Tequila meets Technology

A new look at Tequila

Robert Gravelle

2023-12-06

Contents

Figure 1: Background Image

Introduction

A brief history of Tequila

Once upon a time in the sun-drenched lands of Mexico, there existed an ancient spirit with roots deeply embedded in the soil and traditions of the indigenous peoples. Long before the arrival of the Spanish conquistadors, the agave plant was revered for its versatile uses. The native people fermented its sap to craft a magical elixir known as "pulque," a beverage that would set the stage for the birth of a legendary spirit.

As the Spanish adventurers, led by the intrepid Hernán Cortés, set foot on Mexican soil in the 16th century, they brought with them the knowledge of distillation. The alchemy of turning humble agave into a distilled spirit began, giving birth to what we now know as tequila.

The stage for tequila's grand performance was set in the picturesque town of Tequila and the surrounding regions of Jalisco. Here, the volcanic soil and ideal climate nurtured the blue agave plant, which flourished under the watchful gaze of the Jalisco sun.

In 1758, a pioneer named José Antonio Cuervo established the first licensed tequila distillery, marking a pivotal moment in the spirit's history. The golden elixir, aged in oak barrels, soon captured the hearts and palates of those who indulged in its magic.

The Mexican government, recognizing the importance of preserving tequila's authenticity, introduced regulations in the 20th century. In 1974, the Denomination of Origin (DO) was established, defining the geographical boundaries where true tequila could be crafted. These regulations specified the types of agave permitted, with the blue agave taking center stage in the tequila tale.

As the years passed, tequila transcended borders, becoming an international sensation synonymous with fiestas and celebrations. The distinct categories of Blanco, Reposado, Añejo, and Extra Añejo added layers to the story, each contributing its unique flavor to the narrative.

Tequila became not just a drink but a cultural symbol, woven into the fabric of Mexican identity. It found its way into the hands of storytellers, artists, and revelers worldwide, leaving an indelible mark on the global spirits scene.

And so, the legend of tequila continues to unfold—a tale of agave fields whispering in the breeze, of distilleries crafting liquid gold, and of glasses raised in celebration. It is a story that intertwines the ancient traditions of Mexico with the modern rhythms of a world captivated by the spirit of tequila. And with each sip, the journey through time and taste goes on, inviting all to savor the magic born from the heart of the agave plant and the spirit of a land called Mexico.

Technology helping traditions

Tequila aficionados are always on the lookout for novel and exceptional spirits, yet the hesitation to invest in a bottle without certainty of appreciation remains a common concern. Enter technology—a transformative force not only for recording and exploring ratings but also for predicting them. The key lies in deciphering the elements that constitute a stellar tequila, encompassing Agave regions, tequila categories, water sources, distillation techniques, production facilities, and the skilled artisans behind each creation.

Anticipating ratings necessitates a comprehensive understanding of current ratings and trends spanning several years. Embracing technology allows us to chart future trends and ratings for tequilas, enabling consumers to make informed decisions about unfamiliar spirits. It's a digital guide, providing more than just personal preferences—a data-driven compass predicting the trajectory of tequila experiences.

This innovative approach not only empowers consumers to navigate uncharted tequila territories with confidence but also exemplifies the harmonious synergy between technology and the beverage industry. As we witness this convergence, technology becomes a beacon for enthusiasts and connoisseurs, offering insightful solutions that elevate the exploration of tequila to new heights.

Loading and Exploring Data

This project revolves around loading insights from three key files: fulldataman.csv, weather.csv, and ybyratings.csv. Picture these files as the culmination of an intricate ballet between data scraping from various websites and the continuous processing by diverse programs. The beauty lies in the fact that these files are not static; they're a reflection of the most recent tequila landscape.

As we ride the wave of data evolution, the files undergo meticulous preprocessing, acting as guardians of data integrity. This ensures that with every update the analysis remains robust, adapting seamlessly to the ever-shifting dynamics of tequila information.

One of the web scraping algorithims I wrote is contained below.

library(dplyr) library(rvest)

brands=NULL for(i in 1:58) { url=paste("https://tequilamatchmaker.com/brands?page=",i,sep="") wc <- read_html(url) urls= wc %>% html_nodes('tr.clickable-row') %>% html_attr('data-href') urls=paste("https://tequilamatchmaker.com",urls,sep="")

brands=c(brands,urls) print(i)

}

brands=unique(brands) write.csv(unique(brands),"brands.csv")

allurls=NULL ua <-"Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:106.0) Gecko/20100101 Firefox/106.0"

for(i in brands) {

ii <- GET(i, user_agent(ua))

wc <- read_html(ii) urls= wc %>% html_nodes('a.product-list__item-link.material-shadow.material-shadow-hover') %>% html_attr('href') urls=paste("https://tequilamatchmaker.com",urls,sep="") allurls=c(allurls,urls) print(i) Sys.sleep(.5) } write.csv(unique(allurls),"allurls.csv")

Data Summary

In the symphony of data, we have a trio of CSV files to elevate the tequila analysis. These live files undergo a preprocessing, into the file fulldataman.csv. This data source encapsulates the essence of Tequila, featuring names, NOM, distillation processes, ratings, flavors, aromas, and more. Imagine it as a distilled blend of information, meticulously collected from various online sources and expertly merged into this data haven.

The second file is the weather.csv, a meteorological data source of tequila and agave growth regions. Location, sunshine hours, humidity, temperature – this file paints a vivid meteorological portrait sourced from multiple online platforms.

Completing the third file is ybyratings.csv, a collection of summarized tequila ratings spanning the years 2020, 2021, and 2022. This file acts as a compass for trend analysis, meticulously filtering ratings based on the review dates. Together, these files form a grand ensemble, ready to serenade your tequila exploration.

Additional Notes

One thing that I always do when I program, compose reports, or anything that gets published on the web I always perform a plagiarism check. One of the websites that I use is https://www.check-plagiarism.com/. The results of the scan report are

PLAGIARISM SCAN REPORT Date December 06, 2023 Unique Content 100% Word Count 5936 Plagiarized Content 0%

Exploratory Data Analysis

First and foremost, we must amass a diverse set of data essential for exploring the myriad facets of Tequila. This includes insights into Tequila by regions, cooking methods, NOM production, extraction methods, distillation methods, and ABV.

Tequila Exploratory Analysis

Regions

Now let's embark on an exploration of the regions where Tequila is produced.

Region	Total
Jalisco (Los Valles)	2498
Jalisco (Los Altos Southern)	1442
Jalisco (Central)	479
Jalisco (Cinega)	389
Guanajuato	44
Jalisco (Southeast)	33
Michoacn	29
Jalisco (Southern)	21
Tamaulpas	13
Jalisco (South Coast)	6
Jalisco (Sierra de Amula)	3
Jalisco (Sierra Occidental)	2
Jalisco (Los Altos)	1
Jalisco (North Coast)	1

Cooking Types

Now let's embark on an exploration of the cooking methods for Tequila

Cooking	Total
Stone/Brick Ovens	2324
Autoclave (high pressure)	1593
Autoclave (low pressure)	179
Stone/Brick Ovens Autoclave (high pressure)	134
Acid-Thermal Hydrolysis	108
Autoclave (post diffuser)	77
Boiling Tank	13
Autoclave (high pressure) Stone/Brick Ovens	12
Stone/Brick Ovens Autoclave (low pressure)	9
Stone/Brick Ovens Boiling Tank	9
Earthen Pit	5
Stone/Brick Ovens Autoclave (post diffuser)	5
Steel Oven (Direct Flame)	4
Earthen Pit Autoclave (high pressure)	2
Stone/Brick Ovens Diffuser	2
Diffuser	1

NOM Top ten producers

Now, let's embark on an exploration of the top ten NOMs, each representing a unique facet of the tequila landscape. NOM stands for "Norma Oficial Mexicana," which translates to "Official Mexican Standard." In the context of tequila, it refers to the official regulations and standards set by the Mexican government to ensure the quality and authenticity of tequila production. Each tequila-producing distillery is assigned a unique NOM number, serving as both an identifier and a testament to the adherence to established standards. With 1,754 tequilas currently in production and only 152 NOMs, it's a testament to the rich diversity within the framework of the Official Mexican Standard.

