

Calidad del vino



Grupo número 12

Integrantes

Juan Simón Abraham

Abril Iglesias

Tomás Lucio Ilari

19 de noviembre de 2024

| | |
|---|-----------|
| Introducción | 3 |
| Hipótesis del Estudio | 3 |
| Análisis | 4 |
| Descripción de las variables | 4 |
| Distribución de las Variables: | 5 |
| Análisis estadístico univariado | 5 |
| Análisis estadístico bivariado | 6 |
| Análisis estadístico Multivariado | 8 |
| Reducción de dimensionalidad | 12 |
| Regresión Logística Múltiple | 14 |
| Desarrollo de Hipótesis | 16 |
| Los vinos de uvas “Viogner” son más ácidos que los vinos de uvas “Merlot”. | 16 |
| Los vinos de uvas “Viogner” son más dulces en comparación a los vinos de uvas “Merlot”. | 17 |
| A mayor calidad los vinos tienen mayor cantidad de alcohol | 18 |
| La densidad del vino es menor a mayor calidad | 18 |
| Conclusiones | 20 |
| Bibliografía | 21 |

Introducción

Este informe se centra en el análisis del conjunto de datos de las pruebas obtenidas sobre vinos en la bodega La Esperanza, que proporciona variables detalladas sobre las características químicas y la calidad de vinos tintos y blancos, más específicamente sobre dos uvas: Merlot (vino tinto) y Viogner (vino blanco). El objetivo de este estudio es identificar los factores clave que influyen en la calidad del vino y encontrar qué factores distinguen a ambos tipos de vino.

El dataset incluye variables como la acidez fija, el pH, el contenido de alcohol, azúcar residual, entre otras, y una evaluación de calidad que va de 3 a 8. Comprender cómo estas variables afectan la calidad del vino puede tener importantes implicaciones económicas y de producción. Además distinguir las características de cada uva podría ayudar a optimizar el proceso de vinificación ó mejorar el maridaje.

Para llevar a cabo este análisis, se utilizarán diversas técnicas estadísticas y de análisis de datos, incluyendo análisis univariado como medidas estadística, bivariado analizando la correlación entre distintas variables y multivariado haciendo reducción de dimensionalidad. Además se incluyeron gráficos para representar mejor los datos y se usar test no paramétricos para la demostración de las hipótesis.

Hipótesis del Estudio

En el presente informe, se evalúan las siguientes hipótesis:

- Los vinos de uvas “Viogner” son más ácidos que los vinos de uvas “Merlot”.
- Los vinos de uvas “Viogner” son más dulces en comparación a los vinos de uvas “Merlot”.
- A mayor calidad un vino tiene mayor cantidad de alcohol.
- La densidad del vino es menor a mayor calidad.

Hipótesis Adicionales:

- El alcohol del vino es menor a mayor densidad (concentración de sólidos disueltos).
- La densidad es menor a menor azúcar residual.
- Los vinos con uva Viognier tienen más alcohol que los que se realizan con la uva Merlot.

Análisis

Descripción de las variables

1. type: tipo de uva con la que se elabora el vino.
2. fixed acidity: cantidad de ácidos no volátiles presentes en el vino, medida en gramos por litro.
3. volatile acidity: cantidad de ácidos volátiles presentes en el vino, medida en gramos por litro.
4. citric acid: contenido de ácido cítrico en el vino, medido en gramos por litro.
5. residual sugar: cantidad de azúcar que queda en el vino después de la fermentación, medida en gramos por litro.
6. chlorides: concentración de cloruros (sales) en el vino, medida en gramos por litro.
7. free sulfur dioxide: cantidad de dióxido de azufre que no está ligado químicamente en el vino, medida en gramos por litro.
8. total sulfur dioxide: suma del dióxido de azufre libre y el combinado en el vino, medida en gramos por litro.
9. density: medida de la masa por unidad de volumen del vino, utilizada para estimar la concentración de sólidos disueltos, medida en gramos por litro.
10. pH: medida de la acidez o alcalinidad del vino.
11. sulphates: concentración de sales de sulfato en el vino, medida en gramos por litro.
12. alcohol: contenido alcohólico del vino, medido en porcentaje de volumen (% vol).
13. quality: puntuación del vino, con una escala que va de 0 a 10.

