# Análisis Inicial

Contents

[Descripción del Dataset Winemag 1](#_Toc57567817)

[Modificación del Dataset 3](#_Toc57567818)

[Análisis de histogramas por Variables 3](#_Toc57567819)

[Países (Countries) 3](#_Toc57567820)

[Calificacion (Points) 5](#_Toc57567821)

[Precio (Price) 6](#_Toc57567822)

[Provincias. 8](#_Toc57567823)

[Region 9](#_Toc57567824)

[Catador (Taster Name) 11](#_Toc57567825)

[Nombre/Título del Vino 12](#_Toc57567826)

[Varietal 14](#_Toc57567827)

[Casa/Vinoteca 16](#_Toc57567828)

[Correlaciones 18](#_Toc57567829)

[Inferencias 18](#_Toc57567830)

## Descripción del Dataset Winemag

* ID → Referencia general de cada entrada del dataset
* Country → Describe el país de origen del vino
* Description → Una descripción general del catador sobre el vino
* Designation → Describe el tipo de vino, si es seco, semi-seco, dulce, etc.
* Points → Calificación que le da el catador al vino
* Price → El precio en dólares del vino
* Province → Información relacionada a la provincia/estado donde el vino fue cultivado y procesado
* Region\_1 → Información adicional donde el vino fue cultivado
* Region\_2 → Información adicional donde el vino fue cultivado
* Taster Name → Nombre del catador que calificó el vino
* Taster Twitter → Twitter del catador del vino
* Title → Nombre (título) del vino
* Variety → (Varietal) Tipo de uva del vino
* Winery → Casa / Vinoteca que creó el vino

Para el objetivo del análisis del dataset de datos estructurados y semi-estructurados, inicialmente tenemos que describir cada variable, y su utilidad en el objetivo:

1. ID: Es solo un identificador general de las entradas. No tiene implicación alguna en el objetivo más que llevar un conteo general de cada entrada valida
   1. Tipo de Variable: Integer
   2. Rango: 1 ↑ …
   3. Valores Nulos: No existen
   4. Dato: Estructurado
2. Country: Es el identificador por país. Tienen implicación media en el objetivo, ya que determina que países son los mejores productores de vinos, correlacionado al Varietal y Región
   1. Tipo de Variable: String
   2. Rango: Países
   3. Valores Nulos: Existen. Para estos casos se debe correlacionar la región o no es válido
   4. Dato: Semi-Estructurado
3. Description: Es la descripción dada por el catador. En este momento no será usado este dato. Es posible que sea usado más adelante en un análisis de palabras
4. Designation: La designación no será usada en este momento. No entrega mayor información.
5. Points: Es la calificación. Tiene una implicación alta, ya que nos permitirá analizar las calificaciones por País/Provincia/Región. También nos permitirá analizar si el Varietal tiene alguna correlación con la calidad del vino.
   1. Tipo de Variable: Integer
   2. Rango: 80 ↔ 100
   3. Valores Nulos: No Existen
   4. Dato: Estructurado
6. Price: Es el valor del vino. Tiene una implicación baja. Se podrá usar para inferir si la región tiene que ver con el precio del producto y su relación con la calidad.
   1. Tipo de Variable: Integer
   2. Rango: 4 ↔ 3300 + Null
   3. Valores Nulos: Como existen valores nulos, se intentará realizar una búsqueda de los precios de los vinos a través de su nombre.
   4. Dato: Estructurado
7. Provincia: Región del vino. Tiene una implicación alta. Se requiere de este dato para poder inferir los datos finales de relación calidad-región-clima. Si no se tiene este dato, se tratará de inferir de los campos de País/Region\_1/Region\_2. De no encontrar datos se intentará realizar una búsqueda de la región a través de su nombre.
   1. Tipo de Variable: Text
   2. Rango: Undefined
   3. Valores Nulos: Como existen valores nulos, se intentará realizar una búsqueda de la región de los vinos a través de su nombre.
   4. Dato: Semi-Estructurado
8. Region1/Region2: Datos más precisos de la región donde fue creado el vino. Se usarán en caso de apoyo para determinar una región más exacta. Tiene una implicación media, como ayuda para realizar una geolocalización más precisa de la región donde fue creado el vino.
   1. Tipo de Variable: Text
   2. Rango: Undefined
   3. Valores Nulos: Se depreciará cualquier valor nulo
   4. Dato: Semi-Estructurado
9. Taster\_Name: Nombre del catador. Solo se tendrá en cuenta para encontrar alguna relación de cantidad de calificaciones por catador. Implicación baja
   1. Tipo de Variable: Text
   2. Rango: Undefined
   3. Valores Nulos: Como existen, se obviará ese dato (la entrada será aun válida)
   4. Dato: Semi-Estructurado
10. Taster\_Twitter: Se depreciará este dato. Implicación Nula.
11. Title: Nombre del Vino. Implicación baja. Se utilizará el dato como valor de apoyo para obtener información nula de otras casillas (País, Región, Varietal, Casa/Vinoteca, Precio).
    1. Tipo de Variable: Text
    2. Rango: Undefined
    3. Valores Nulos: No Existen
    4. Dato: Semi-Estructurado
12. Variety: Varietal del Vino. Implicación media. Se utilizará obtener alguna correlación del tipo de Uva, con la región y la calidad.
    1. Tipo de Variable: Texto
    2. Rango: Undefined
    3. Valores Nulos: Existe en un caso. Se usará el título para completar este dato
    4. Dato: Semi-Estructurado
13. Winery: Es la casa productora del vino. Implicación Baja. Se usará solo como dato de apoyo para obtener información nula de otras casillas (País, Región)