#Count by NOM top ten

Table 1: Count by NOM (Top 10)

NOM	Total
1438	249
1459	226
1137	223
1499	160
1480	134
1466	130
1477	115
1173	109
1479	109
1414	99

Count by Extraction methods

Now let's see the Extraction method of Tequila by method

Table 2: Count by Extraction Methods

Extraction	Total
Roller Mill	3894
-	346
Diffuser	232
Screw Mill	220
Tahona	103
Screw Mill Roller Mill	64
Tahona Roller Mill	39
Roller Mill Screw Mill	26
Tahona Screw Mill	10
Roller Mill Tahona	9
Diffuser Roller Mill	7
Diffuser Screw Mill	5
Roller Mill Diffuser	5
Hand Crushed	1

Tequila Production by Distillation Method

Now let's see the distillation methods of Tequila ranked by order

Table 3: Tequila Production by Distillation Method

Distillation	Total
2x distilled	4457
	271
3x distilled	227
4x distilled	3
5x distilled	2
7x distilled	1

Count by Still Methods

Table 4: Count by Still Methods

Still	Total
Stainless Steel Pot	1928
Copper Pot	1166
Stainless Pot w/Copper Coil	999
-	587
Column	195
Column Stainless Pot w/Copper Coil	25
Copper Pot Column	22
Copper Pot Stainless Steel Pot	14
Stainless Steel Pot Column	10
Column Copper Pot	8
Stainless Steel Pot Copper Pot	4
Copper Pot Stainless Pot w/Copper Coil	2
Filipino Style (wood/copper) Stainless Pot w/Copper Coil	1

Count by Water Source

Table 5: Tequila Production by Water Source

Water_Source	Total
Deep well water	1159
Natural spring water	319
Natural spring water Rain water	8
Deep well water Natural spring water Rain water	4
Distilled water	4
Deep well water Rain water	3
Deep well water Natural spring water	1
Rain water Natural spring water	1

Count by ABV

Table 6: Count by ABV

ABVorProof	Total
40% abv (80-proof)	3318
35-39% abv (70-78 proof)	626
38% abv (76 proof)	84
35% abv (70 proof)	78
35-39% abv (70-78 proof) 40% abv (80-proof)	47
50% abv (100-proof)	38
40% abv (80-proof) 35-39% abv (70-78 proof)	34
42% abv (84-proof)	32
40% abv (80-proof) 38% abv (76 proof)	30
46% abv (92-proof)	23
55% abv (110-proof)	20
43% abv (86-proof)	18
45% abv (90-proof)	12
40% abv (80-proof) 35% abv (70 proof)	8
41.5% abv (83-proof)	8

Data Analysis and Modeling

Ratings for overall Regions

In this analysis, we delve into the distinctions between panel and community ratings and explore the variations they exhibit. To discern these differences, I partitioned the data into two distinct sets. This segregation arose from disparities observed in how Tequila Sommeliers, industry professionals, and the general public assess and grade tequilas. Notably, I amalgamated Sommeliers and industry professionals into a unified category due to the remarkable similarity in their results, whereas the general public's ratings demonstrated considerable divergence from those of industry professionals.

Here, we commence the breakdown of the data to examine how regions are typically rated, considering panel, community, and aggregated ratings. This approach provides a comprehensive overview of the ratings attributed to each region, encompassing evaluations from diverse perspectives.

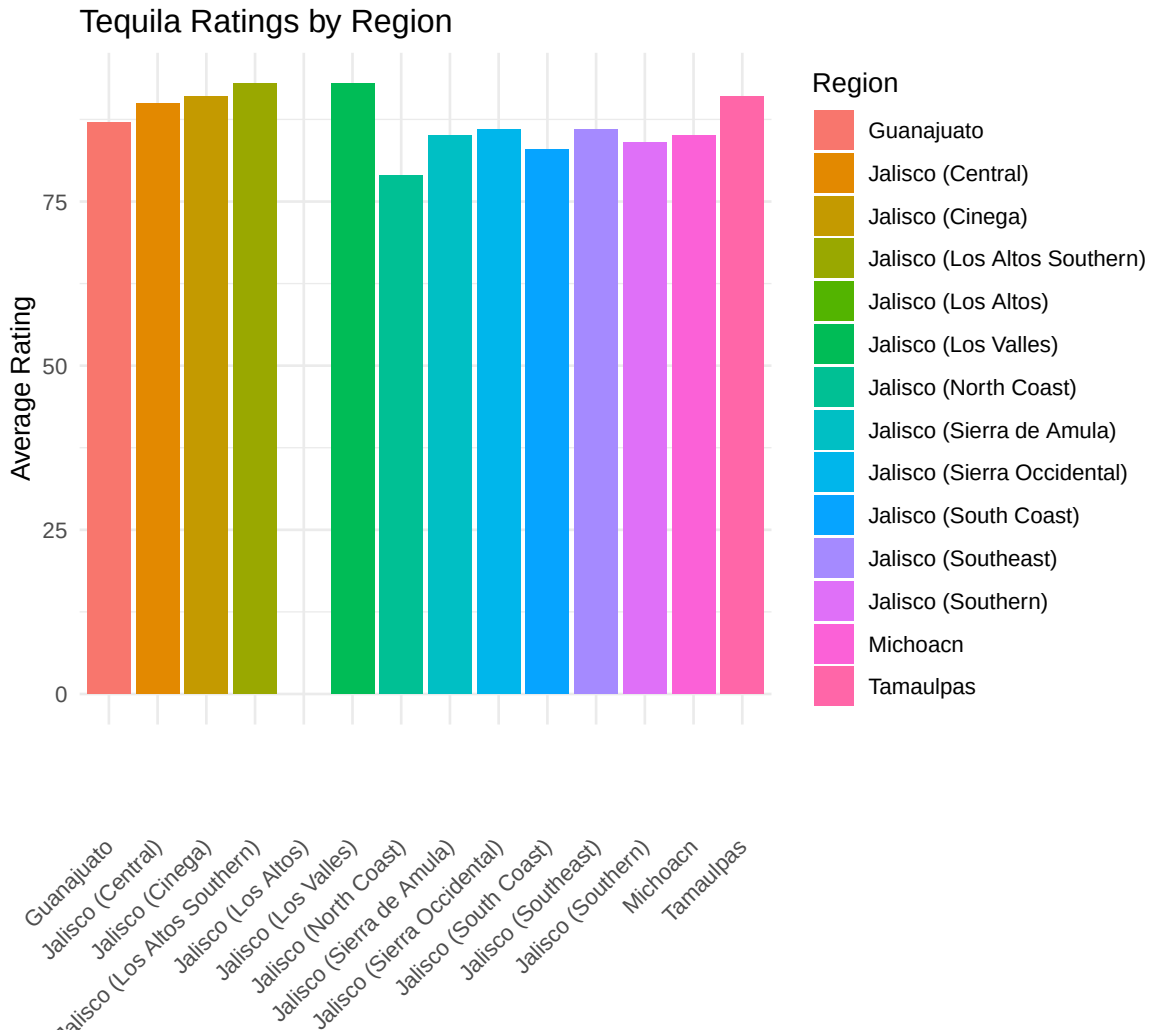


Figure 2: Tequila Ratings by Region

Table 7: Average Rating for Community Rating by Region

Region	AverageRatingCommunity
Guanajuato	80.56
Jalisco (Central)	77.30
Jalisco (Cinega)	79.87
Jalisco (Los Altos Southern)	79.41
Jalisco (Los Altos)	NaN
Jalisco (Los Valles)	80.23
Jalisco (North Coast)	76.00
Jalisco (Sierra Occidental)	85.00
Jalisco (Sierra de Amula)	89.50
Jalisco (South Coast)	79.00
Jalisco (Southeast)	78.68
Jalisco (Southern)	81.58
Michoacn	79.33
Tamaulpas	84.17

Table 8: Highest and Lowest Community Ratings by Region

Region	HighestCommunityRating	LowestCommunityRating
Guanajuato	94	45
Jalisco (Central)	100	29
Jalisco (Cinega)	98	12
Jalisco (Los Altos Southern)	100	1
Jalisco (Los Altos)	-Inf	Inf
Jalisco (Los Valles)	100	1
Jalisco (North Coast)	76	76
Jalisco (Sierra Occidental)	85	85
Jalisco (Sierra de Amula)	91	88
Jalisco (South Coast)	80	78
Jalisco (Southeast)	87	64
Jalisco (Southern)	99	61
Michoacn	88	63
Tamaulpas	94	71

Ratings from Commumity and Panel raters

Here, we examine the community ratings for each region. Initially, I suspected errors in the data due to the wide spread of reported ratings. Upon careful examination, I confirmed the accuracy of the ratings. Some individuals assigned exceptionally low ratings, as certain tequilas were labeled as Mixtos rather than genuine tequila. I opted to retain these ratings, providing a realistic reflection of the diverse perspectives captured in the data.

Table 9: Average Rating for Panel Rating by Region

Region	AverageRatingPanel
Guanajuato	77.00
Jalisco (Central)	74.09
Jalisco (Cinega)	79.38
Jalisco (Los Altos Southern)	79.00
Jalisco (Los Altos)	NaN
Jalisco (Los Valles)	77.32
Jalisco (North Coast)	79.00
Jalisco (Sierra Occidental)	85.00
Jalisco (Sierra de Amula)	85.00
Jalisco (South Coast)	74.60
Jalisco (Southeast)	77.65
Jalisco (Southern)	79.62
Michoacn	75.67
Tamaulpas	77.78

Table 10: Highest and Lowest Panel Ratings by Region

Region	HighestPanelRating	LowestPanelRating
Guanajuato	87	66
Jalisco (Central)	90	9
Jalisco (Cinega)	91	28
Jalisco (Los Altos Southern)	93	17
Jalisco (Los Altos)	-Inf	Inf
Jalisco (Los Valles)	93	10
Jalisco (North Coast)	79	79
Jalisco (Sierra Occidental)	86	84
Jalisco (Sierra de Amula)	85	85
Jalisco (South Coast)	83	63
Jalisco (Southeast)	86	63
Jalisco (Southern)	84	72
Michoacn	85	38
Tamaulpas	91	57

Density plot ratings for the panel and community

In this analysis, we delve into the panel ratings for each region. Initially, there were concerns about data accuracy due to the extensive range of reported ratings just as reported in the community raters. However, upon meticulous examination, the ratings were verified to be accurate. It was observed that some individuals assigned notably low ratings but not as low as the community raters, attributing them to tequilas labeled as Mixtos rather than genuine tequila. Despite this variation, these ratings have been retained to offer a genuine portrayal of the diverse perspectives captured in the data.