A continuación, exploramos las variables según su tipo, y analizaremos cómo se comportan por separado.

Variables Cuantitativas Continuas: Fixed Acidity, Volatile Acidity, Citric Acid, Residual Sugar, Chlorides, Free Sulfur Dioxide, Total Sulfur Dioxide, Density, pH, Sulphates, Alcohol.

Estas variables cuantitativas continuas representan mediciones específicas que pueden tomar cualquier valor dentro de un rango, permitiendo un análisis detallado de las características del vino.

Variable Cuantitativa Discreta: Quality.

La calidad se registra como una variable cuantitativa discreta, ya que toma valores enteros que representan diferentes niveles de calidad del vino dentro del rango.

La única variable cualitativa en el conjunto de datos es la variable **Type**, que clasifica las uvas en dos tipos, Viognier o Merlot. Esta es una variable nominal ya que categoriza las uvas en grupos que no tienen un orden específico.

Distribución de las Variables:

Cabe destacar que ninguna de estas variables sigue una distribución normal. Esto implica que en el análisis de datos, es importante considerar técnicas estadísticas que no asuman normalidad para obtener resultados más precisos.

Análisis estadístico univariado

El conjunto de datos de trabajo está dividido principalmente en dos, los vinos provenientes de uvas Merlot y los provenientes de uvas Viogner, y por lo que se puede ver en Gráfico 1 la división de los datos es equitativa.

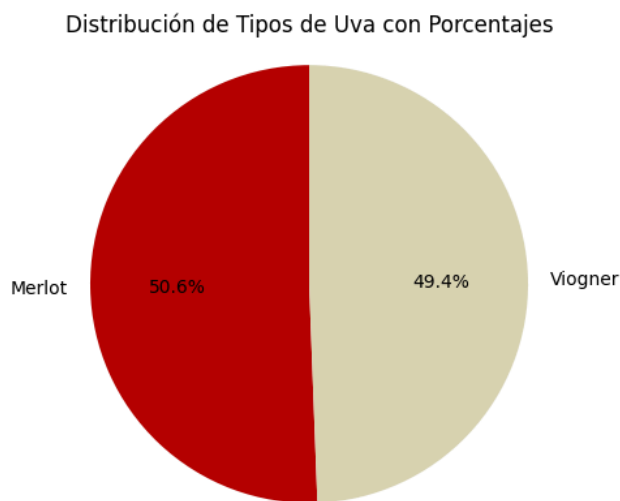


Gráfico 1: Frecuencia variable "type".

Sin embargo si examinamos las muestras de mayor calidad y las comparamos con las de menor calidad, encontramos que la distribución de tipo de uvas cambia y deja marcada una preferencia clara de los consumidores, ver Gráfico 2.