Con esto se usarán los siguientes campos:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **ID** | **country** | **points** | **price** | **province** | **region\_1** | **region\_2** | **taster\_name** | **title** | **variety** | **winery** |
| Importancia | ↓ | ↔ | ↑ | ↔ | ↑ | ↔ | ↔ | ↓ | ↔ | ↔ | ↔ |

## Modificación del Dataset

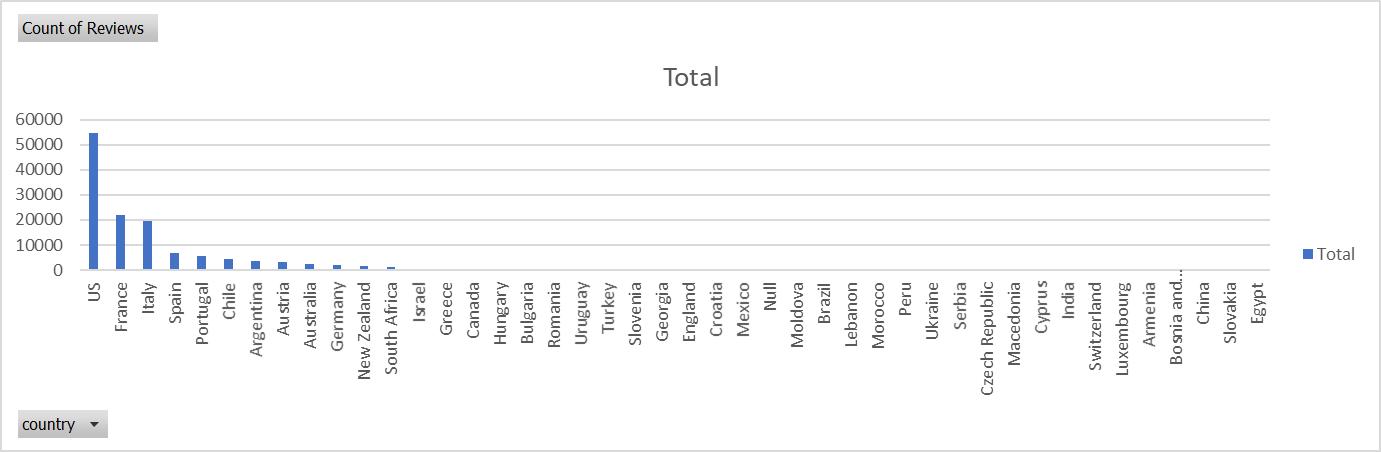
El primer paso es cambiar todos los valores nulos por el valor Null, excepto por los valores numéricos que tomarán el valor 0.

## Análisis de histogramas por Variables

### Países (Countries)

En la siguiente tabla se encuentra el conteo de datos por países

|  |  |  |
| --- | --- | --- |
| **Countries** | **Count of Reviews** | **Percentage Of Total** |
| US | 54504 | 41.94% |
| France | 22093 | 17.00% |
| Italy | 19540 | 15.03% |
| Spain | 6645 | 5.11% |
| Portugal | 5691 | 4.38% |
| Chile | 4472 | 3.44% |
| Argentina | 3800 | 2.92% |
| Austria | 3345 | 2.57% |
| Australia | 2329 | 1.79% |
| Germany | 2165 | 1.67% |
| New Zealand | 1419 | 1.09% |
| South Africa | 1401 | 1.08% |
| Israel | 505 | 0.39% |
| Greece | 466 | 0.36% |
| Canada | 257 | 0.20% |
| Hungary | 146 | 0.11% |
| Bulgaria | 141 | 0.11% |
| Romania | 120 | 0.09% |
| Uruguay | 109 | 0.08% |
| Turkey | 90 | 0.07% |
| Slovenia | 87 | 0.07% |
| Georgia | 86 | 0.07% |
| England | 74 | 0.06% |
| Croatia | 73 | 0.06% |
| Mexico | 70 | 0.05% |
| Null | 63 | 0.05% |
| Moldova | 59 | 0.05% |
| Brazil | 52 | 0.04% |
| Lebanon | 35 | 0.03% |
| Morocco | 28 | 0.02% |
| Peru | 16 | 0.01% |
| Ukraine | 14 | 0.01% |
| Serbia | 12 | 0.01% |
| Czech Republic | 12 | 0.01% |
| Macedonia | 12 | 0.01% |
| Cyprus | 11 | 0.01% |
| India | 9 | 0.01% |
| Switzerland | 7 | 0.01% |
| Luxembourg | 6 | 0.00% |
| Armenia | 2 | 0.00% |
| Bosnia and Herzegovina | 2 | 0.00% |
| China | 1 | 0.00% |
| Slovakia | 1 | 0.00% |
| Egypt | 1 | 0.00% |
| **Grand Total** | **129971** | **100.00%** |