We will now explore the distinctions in ratings between the community and the professional panel using a density plot graph. This visual representation unveils the nuances in rating patterns. The community tends to assign higher ratings, showcasing a tendency towards positivity, while the professional panel exhibits a more discerning approach with a broader range of ratings, including lower scores.

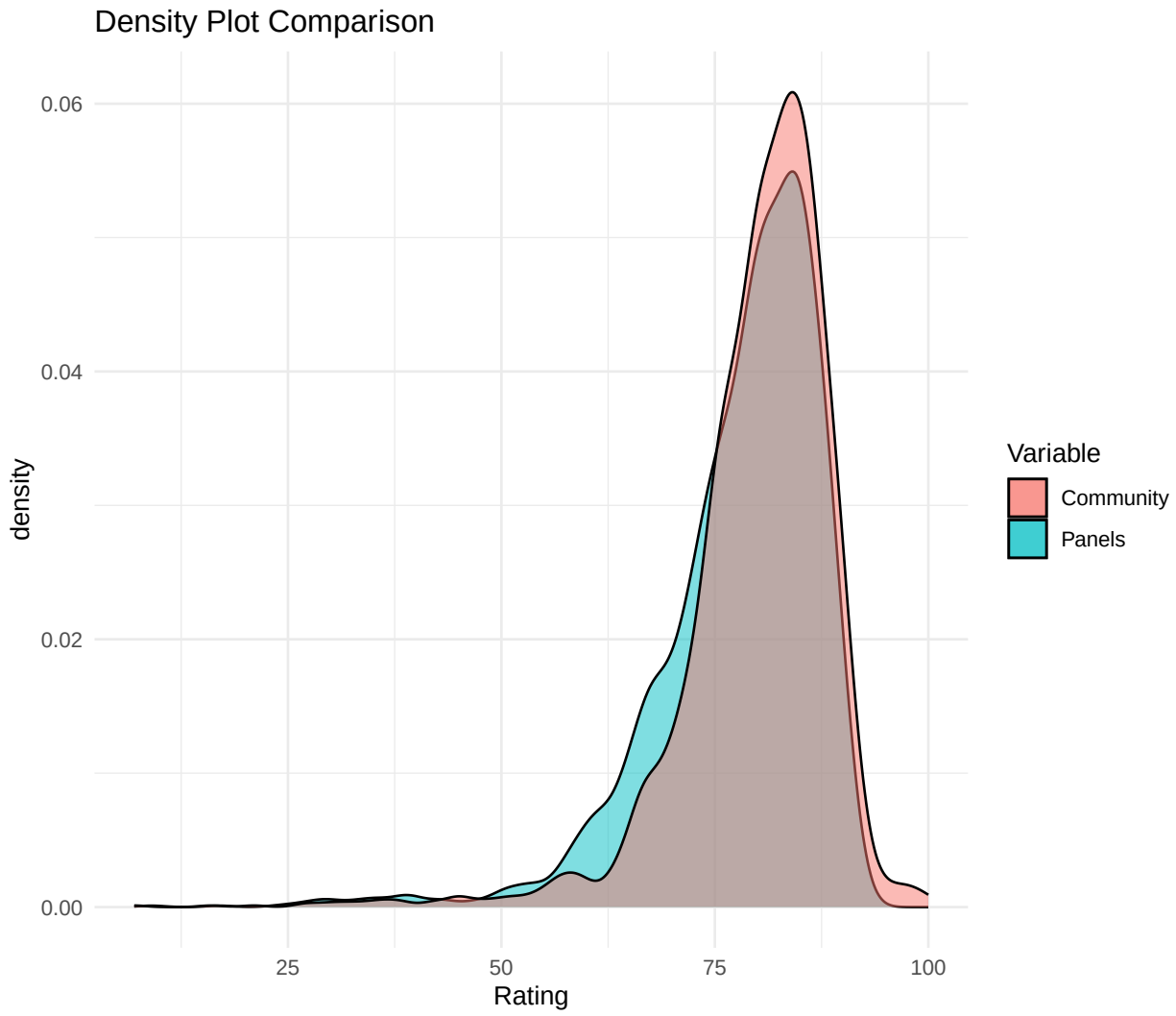


Table 11: Top 25 for Community Ratings

community	Region	title
100	Jalisco (Central)	Tequila Deleite Blanco
100	Jalisco (Los Altos Southern)	Atotonilco Reposado
100	Jalisco (Los Altos Southern)	Campo Azul Selecto Cristal Reposado
100	Jalisco (Los Altos Southern)	Celosa Rose Joven Rosa
100	Jalisco (Los Altos Southern)	Clandestina Tequila Blanco
100	Jalisco (Los Altos Southern)	Country Class Tequila Blanco
100	Jalisco (Los Altos Southern)	Presa Vieja Tequila Blanco
100	Jalisco (Los Altos Southern)	La Pitaya Blanco
100	Jalisco (Los Altos Southern)	Fiesta Brava Reposado
100	Jalisco (Los Valles)	Black 50 Tequila Reposado
100	Jalisco (Los Valles)	Burgues Azteca Tequila Reposado
100	Jalisco (Los Valles)	El Guerrillero Tequila Reposado
100	Jalisco (Los Valles)	El Mirador Reposado
100	Jalisco (Los Valles)	Las Potrancas Joven
100	Jalisco (Los Valles)	Majalca Tequila Extra A<f1>ejo
100	Jalisco (Los Valles)	Mariangel Tequila Reposado
100	Jalisco (Los Valles)	Miramontes A<f1>ejo Reserva de Don Francisco
100	Jalisco (Los Valles)	Nock Tequila Plata Premium
100	Jalisco (Los Valles)	Pink Pig Cristalino A<f1>ejo
100	Jalisco (Los Valles)	Tequila de Don Jes<fa>s Blanco
100	Jalisco (Los Valles)	Tequilador Crema de Chocolate Belga
100	Jalisco (Los Valles)	Tridente Reposado
100	Jalisco (Los Valles)	V Rey Supremo Reposado
100	Jalisco (Los Valles)	Amanecer Ranchero Reposado
100	Jalisco (Los Valles)	Gran Ciervo A<f1>ejo

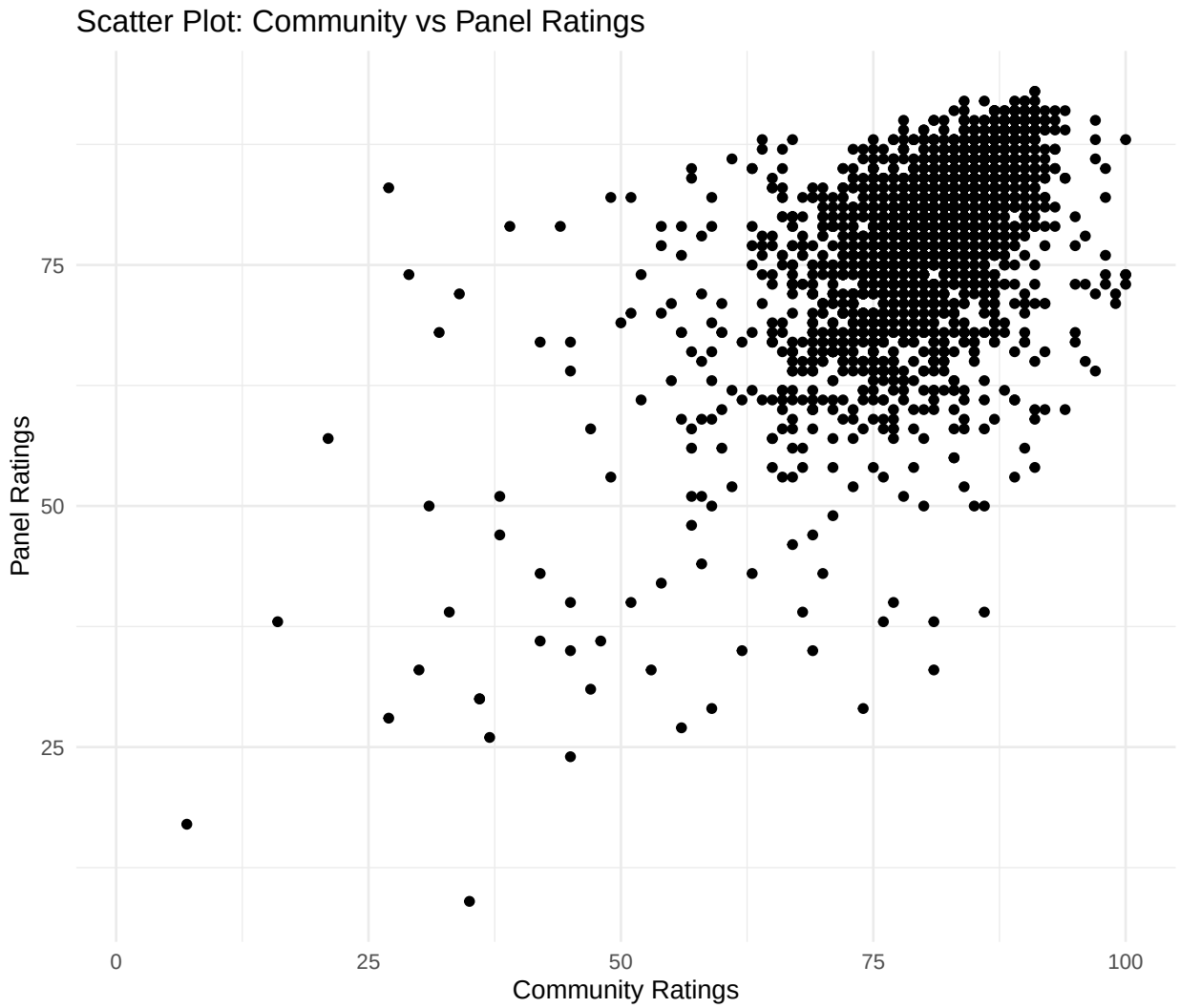
Let's explore the top 25 tequilas across various regions, as evaluated by community ratings. Notably, each of these tequilas boasts a perfect score of 100. Now, let's shift our focus to the Panel ratings and compare them to the community assessments.