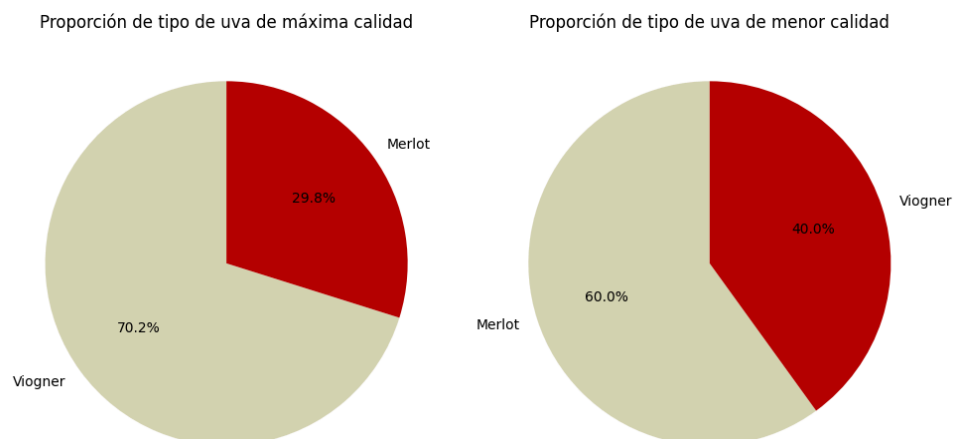


Gráfico 2: Frecuencia variable "type" en vinos de máxima y mínima calidad.

Aunque esto parece indicar un favoritismo claro, hay que tener en cuenta que la cantidad de muestra no es la mejor, ya que, el conjunto de datos no se distribuye igual para todos los tipos de calidades, es decir, hay muchas más muestras en las calidades intermedias 5 ó 6 y muchas menos en los extremos, esto queda bien reflejado en el Gráfico 3.

Esta distribución puede explicarse por la tendencia a elegir valores centrales, ya que, al realizar clasificaciones, suele observarse que la moda tiende a concentrarse en el valor más "neutro" de la escala. Esto ocurre porque, a veces, es difícil determinar si un valor pertenece a las categorías superiores o inferiores, por lo que se opta por colocarlo en el centro.

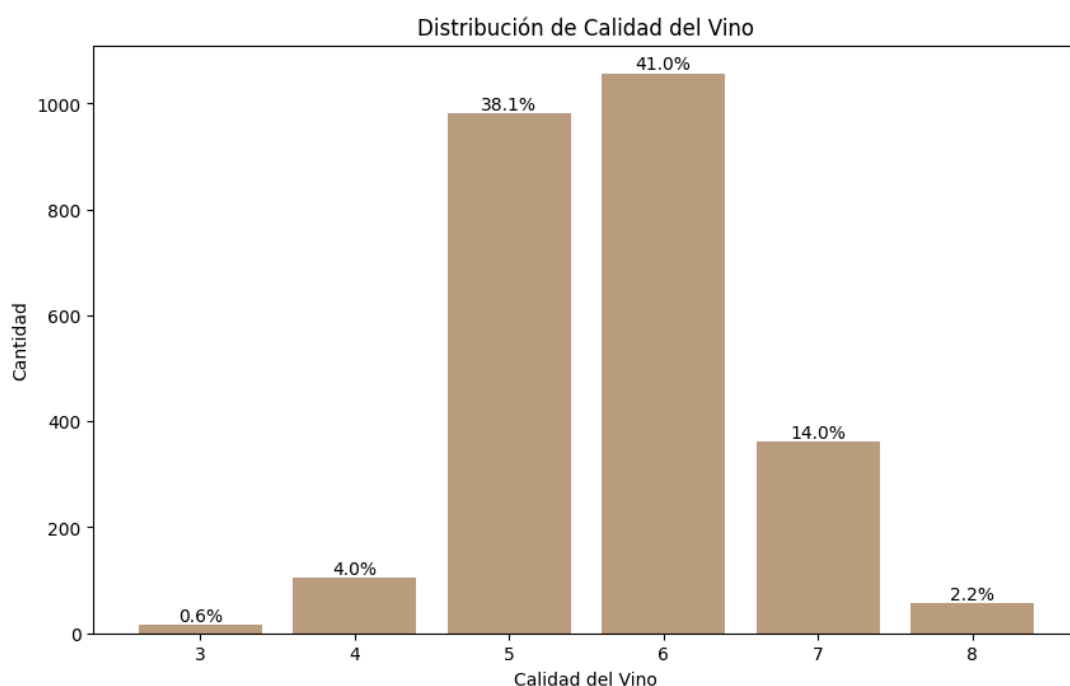


Gráfico 3: Frecuencia variable "quality".

