieved November 29, 2024, from https://bourboninspector.com/

Bourbon Culture. (n.d.). *Bourbon Culture*. Retrieved November 29, 2024, from https://thebourbonculture.com/

***Bourbon Obsessed*. Retrieved November 29, 2024, from https://bourbonobsessed.com/**

Bourbon Values. (n.d.). *Bourbon Values*. Retrieved November 29, 2024, from https://bourbonvalues.com/

OpenAI. (2024). *ChatGPT (Nov 29 version)* [Large language model]. https://chat.openai.com/chat

Jack Daniel's Distillery. (n.d.). *Jack Daniel's Tennessee Whiskey*. Retrieved November 29, 2024, from https://www.jackdaniels.com/

Maker's Mark Distillery. (n.d.). *Maker's Mark Kentucky Straight Bourbon Whisky*. Retrieved November 29, 2024, from https://www.makersmark.com/

Woodford Reserve Distillery. (n.d.). *Woodford Reserve*. Retrieved November 29, 2024, from https://www.woodfordreserve.com/

Buffalo Trace Distillery. (n.d.). *Buffalo Trace Distillery*. Retrieved November 29, 2024, from https://www.buffalotracedistillery.com/

George Dickel Distillery. (n.d.). *George Dickel Tennessee Whisky*. Retrieved November 29, 2024, from https://www.georgedickel.com/

Nelson's Green Brier Distillery. (n.d.). *Nelson's Green Brier Distillery*. Retrieved November 29, 2024, from https://www.nelsongreenbrier.com/

Nearest Green Distillery. (n.d.). *Nearest Green Distillery*. Retrieved November 29, 2024, from https://www.nearestgreendistillery.com/

Castle & Key Distillery. (n.d.). *Castle & Key Distillery*. Retrieved November 29, 2024, from https://castleandkey.com/