We now explore the top 25 tequilas across various regions, as evaluated by community ratings. Notably, each of these tequilas boasts a perfect score of 100. Now lets shift our focus to the Panel ratings and compare them to the community assessments.

Table 12: Top 25 for Panel Ratings

Panel	Region	title
93	Jalisco (Los Altos Southern)	Real Hacienda <da>nico Blanco
93	Jalisco (Los Altos Southern)	De La Riva Tequila A<f1>ejo
93	Jalisco (Los Valles)	Los Abuelos A<f1>ejo
92	Jalisco (Los Altos Southern)	Rey Sol "La Cata" Single Barrel Extra A<f1>ejo
92	Jalisco (Los Altos Southern)	Rincon Alte<f1>o Extra A<f1>ejo
92	Jalisco (Los Altos Southern)	Romance Reposado-A<f1>ejo
92	Jalisco (Los Valles)	Puro Tequila One-Ten
92	Jalisco (Los Valles)	Cascahuin Cerro De Luz Blanco
92	Jalisco (Los Valles)	Fuenteseca Cosecha 2018 - Huerta "Las Antenas"
91	Jalisco (Cinega)	Calle 23 Criollo
91	Jalisco (Los Altos Southern)	Tequila Cabeza Sagrado Extra Anejo
91	Jalisco (Los Altos Southern)	Agave Dos Mil Grand Reserve Tequila Reposado
91	Jalisco (Los Altos Southern)	Tequila G4 Extra A<f1>ejo Reserva Especial
91	Jalisco (Los Altos Southern)	Pasote Extra A<f1>ejo
91	Jalisco (Los Altos Southern)	Tapatio Excelencia Extra A<f1>ejo (Lot 1, MX only)
91	Jalisco (Los Altos Southern)	Ocho Tequila A<f1>ejo - Cask Finish - Plantation Barbados
91	Jalisco (Los Altos Southern)	Volans Extra A<f1>ejo (6 Year)
91	Jalisco (Los Altos Southern)	KOKORO Tequila Limitada
91	Jalisco (Los Valles)	Berrueco Reposado Tequila
91	Jalisco (Los Valles)	El Caballito Cerrero Blanco (46%)
91	Jalisco (Los Valles)	El Caballito Cerrero A<f1>ejo 9 A<f1>os (53%)
91	Jalisco (Los Valles)	Caballito Cerrero Chato Reposado
91	Jalisco (Los Valles)	Caballito Cerrero Chato A<f1>ejo
91	Jalisco (Los Valles)	Tinta Negra Supreme Extra A<f1>ejo
91	Jalisco (Los Valles)	Cascahu<ed>n Extra A<f1>ejo

Diving into the realm of Panel ratings for the top 25 tequilas across diverse regions, an interesting pattern emerges. Unlike community ratings, none of these tequilas claim a perfect 100. The highest accolade stands at 93, subtly hinting at a trend where community evaluations tend to be more generous compared to their Panel counterparts.

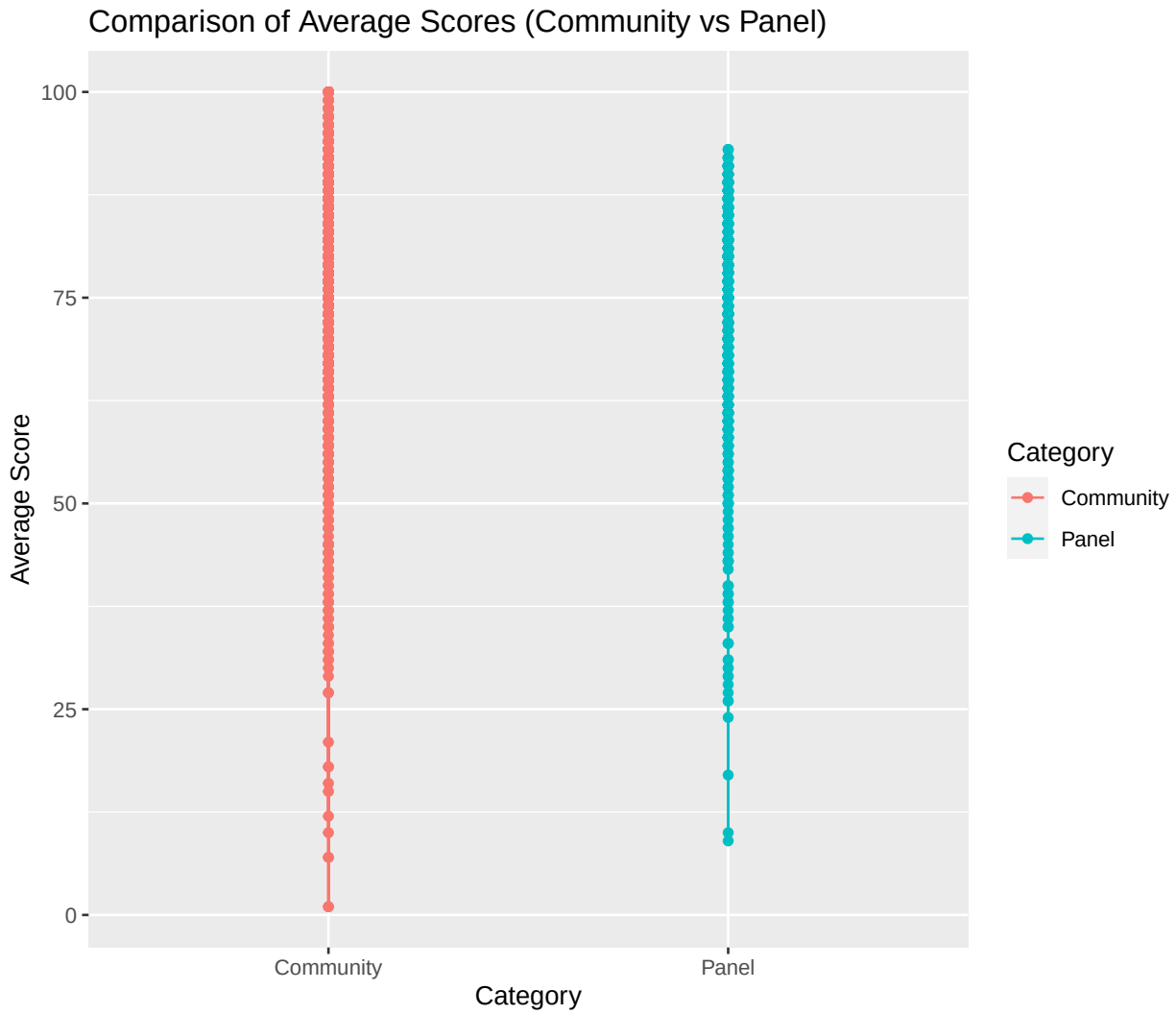


Now equipped with both panel and community ratings, we embark on a statistical journey, exploring the correlation between the two and unveiling the confidence intervals. These metrics will offer insights into the interplay of opinions from the tequila community and expert panels, enriching our understanding of the diverse perspectives surrounding the top 25 tequilas.

Here now we can see the results of 95% confidence interval and the correlation between the community and panel ratings.