Análisis estadístico bivariado

En el análisis bivariado, buscamos entender las relaciones entre dos variables y cómo estas interactúan para influir en la calidad del vino. Esto mediante el cálculo de la correlación entre las variables y usando gráficos de dispersión para encontrar relaciones lineales. Esta sección explora correlaciones que pueden revelar factores clave que afectan las propiedades del vino, como el contenido de alcohol, la densidad y otras características.

Por ejemplo, al analizar la relación entre el nivel de alcohol y la densidad del vino, observamos que estas dos variables suelen presentar una relación inversa (ver Gráfico 4): a medida que el contenido de alcohol aumenta, la densidad tiende a disminuir. Este comportamiento puede estar relacionado con el efecto del proceso de fermentación, donde a mayor producción de alcohol, disminuye la densidad del vino.

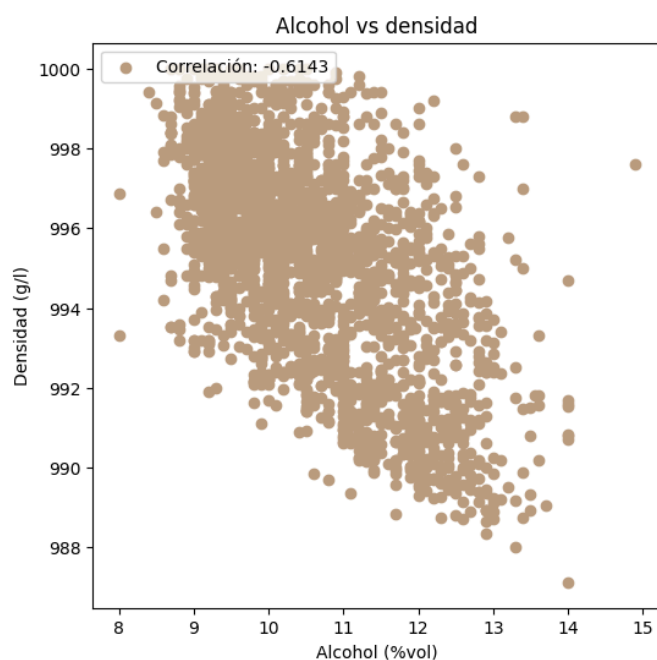


Gráfico 4: Diagrama de dispersión de la variable “Alcohol” y “Density”.

A partir del estudio las variables revelaron que hay una importante relación entre el dióxido de azufre libre y total, esto es lógico ya que ambas representan cantidades de una misma sustancia química. Esto podría ser relevante a la hora de estudiar si el dióxido de azufre está ligado a la calidad del vino o si es definitorio en la calificación de la misma ya que el mismo es un componente agregado para prevenir el olor avinagrado producido por la acidez volátil. Si bien no se ve una clara correlación entre la acidez volátil y el dióxido de azufre, estos podrían estar relacionados a partir de otras variables o bien a través de variables no estudiadas, como el tiempo de conservación, el tiempo de estacionamiento, entre otras.

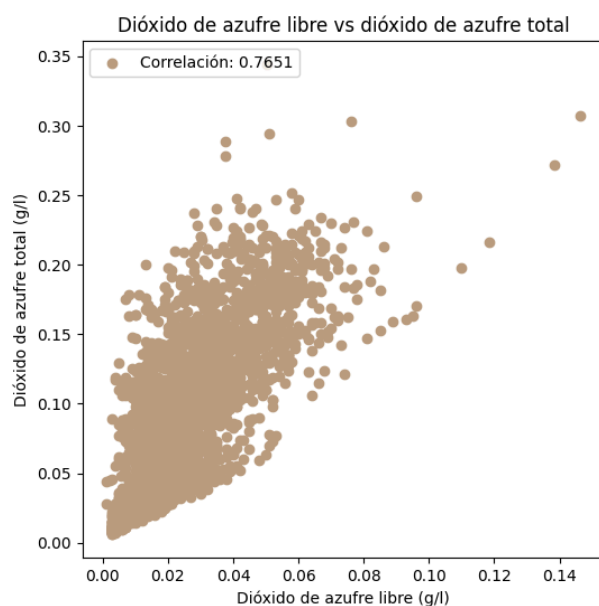


Gráfico 3: Diagrama de dispersión de la variable “total sulfur dioxide” y “free sulfur dioxide”.