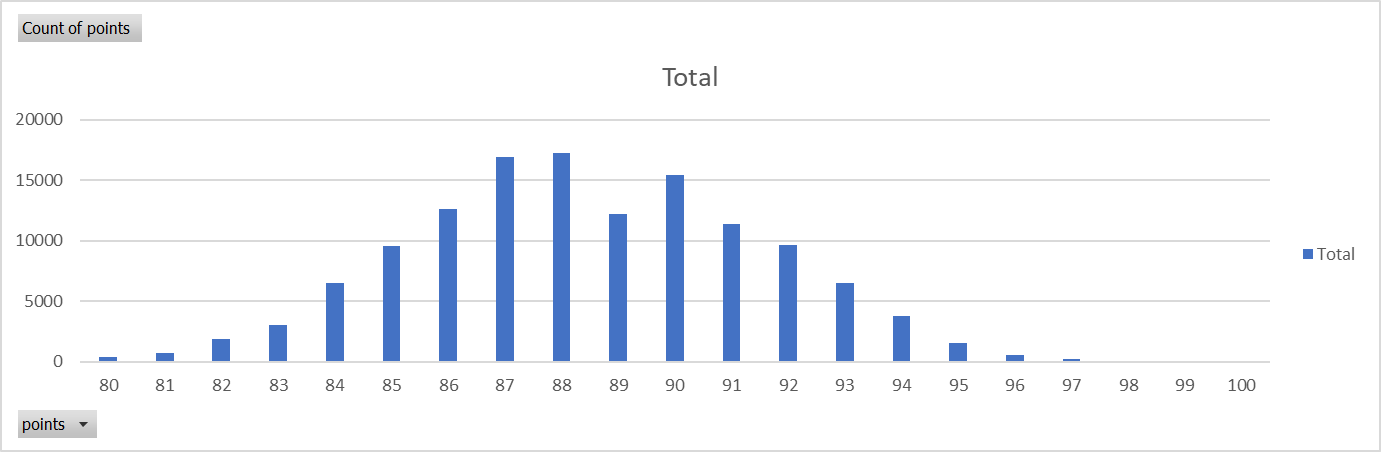
Esta tabla la lidera Estados Unidos con casi el 42% de las calificaciones, mientras otros países aparecen con valores muy bajos. Esto muestra que gran parte de nuestro análisis se basará en vinos de Norte América, pero no quiere decir que otros países con porcentajes bajos no puedan ser evaluados.

Países con muy pocos conteos de vinos no serán tomados en cuenta ya que se puede tener un conteo de calificaciones lo suficientemente alto para poder tener una inferencia valiosa. Para esto se tomará un análisis los países con un conteo igual o mayor a 100.

### Calificacion (Points)

El conteo e histograma de puntos se ve a continuación:

|  |  |
| --- | --- |
| **Points** | **Count of points** |
| 80 | 397 |
| 81 | 692 |
| 82 | 1836 |
| 83 | 3025 |
| 84 | 6480 |
| 85 | 9530 |
| 86 | 12600 |
| 87 | 16933 |
| 88 | 17207 |
| 89 | 12226 |
| 90 | 15410 |
| 91 | 11359 |
| 92 | 9613 |
| 93 | 6489 |
| 94 | 3758 |
| 95 | 1535 |
| 96 | 523 |
| 97 | 229 |
| 98 | 77 |
| 99 | 33 |
| 100 | 19 |
| **Grand Total** | **129971** |