The 95% confidence interval for the difference in means is: 4.561183 6.238817

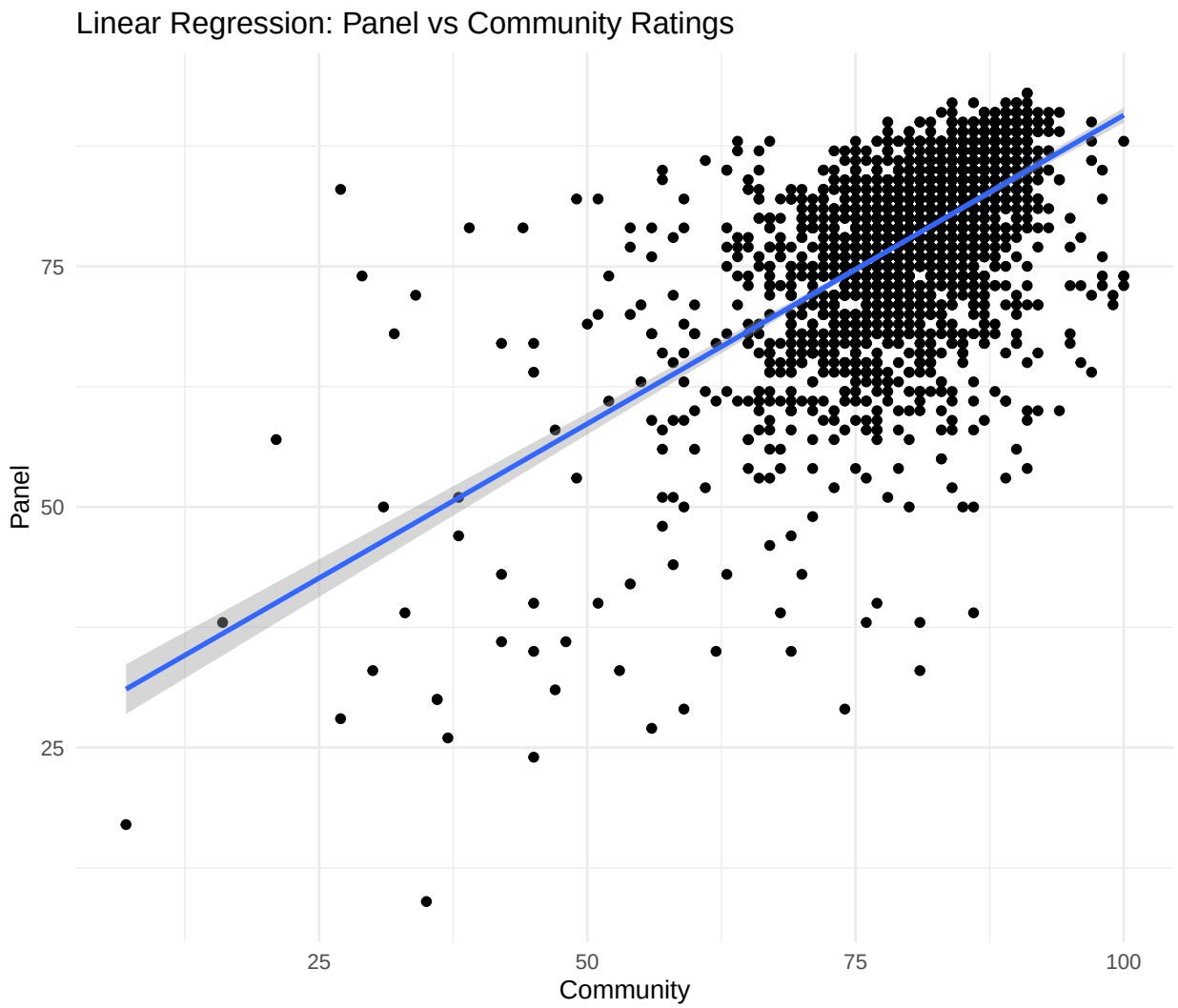
The correlation between community and panel ratings for the top 25 tequilas is: 0.7891467



In this expanded analysis, we delve into all tequilas, meticulously examining the Community and Panel ratings in tandem. The resulting diagram provides a comprehensive visualization of the rating distribution for each group. Interestingly, community ratings tend to be higher and more rating-intense, while panel ratings exhibit a broader spectrum, starting higher and ending lower, indicating a more mid-range intensity.

While this method may not be ideal for predicting future ratings, it serves as a valuable foundation. The observed patterns can inform adjustments to ratings, laying the groundwork for future predictive models that cater to the distinct preferences of both the community and the panel.

This is demonstrated in the linear regression graph with scatter plot to show how far out the Panel and Community would in predicting future ratings of Tequila.


Linear Regression: Panel vs Community Ratings
Panel
75
50
25
25
50
75
100
Community

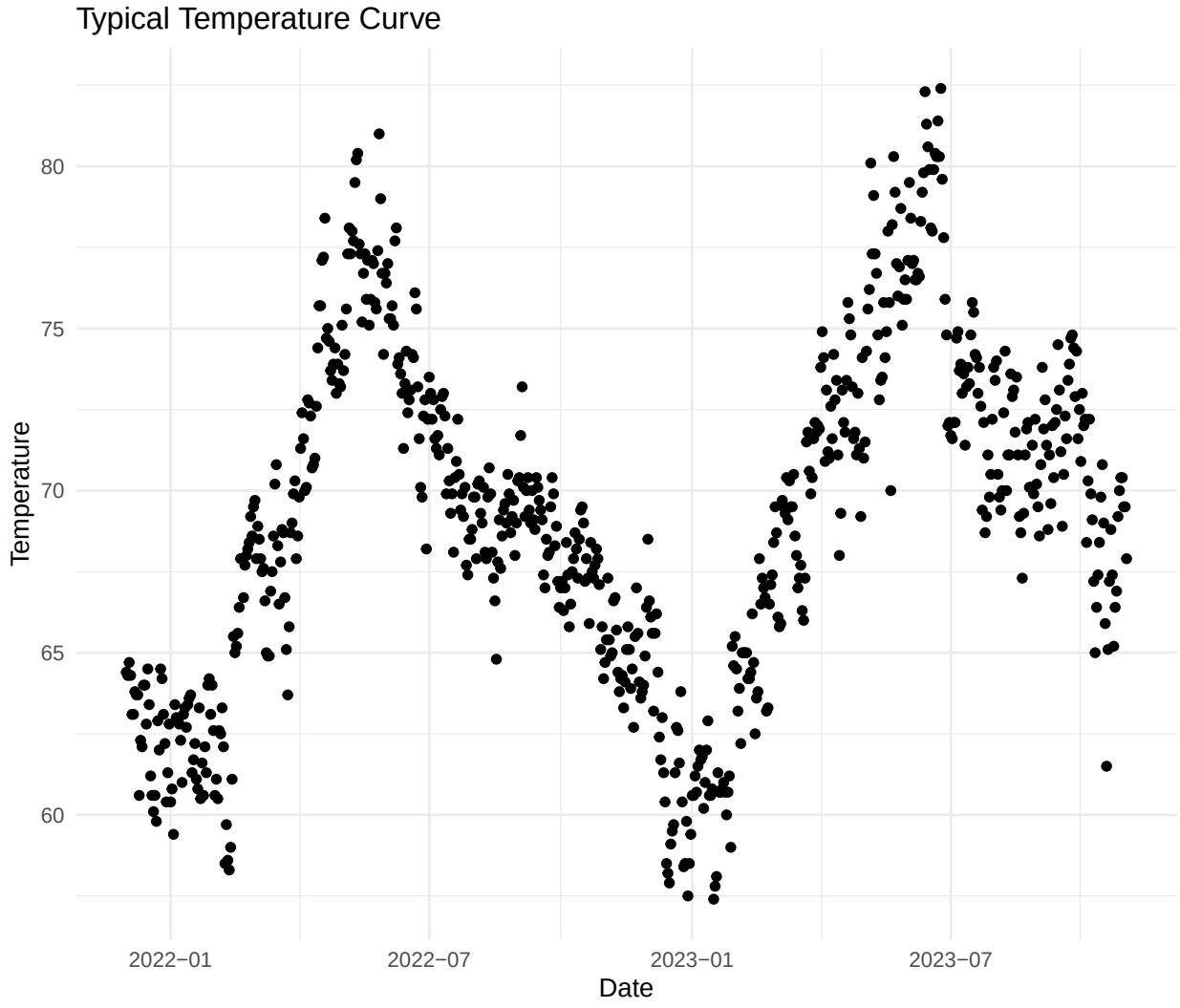
Now lets turn our focus to the weather and specifically the temperaure. What afect does temperature play in tequila? Imagine the sunlit landscapes where the resilient Blue agave, the maestro of tequila, unfolds its story. In the gentle warmth, it gracefully grows and matures, reaching the pinnacle of sugar content for crafting exceptional tequila.

Temperature, the unseen artist, orchestrates this alchemy, coaxing forth elevated sugar levels, especially the prized fructose. This infusion of sweetness paints the tequila canvas with a symphony of flavors, a testament to the sun's embrace.