Análisis estadístico Multivariado

En esta sección, exploramos cómo múltiples variables interactúan simultáneamente para proporcionar una visión más completa de las características del vino y su calidad. El análisis multivariado permite observar cómo ciertas propiedades varían entre distintos tipos de vinos y si estas diferencias se asocian con factores que afectan la percepción y clasificación del vino. Para ellos usaremos: distintos gráficos que resuman los datos obtenidos, técnicas de disminución de dimensionalidad para poder ver las muestras en el plano, clustering para poder realizar agrupaciones y definir qué características se diferencian entre los grupos y por último usaremos regresión logística para poder cuantificar cuánto afectan las variables dependientes a la variable objetivo.

Para seguir explorando la relación de la densidad con el vino examinamos las diferencias entre densidades de los distintos tipos de uvas del conjunto de datos.

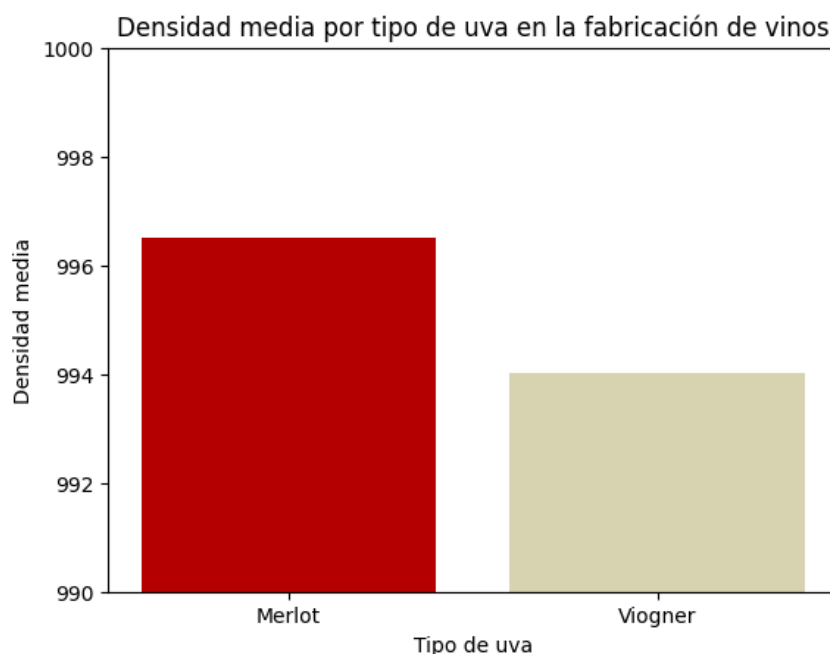


Gráfico 4: Gráfico de barras de la variable “Density” respecto al tipo de uva.

Para empezar, presentamos un gráfico de barras que ilustra la diferencia en la densidad media entre vinos blancos y vinos tintos. Los vinos blancos tienen una densidad media menor que los vinos tintos.

Esta comparación es relevante porque la densidad puede influir en la textura y el cuerpo del vino, atributos que suelen diferir notablemente entre ambos tipos y que impactan en la percepción general de su calidad. Esto puede deberse a el pH, la cantidad de azúcar residual, el porcentaje de alcohol, la madurez de la uva entre muchos otros factores.