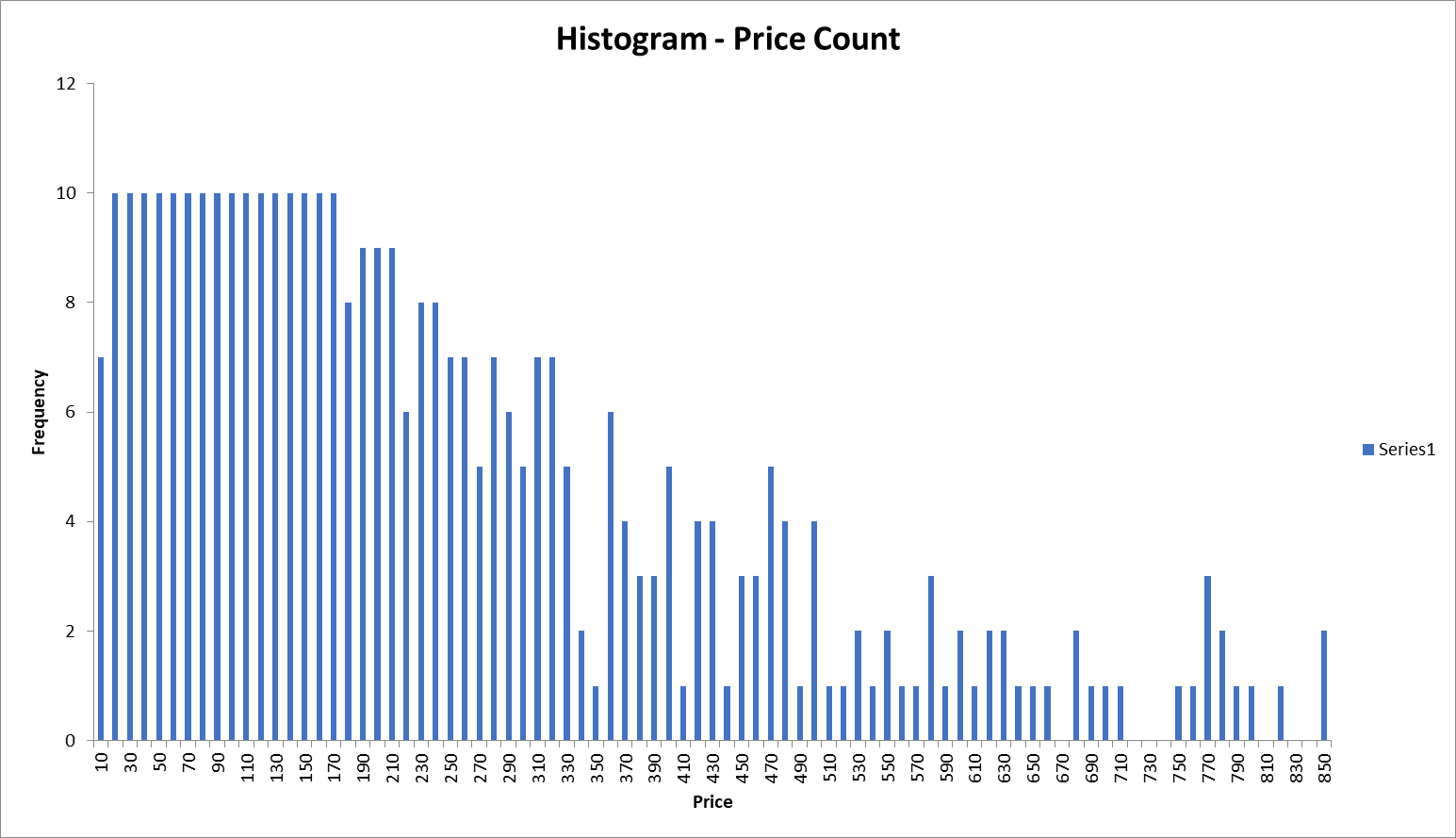


Como se puede observar, se encuentra una distribución similar a la distribución normal. Esto permitirá hacer ciertos cálculos con los valores estándar.

### Precio (Price)

La distribución de precios tiene una tendencia “Skew to the right”, donde la tendencia de los precios es mayor en la parte inferior del rango total.

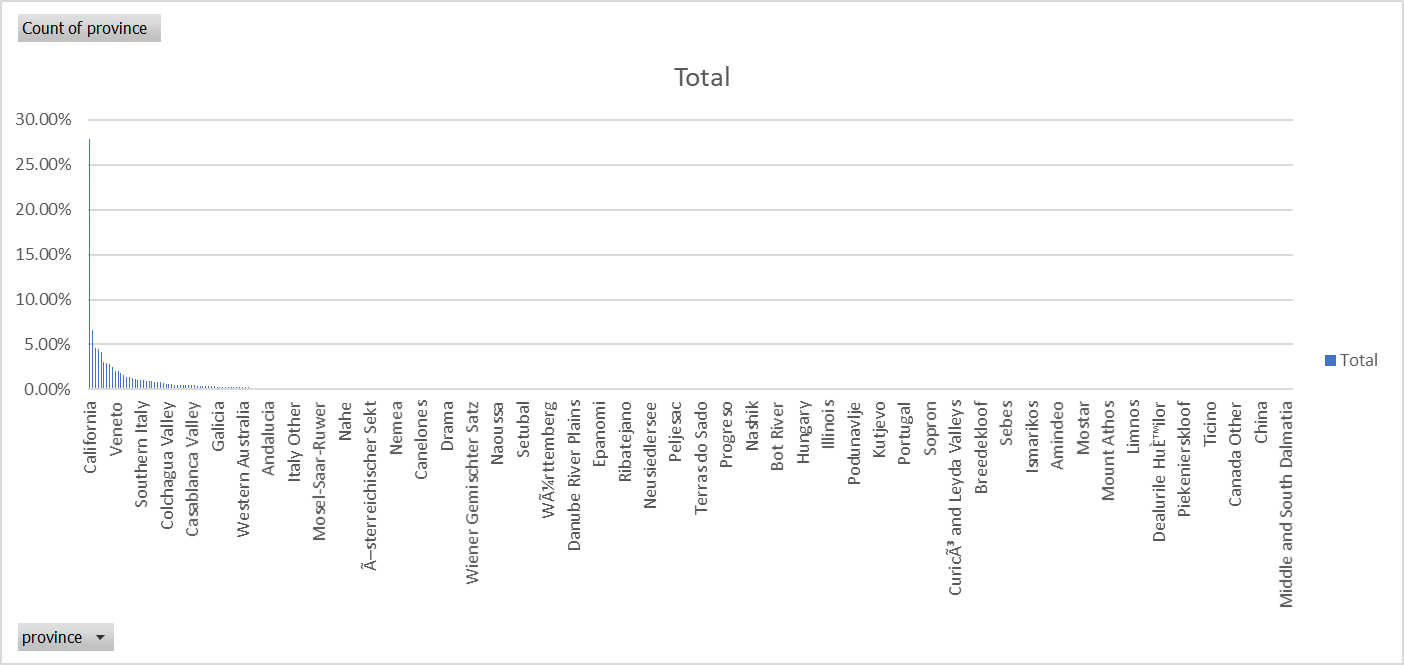
Como se puede ver en el histograma, la mayor parte de valores se encuentran hasta los 850 dólares aproximadamente, por lo cual, haciendo este mismo histograma en un menor rango, se tiene:



Esto nos muestra que la mayor parte de los vinos tienen un costo de 400 dólares o menor.

### Provincias.

Al realizar un histograma de este valor, se encuentran muchos datos que no ayudan al análisis final, ya sea por el nombre usado o por que es muy general. En este caso, lo que se propone es modificar este valor, por valores de longitud y latitud, que permitan ubicar con mayor facilidad la zona, y con rangos establecer zonas de análisis más generales o menos específicas, según sea el caso.