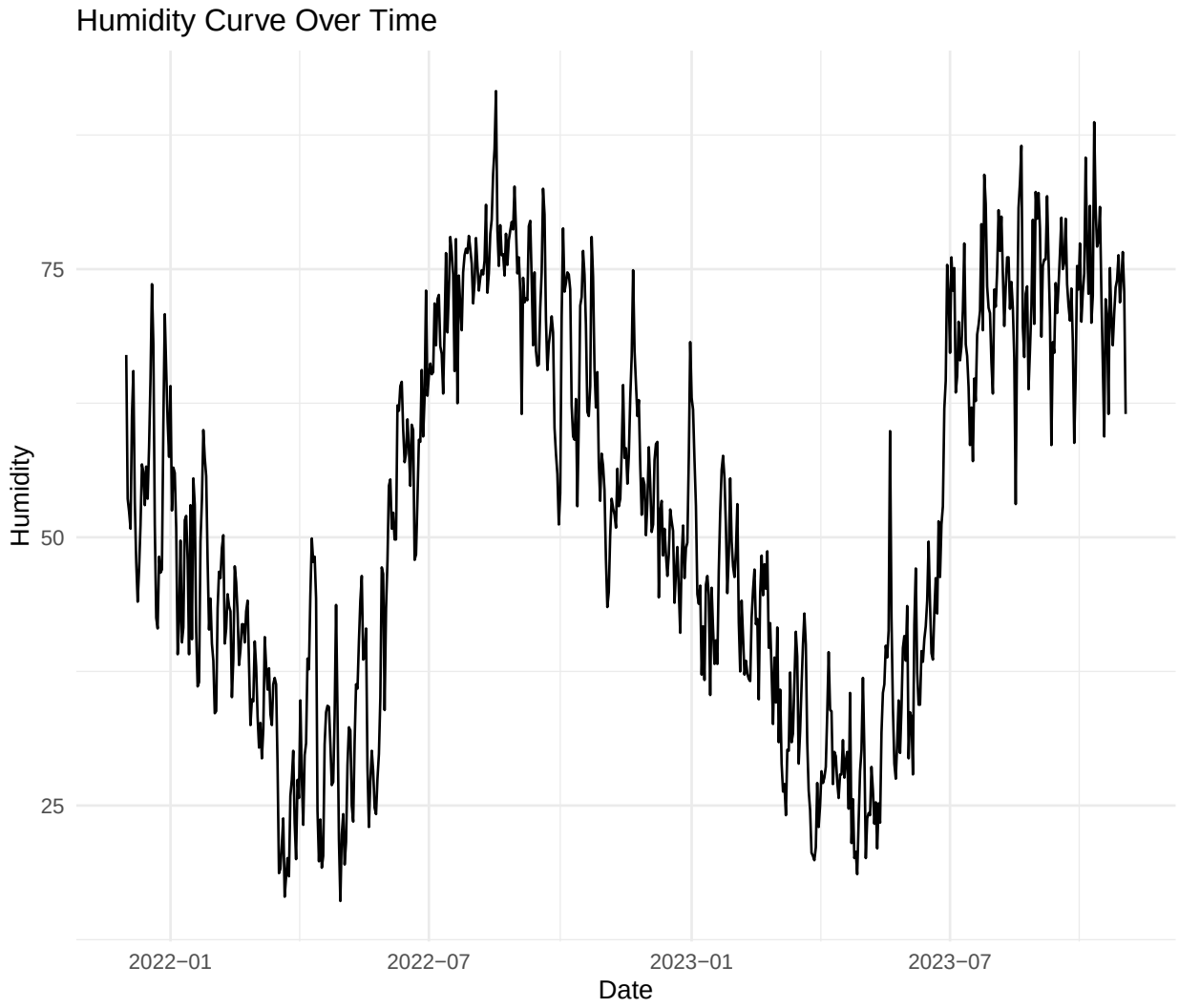
In this dance of nature, temperature becomes the choreographer, determining the timing of the agave harvest. A warmer touch quickens the maturation tempo, leading to a nuanced harvest precisely orchestrated to capture tequila's signature sweetness.

Temperature emerges as the silent conductor, guiding agave growth, sculpting sugar content, and dictating the poetic timing of the harvest. Together, they compose the melody of tequila production, where each note resonates with the warmth of the sun and the artistry of the agave.

Lets plot what the typical temperature curve is in the Tequila production and agave growing regions.

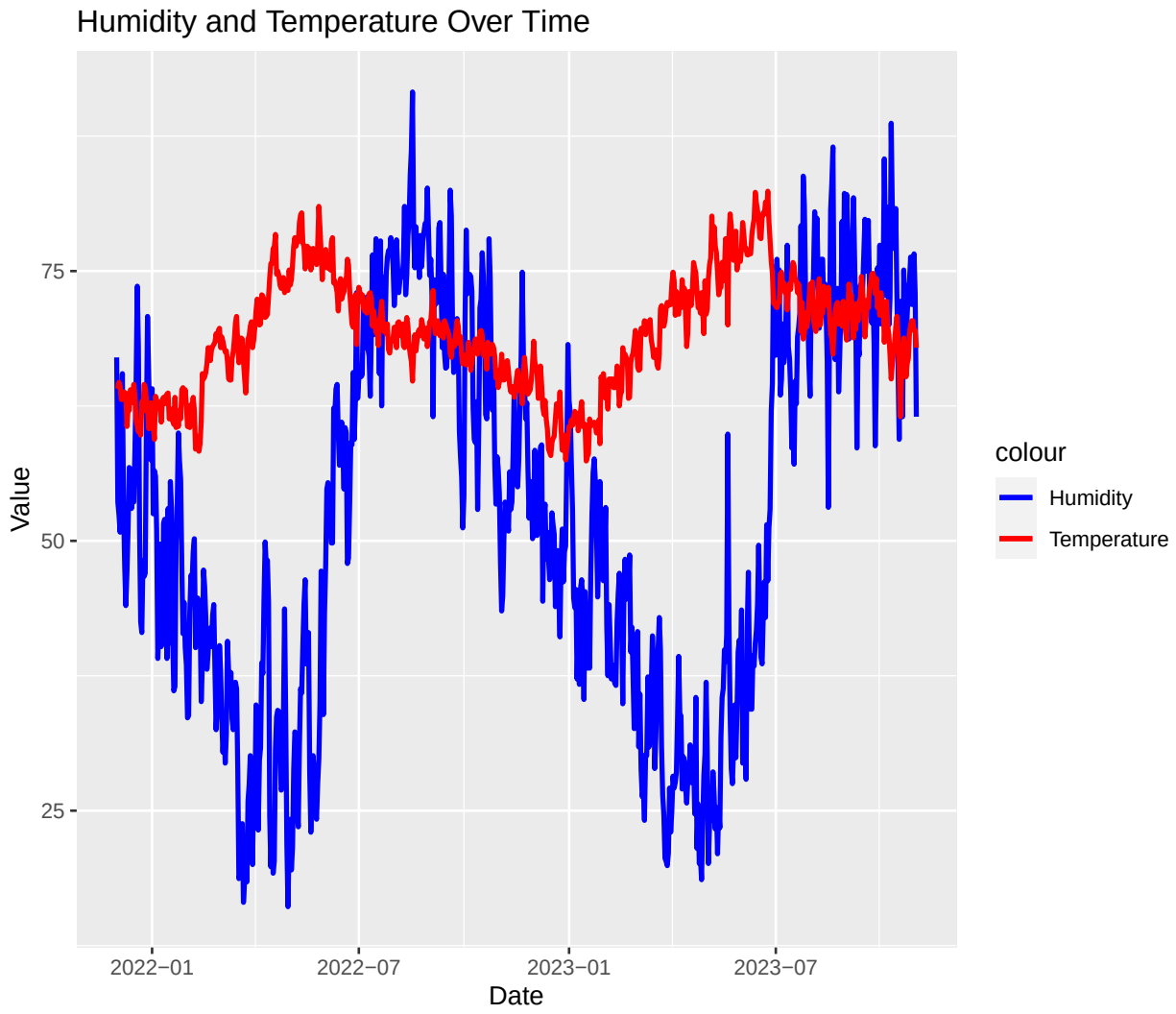


Now let us look at a similar graph but look at the humidity curve for the same time period.



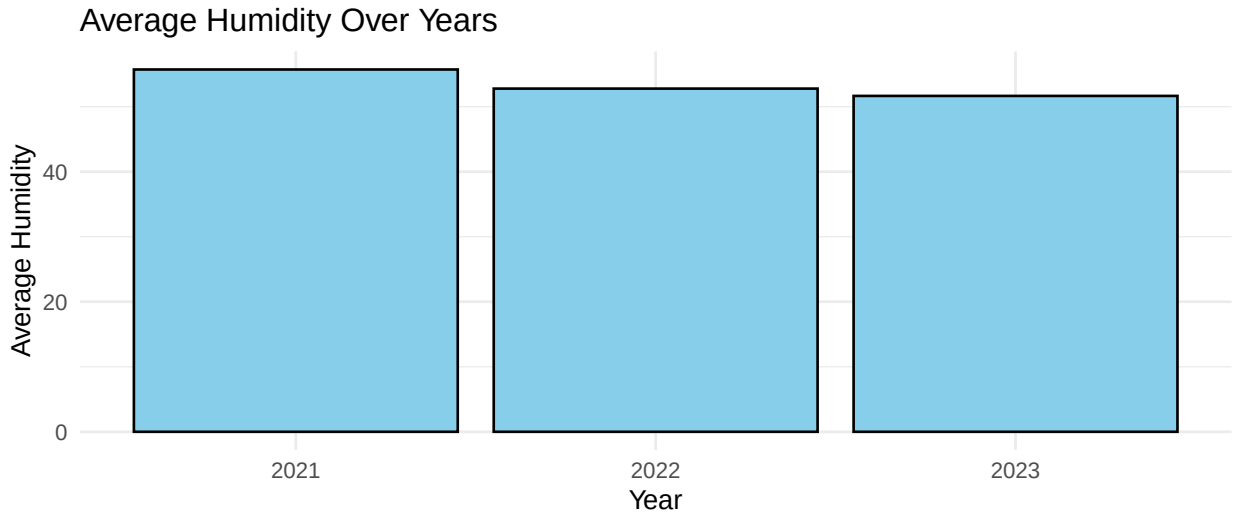
The effects of humidty on Agave growing is widespread. Humidity levels in the agave cultivation area can influence the growth of agave plants. Higher humidity may slow down evaporation, leading to a potentially slower maturation process for the agave. This could impact the sugar accumulation in agave plants, affecting the sweetness of the tequila.

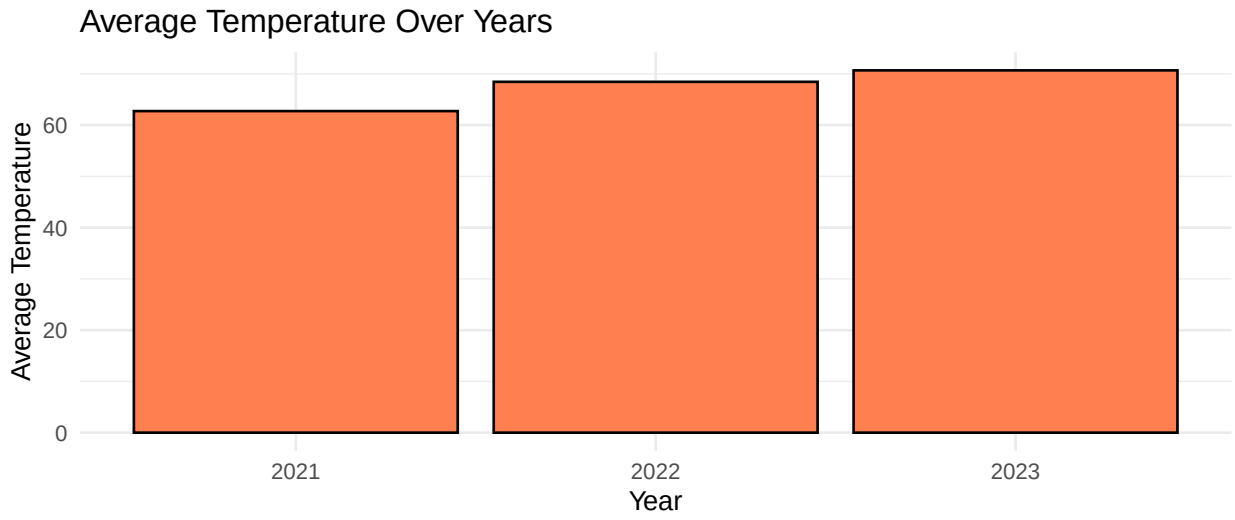
We can see that is no real correlation between humidity and temperature. If we notice anything we do see that as the temperature goes up the humidity trends the opposite direction.

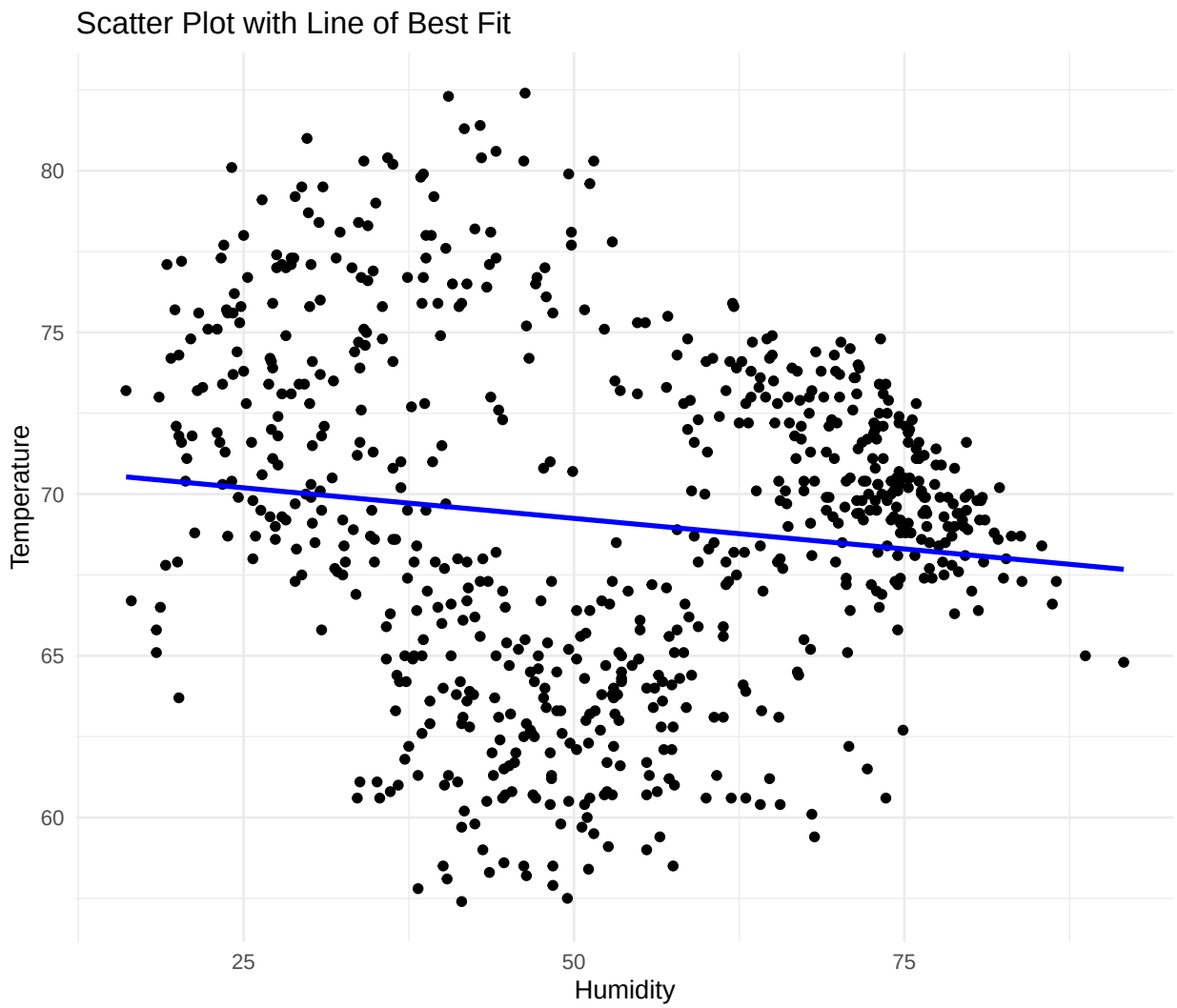


Now we should look at the correlation between the average humidity and temperature for the previous three years.