Principales 50 provincias de análisis

|  |  |
| --- | --- |
| Province | Percentage of Total |
| California | 27.89% |
| Washington | 6.65% |
| Bordeaux | 4.57% |
| Tuscany | 4.54% |
| Oregon | 4.13% |
| Burgundy | 3.06% |
| Northern Spain | 2.96% |
| Piedmont | 2.87% |
| Mendoza Province | 2.51% |
| Veneto | 2.09% |
| New York | 2.07% |
| Alsace | 1.88% |
| Northeastern Italy | 1.64% |
| Loire Valley | 1.43% |
| Sicily & Sardinia | 1.38% |
| Champagne | 1.24% |
| Southwest France | 1.16% |
| South Australia | 1.04% |
| Southern Italy | 1.04% |
| Provence | 1.04% |
| Douro | 0.99% |
| Central Italy | 0.95% |
| Catalonia | 0.90% |
| RhÃ´ne Valley | 0.83% |
| Beaujolais | 0.80% |
| Mosel | 0.80% |
| Alentejano | 0.72% |
| Colchagua Valley | 0.64% |
| Marlborough | 0.61% |
| Virginia | 0.60% |
| Tejo | 0.53% |
| France Other | 0.52% |
| Burgenland | 0.52% |
| Maipo Valley | 0.51% |
| Languedoc-Roussillon | 0.47% |
| Port | 0.46% |
| Casablanca Valley | 0.46% |
| NiederÃ¶sterreich | 0.46% |
| Other | 0.41% |
| Lombardy | 0.41% |
| Lisboa | 0.39% |
| Central Spain | 0.37% |
| Maule Valley | 0.36% |
| Levante | 0.35% |
| Stellenbosch | 0.34% |
| Galicia | 0.32% |
| Vinho Verde | 0.32% |
| Kamptal | 0.31% |
| Central Valley | 0.27% |
| Wachau | 0.27% |

### Region

Para la parte de Region\_1, da una mejor información sobre donde están ubicados los productores de vino. Aunque hay un 16,35% de valores nulos que toca complementar a través de otros datos, las demás regiones están demarcadas, lo que permite hacer un mejor análisis de detalle de la región/zona de producción del vino. Así mismo, se encuentra que Región\_2 no sumó mayores detalles, por lo cual no se utilizará en el respectivo análisis.

50 primeros valores de regiones

|  |  |
| --- | --- |
| **Region** | **Percentage of Total Count per Region** |
| Null | 16.35% |
| Napa Valley | 3.45% |
| Columbia Valley (WA) | 3.17% |
| Russian River Valley | 2.38% |
| California | 2.02% |
| Paso Robles | 1.81% |
| Willamette Valley | 1.77% |
| Mendoza | 1.77% |
| Alsace | 1.66% |
| Champagne | 1.24% |
| Barolo | 1.23% |
| Finger Lakes | 1.20% |
| Sonoma Coast | 1.13% |
| Brunello di Montalcino | 1.13% |
| Rioja | 1.13% |
| Sonoma County | 0.96% |
| Toscana | 0.92% |
| Walla Walla Valley (WA) | 0.82% |
| Chianti Classico | 0.82% |
| Santa Barbara County | 0.77% |
| Carneros | 0.77% |
| Yakima Valley | 0.77% |
| Sta. Rita Hills | 0.75% |
| Dry Creek Valley | 0.72% |
| Sicilia | 0.71% |
| CÃ´tes de Provence | 0.66% |
| Santa Ynez Valley | 0.62% |
| Santa Lucia Highlands | 0.62% |
| Lodi | 0.62% |
| Barbaresco | 0.61% |
| Alexander Valley | 0.60% |
| Central Coast | 0.58% |
| Ribera del Duero | 0.56% |
| Red Mountain | 0.56% |
| Santa Maria Valley | 0.54% |
| Alto Adige | 0.53% |
| Sonoma Valley | 0.51% |
| Anderson Valley | 0.50% |
| North Fork of Long Island | 0.46% |
| Saint-Ã‰milion | 0.43% |
| Horse Heaven Hills | 0.43% |
| Chablis | 0.43% |
| Virginia | 0.43% |
| Sancerre | 0.43% |
| Dundee Hills | 0.43% |
| Santa Cruz Mountains | 0.42% |
| Bordeaux | 0.40% |

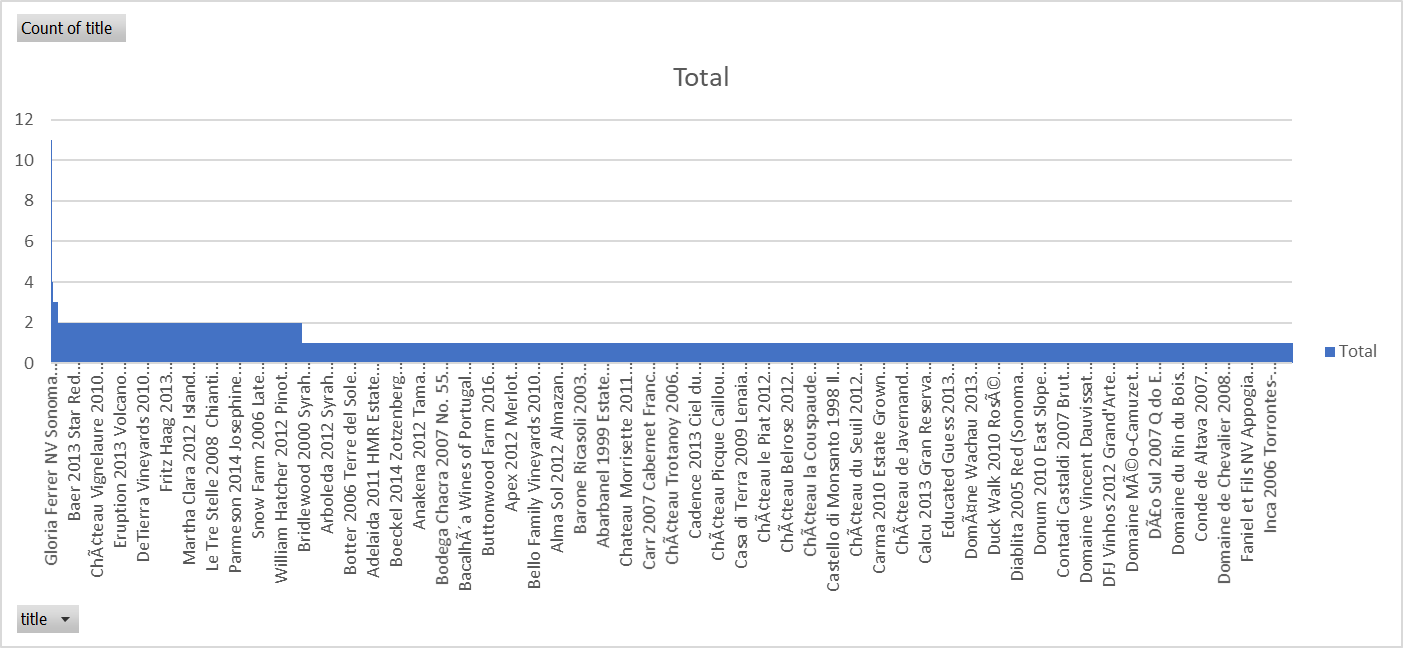
### Catador (Taster Name)

En este análisis, se encuentra que muchas de las calificaciones no tienen un nombre del catador, pero los demás valores se encuentran entre catadores renombrados, con un conteo bajo de catadores, como se muestra a continuación:

|  |  |
| --- | --- |
| **Points** | **Count of taster\_name** |
| Null | 26244 |
| Roger Voss | 25514 |
| Michael Schachner | 15134 |
| Kerin Oâ€™Keefe | 10776 |
| Virginie Boone | 9537 |
| Paul Gregutt | 9532 |
| Matt Kettmann | 6332 |
| Joe Czerwinski | 5147 |
| Sean P. Sullivan | 4966 |
| Anna Lee C. Iijima | 4415 |
| Jim Gordon | 4177 |
| Anne KrebiehlÂ MW | 3685 |
| Lauren Buzzeo | 1835 |
| Susan Kostrzewa | 1085 |
| Mike DeSimone | 514 |
| Jeff Jenssen | 491 |
| Alexander Peartree | 415 |
| Carrie Dykes | 139 |
| Fiona Adams | 27 |
| Christina Pickard | 6 |
| **Grand Total** | **129971** |

### Nombre/Título del Vino

Como era de esperarse, los nombres de los vinos no se repiten mucho. Es un estudio general donde solo pocos vinos son calificados múltiples veces, como se muestra a continuación:



Conteo de calificaciones de vinos, con 4 calificaciones o más (77 entradas)

|  |  |
| --- | --- |
| **Points** | **Count of title** |
| Gloria Ferrer NV Sonoma Brut Sparkling (Sonoma County) | 11 |
| Korbel NV Brut Sparkling (California) | 9 |
| Segura Viudas NV Extra Dry Sparkling (Cava) | 8 |
| Gloria Ferrer NV Blanc de Noirs Sparkling (Carneros) | 7 |
| Ruinart NV Brut RosÃ© (Champagne) | 7 |
| Segura Viudas NV Aria Estate Extra Dry Sparkling (Cava) | 7 |
| Bailly-Lapierre NV Brut (CrÃ©mant de Bourgogne) | 6 |
| Boizel NV Brut RÃ©serve (Champagne) | 6 |
| Jacquart NV Brut MosaÃ¯que (Champagne) | 6 |
| J Vineyards & Winery NV Brut RosÃ© Sparkling (Russian River Valley) | 6 |
| Korbel NV Sweet RosÃ© Sparkling (California) | 6 |
| Pierre Sparr NV Brut RÃ©serve Sparkling (CrÃ©mant d'Alsace) | 6 |
| Mumm Napa NV Brut Prestige Sparkling (Napa Valley) | 6 |
| Breathless NV Brut Sparkling (North Coast) | 5 |
| Chandon NV RosÃ© Sparkling (California) | 5 |
| J. Dumangin Fils NV Le RosÃ© Premier Cru Brut (Champagne) | 5 |
| Freixenet NV Cordon Negro Extra Dry Sparkling (Cava) | 5 |
| G. H. Mumm NV Cordon Rouge Brut (Champagne) | 5 |
| Henri Abele NV Brut (Champagne) | 5 |
| Freixenet NV Cordon Negro Brut Sparkling (Cava) | 5 |
| Jean Laurent NV Blanc de Noirs Brut Pinot Noir (Champagne) | 5 |
| Heidsieck & Co Monopole NV Blue Top Brut (Champagne) | 5 |
| Korbel NV Blanc de Noirs Sparkling (California) | 5 |
| Mailly Grand Cru NV DÃ©lice Demi-Sec (Champagne) | 5 |
| Mailly Grand Cru NV Blanc de Noirs Brut Pinot Noir (Champagne) | 5 |
| Pommery NV Brut Royal (Champagne) | 5 |
| Mumm Napa NV CuvÃ©e M Sparkling (Napa Valley) | 5 |
| Roederer Estate NV Brut RosÃ© Sparkling (Anderson Valley) | 5 |
| Segura Viudas NV Reserva Heredad Sparkling (Cava) | 5 |
| ThiÃ©not NV Brut (Champagne) | 5 |
| ThiÃ©not NV Brut RosÃ© (Champagne) | 5 |
| Besserat de Bellefon NV CuvÃ©e des Moines Brut RosÃ© (Champagne) | 4 |
| A.R. LenobleÂ  NV Intense Brut (Champagne) | 4 |
| Billecart-Salmon NV Brut RosÃ© (Champagne) | 4 |
| Bodegas Dios Baco S.L. NV Amontillado Sherry (Jerez) | 4 |
| Charles Heidsieck NV Brut RÃ©serve (Champagne) | 4 |
| Cave de RibeauvillÃ© NV Giersberger Brut Sparkling (CrÃ©mant d'Alsace) | 4 |
| Chandon NV Blanc de Noirs Sparkling (California) | 4 |
| Chanoine NV Tsarine CuvÃ©e Premium Brut (Champagne) | 4 |
| Canard-DuchÃªne NV CuvÃ©e LÃ©onie Brut (Champagne) | 4 |
| Canard-DuchÃªne NV Authentic Brut (Champagne) | 4 |
| Chandon NV Brut Classic Sparkling (California) | 4 |
| ChÃ¢teau Lestage Simon 2012 Haut-MÃ©doc | 4 |
| Delamotte NV Brut (Champagne) | 4 |
| Domaine Vacheron 2015 Sancerre | 4 |
| Drappier NV Carte d'Or Brut (Champagne) | 4 |
| Domaine Charles Baur NV Brut Sparkling (CrÃ©mant d'Alsace) | 4 |
| G. H. Mumm NV Brut RosÃ© (Champagne) | 4 |
| Ferrari NV RosÃ© Sparkling (Trento) | 4 |
| Jean-Baptiste Adam NV Brut RosÃ© Sparkling (CrÃ©mant d'Alsace) | 4 |
| Freixenet NV Carta Nevada Brut Sparkling (Cava) | 4 |
| Korbel NV Extra Dry Sparkling (California) | 4 |
| Maria Casanovas NV GlaÃ§ Brut Nature Sparkling (Cava) | 4 |
| Korbel NV Sweet CuvÃ©e Sparkling (California) | 4 |
| Masottina NV Brut (Conegliano Valdobbiadene Prosecco Superiore) | 4 |
| Lanson NV Black Label Brut (Champagne) | 4 |
| Korbel NV Brut RosÃ© Sparkling (California) | 4 |
| Louis Roederer NV Brut Premier (Champagne) | 4 |
| Joseph Perrier NV CuvÃ©e Royale Brut (Champagne) | 4 |
| Le Manzane NV Extra Dry (Conegliano Valdobbiadene Prosecco Superiore) | 4 |
| Nicolas Feuillatte NV Brut (Champagne) | 4 |
| Mumm Napa NV Brut RosÃ© Sparkling (Napa Valley) | 4 |
| MoÃ«t & Chandon NV ImpÃ©rial Brut (Champagne) | 4 |
| Pierre Gimonnet et Fils NV Cuis Premier Cru Blanc de Blancs Brut Chardonnay (Champagne) | 4 |
| Perrier JouÃ«t NV Grand Brut (Champagne) | 4 |
| Nicolas Feuillatte NV Brut RosÃ© (Champagne) | 4 |
| Ruinart NV Blanc de Blancs Brut Chardonnay (Champagne) | 4 |
| Taittinger NV Nocturne Sec (Champagne) | 4 |
| Sommariva NV Brut (Conegliano Valdobbiadene Prosecco Superiore) | 4 |
| Segura Viudas NV Brut Reserva Sparkling (Cava) | 4 |
| Segura Viudas NV Aria Estate Brut Sparkling (Cava) | 4 |
| Vignerons des Pierres DorÃ©es 2015 Salamandre d'Or (Coteaux Bourguignons) | 4 |
| Vollereaux NV Brut RÃ©serve (Champagne) | 4 |
| Terre di San Venanzio Fortunato NV Extra Dry (Conegliano Valdobbiadene Prosecco Superiore) | 4 |
| Vollereaux NV Blanc de Blancs Brut Chardonnay (Champagne) | 4 |
| Villa Sandi NV Extra Dry (Conegliano Valdobbiadene Prosecco Superiore) | 4 |