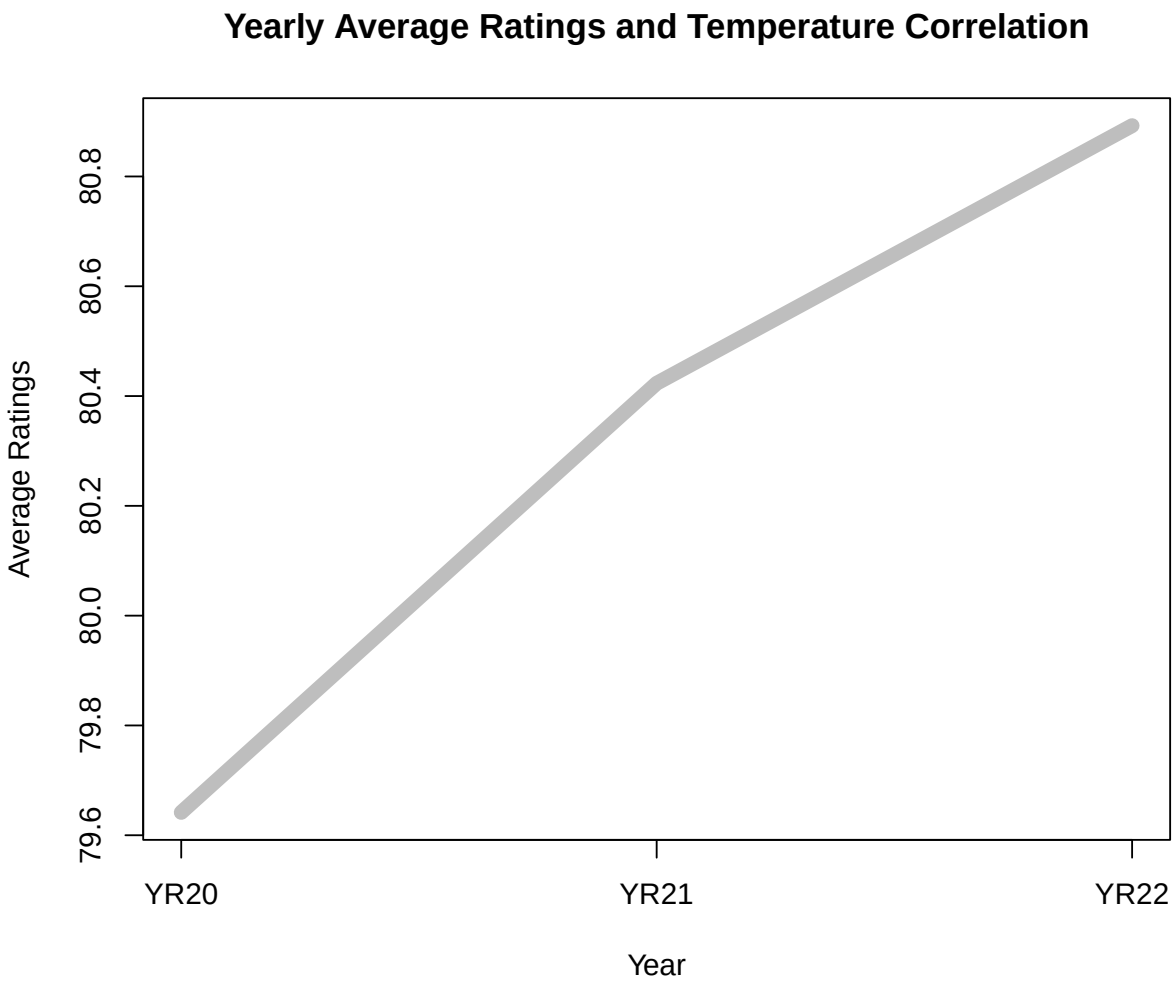
```
## [1] "2021" "2022" "2023"
```







This graph shows there is no correlation between humidity and temperature and ratings.



In exploring the relationship between temperature and yearly ratings, we've uncovered a noteworthy result—the correlation coefficient.

The correlation coefficient between average temperature and average ratings is 0.9945449.

This coefficient is essential in understanding the strength and nature of the relationship between the two variables. Here's a quick guide to interpreting correlation results:

Correlation Coefficient = 0: No discernible relationship. Changes in one variable don't predict changes in the other.

Correlation Coefficient = 0.6: Indicates a moderate positive relationship.

Correlation Coefficient = 0.8: Points to a fairly strong positive relationship.

Correlation Coefficient = +1: Represents a perfect positive relationship.

These insights help us grasp the dynamics between temperature and ratings. We can without question see that the temperature and the ratings of the Tequila are indefinately linked.

Results

RMSE Results

Display the RMSE Results

```
## [1] "Root Mean Squared Error (RMSE) between Temperature and Ratings: 0.436586103219728"
```

```
## [1] "Root Mean Squared Error (RMSE) between Panel and Predicted Ratings: 0.431024547450851"
```

NMRSE Results

```
## There are missing values in the AC20 column. Handling missing values...
## Range of AC20 column after handling missing values:
## [1]  21 100
```

```
## There are missing values in the AP20 column. Handling missing values...
## Range of AP20 column after handling missing values:
## [1]  9 95
```

Display the NRMSE values

In our analysis, we found the NRMSE for Average Ratings and Panel Ratings (nrmse_ac) to be 0.02093522 and 0.004396397, respectively. Similarly, the NRMSE for Panel and Predicted Ratings (nrmse_ap) was calculated to be 0.04789162 and 0.0045371.

Table 13: Top 5 Tequilas and Ratings

	Tequila Title	Panel Rating	Community Rating	Average Rating
1651	NYLA Organic Tequila Blanco	88	100	94.0
1933	Tequila Aguila Superior Reposado	90	97	93.5
204	Penca Azul Extra Aejo	88	97	92.5
2194	Volans Extra Aejo (6 Year)	91	94	92.5
389	Real Hacienda nico Blanco	93	91	92.0

We should take a look at the ratings of the top tequilas and the projected values.

Table 14: Top 25 Tequilas with Highest ACT23 Ratings and PROJ23 Ratings

	Tequila Title	ACT23	PROJ23
458	Cascahuin Cerro De Luz Blanco	95	96
1400	Los Abuelos Aejo	95	95
460	Cascahun Aniversario Blanco	94	94
874	El Caballito Cerrero Aejo 9 Aos (53%)	94	95
1579	Nueveuno Tequila Aejo	94	95
2259	Toro Dorado Aejo	94	94
2394	YaVe Coconut Tequila	94	94
319	Caballito Cerrero Chato Aejo	93	94
421	Casa Noble Alta Belleza Extra Aejo	93	93
466	Cascahun Tahona Blanco	93	93
1153	Gran Dovejo Blanco High Proof	93	93
1403	Los Abuelos Reposado	93	93
1669	Patrn Cask Collection Sherry Aejo (2017)	93	93
1749	Puro Tequila One-Ten	93	93
1918	Siembra Azul Elisa Extra Aejo	93	93
1945	Siete Leguas Blanco	93	94
2300	Tres Vidas Organic Tequila Aejo	93	93
2354	Viva Los Sanchos Aejo	93	93
7	123 Organic Tequila Diablito Rojo Extra Aejo	92	92
22	1824 Blanco	92	92
225	ArteNOM Seleccin de 1123 Blanco	92	92
280	Berrueco Reposado Tequila	92	92
876	El Caballito Cerrero Blanco (46%)	92	92
1016	El Tesoro de Don Felipe 75th Anniv	92	92
1100	Fortaleza Aejo	92	92

Conclusion

In conclusion, this project embarked on a quest to anticipate future tequila ratings and discern the key factors influencing these ratings. The journey led us to a remarkable correlation coefficient of 0.9945449 between average temperature and ratings, signifying an almost perfect relationship.

The dynamic nature of this report hinges on diverse data sources that constantly evolve in the live version. Employing various Data Wrangling techniques, I gathered information from a myriad of tequila, wine, and spirit websites, incorporating reviews and ratings. Additionally, data from multiple weather sites converges into a central weather sheet, ensuring an ever-expanding dataset.

The obtained RMSE results are indeed commendable:

Root Mean Squared Error (RMSE) between Temperature and Ratings: 0.436586103219728 Root Mean Squared Error (RMSE) between Panel and Predicted Ratings: 0.431024547450851 Upon scrutinizing the projected tequila ratings, none deviated by more than a rounded 1 point from their anticipated average in 2023. It's crucial to acknowledge the potential variability in these results, considering the ongoing influx of 2023 data, updated monthly for reviews.

Future Improvements

In the coming phases of this project, I aim to enhance the precision of the predictive model by incorporating a user-driven rating bias. By analyzing users' preferences for specific Tequila characteristics, such as aromas and flavors, I intend to discern patterns that correlate with higher ratings. This approach will enable me to tailor the predictions based on the distinct tastes of individual users.

Furthermore, I plan to introduce a feature that allows for a granular breakdown of reviews by the year of assessment and the moment the Tequila is introduced to the market. This temporal segmentation will provide valuable insights into how perceptions and preferences evolve over time.

Expanding the repository of weather information is a pivotal enhancement on the horizon. Given the lengthy growth cycle of agaves, spanning 7 to 12 years, a comprehensive understanding of weather conditions during this period will be instrumental in refining the accuracy of predicted ratings. The additional weather data will contribute to a more nuanced and informed prediction model, capturing the intricacies of agave cultivation and its impact on Tequila quality.

Technology Improvements

Concerning technology improvements, I will continue to run the automated web scraping and data wrangling processes to consistently expand the database. Enlarging the dataset opens up the possibility of training distinct sets—one for panel ratings and another for community ratings.

Furthermore, I plan to enhance the functionalities related to the production process, particularly focusing on the diverse aromas and flavors associated with tequilas. An illustrative example can be found below. With an increasing number of individuals providing ratings on the aromas and tastes of tequilas, this category has the potential to evolve into a new avenue for exploration and study.

	Aroma	Count
Roller Mill	Agave (cooked)	838
Tahona	Agave (cooked)	65
Diffuser	Agave (cooked)	40
Screw Mill	Agave (cooked)	36
Screw Mill Roller Mill	Agave (cooked)	23
Tahona Roller Mill	Agave (cooked)	18
Roller Mill Screw Mill	Agave (cooked)	9
Roller Mill Tahona	Agave (cooked)	7
Tahona Screw Mill	Agave (cooked)	7
Diffuser Roller Mill	Alcohol	3
Roller Mill Diffuser	Agave (cooked)	3
Diffuser Screw Mill	Peppermint	1
Hand Crushed	Smoke	1

Resources and References

1. Rafael Irizarry. 2018. Introduction to Data Science.
2. Jared Lander, 2017, R for Everyone Advanced Analytics and Graphics.
3. Norman Matloff, 2011 & 2019, The Art of R Programming