### Varietal

Entre los valores de varietales, se encuentran reunidos los tipos más conocidos:

|  |  |
| --- | --- |
| **Points** | **Percentage of Total variety** |
| Pinot Noir | 10.21% |
| Chardonnay | 9.04% |
| Cabernet Sauvignon | 7.29% |
| Red Blend | 6.88% |
| Bordeaux-style Red Blend | 5.32% |
| Riesling | 3.99% |
| Sauvignon Blanc | 3.82% |
| Syrah | 3.19% |
| RosÃ© | 2.74% |
| Merlot | 2.39% |
| Nebbiolo | 2.16% |
| Zinfandel | 2.09% |
| Sangiovese | 2.08% |
| Malbec | 2.04% |
| Portuguese Red | 1.90% |
| White Blend | 1.82% |
| Sparkling Blend | 1.66% |
| Tempranillo | 1.39% |
| RhÃ´ne-style Red Blend | 1.13% |
| Pinot Gris | 1.12% |
| Champagne Blend | 1.07% |
| Cabernet Franc | 1.04% |
| GrÃ¼ner Veltliner | 1.03% |
| Portuguese White | 0.89% |
| Bordeaux-style White Blend | 0.82% |
| Pinot Grigio | 0.81% |
| Gamay | 0.79% |
| GewÃ¼rztraminer | 0.78% |
| Viognier | 0.77% |
| Shiraz | 0.64% |
| Petite Sirah | 0.59% |
| Sangiovese Grosso | 0.58% |
| Barbera | 0.55% |
| Glera | 0.55% |
| Port | 0.51% |
| Grenache | 0.50% |
| Corvina, Rondinella, Molinara | 0.48% |
| Chenin Blanc | 0.45% |
| Tempranillo Blend | 0.45% |
| CarmenÃ¨re | 0.44% |
| AlbariÃ±o | 0.37% |
| Pinot Blanc | 0.34% |
| RhÃ´ne-style White Blend | 0.33% |
| Nero d'Avola | 0.28% |
| Aglianico | 0.28% |
| Moscato | 0.28% |
| Garnacha | 0.25% |
| Sauvignon | 0.24% |
| Verdejo | 0.23% |

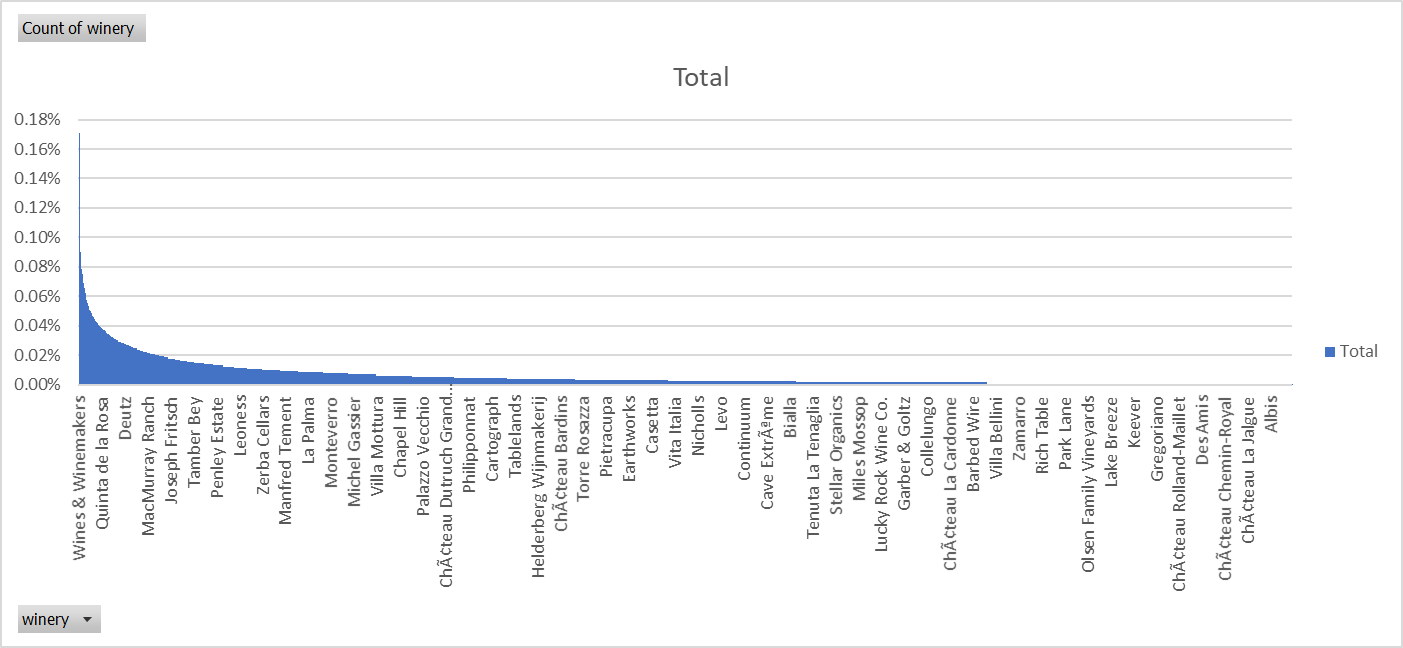
Estos valores si pueden afectar el objetivo final, ya que se debe tener muy en cuenta la relación que hay en una región con el tipo de varietal. Es también muy importante ya que una decisión de inversión puede ser sobre un varietal en una región específica.

### Casa/Vinoteca

Es importante encontrar que los valores de las vinotecas en todo el dataset tienen una repetición muy baja, para así tener en cuenta el tipo de uva en la región y no la casa matriz que los elabora y sus procesos, que pueden afectar al final la calidad. También es interesante analizar como una vinoteca puede elaborar mejores vinos por sus procesos.

Primeros 50 valores de conteo de vinotecas

|  |  |
| --- | --- |
| **Points** | **Count of winery** |
| Wines & Winemakers | 0.17% |
| Testarossa | 0.17% |
| DFJ Vinhos | 0.17% |
| Williams Selyem | 0.16% |
| Louis Latour | 0.15% |
| Georges Duboeuf | 0.15% |
| Chateau Ste. Michelle | 0.15% |
| Concha y Toro | 0.13% |
| Columbia Crest | 0.12% |
| Kendall-Jackson | 0.10% |
| Siduri | 0.10% |
| Gary Farrell | 0.10% |
| Lynmar | 0.09% |
| Montes | 0.09% |
| Albert Bichot | 0.09% |
| Trapiche | 0.09% |
| Undurraga | 0.09% |
| Jean-Luc and Paul Aegerter | 0.09% |
| Casa Santos Lima | 0.09% |
| Santa Ema | 0.09% |
| Robert Mondavi | 0.09% |
| Chanson PÃ¨re et Fils | 0.08% |
| Fess Parker | 0.08% |
| V. Sattui | 0.08% |
| Iron Horse | 0.08% |
| Foxen | 0.08% |
| Louis Jadot | 0.08% |
| Maryhill | 0.08% |
| Feudi di San Gregorio | 0.08% |
| Chehalem | 0.08% |
| Kunde | 0.08% |
| Domaine Zind-Humbrecht | 0.08% |
| BrÃ¼ndlmayer | 0.08% |
| Terre Rouge | 0.08% |
| Dutton-Goldfield | 0.08% |
| D'Arenberg | 0.08% |
| Errazuriz | 0.08% |
| Seven Hills | 0.08% |
| Martin Ray | 0.08% |
| J. Portugal Ramos | 0.08% |
| Henri de Villamont | 0.08% |
| Kenwood | 0.08% |
| Cameron Hughes | 0.08% |
| Rodney Strong | 0.07% |
| Dr. Loosen | 0.07% |
| Georges Vigouroux | 0.07% |
| Naggiar | 0.07% |
| Olivier Leflaive | 0.07% |
| Fenestra | 0.07% |



### Correlaciones

### Inferencias