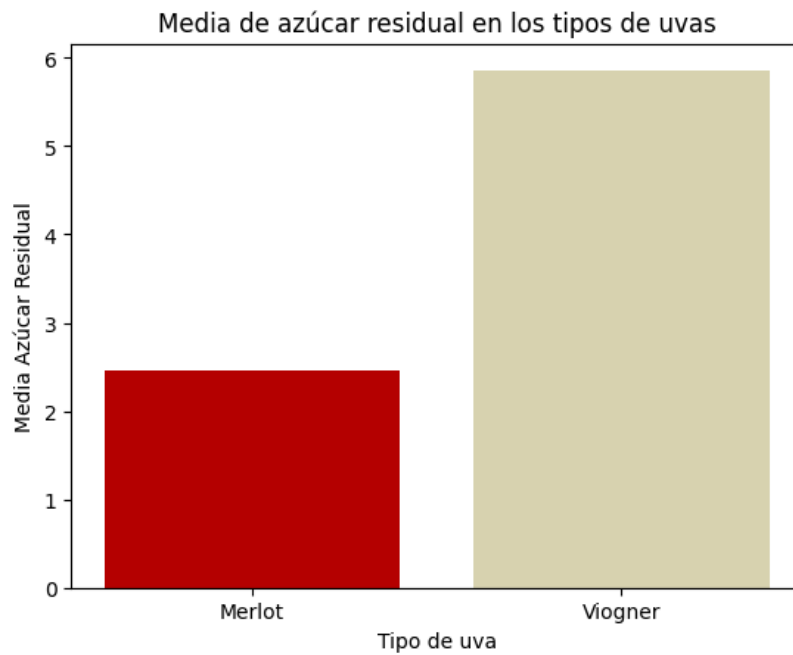


Gráfico 5: Gráfico de barras de la variable "Residual sugar" respecto al tipo de uva.

Como se puede ver, los vinos blancos tienen mayor azúcar residual en promedio, es decir, son más dulces que los vinos tintos y aún así son menos densos, por lo que hay que analizar el pH entre ellos.

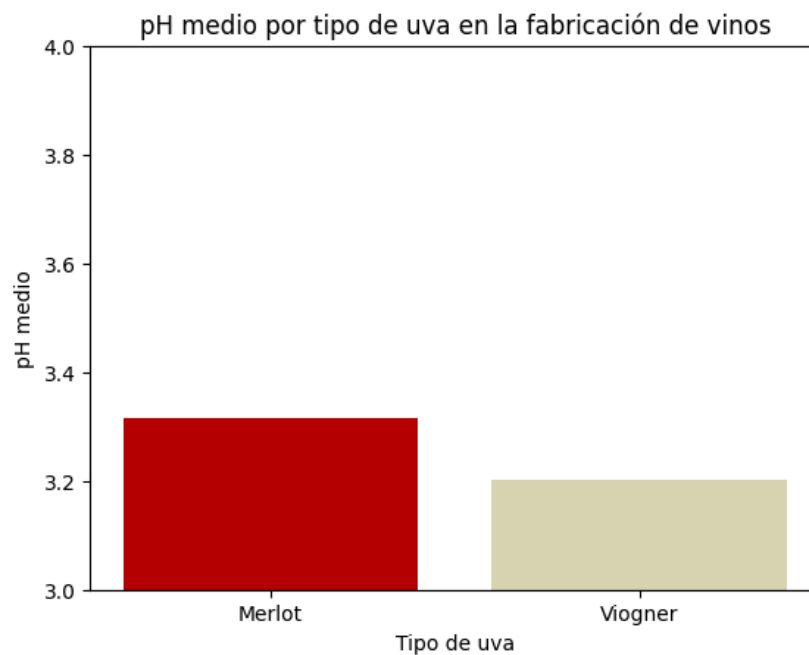


Gráfico 6: Gráfico de barras de la variable "pH" respecto al tipo de uva.

El pH de los vinos "Merlot" es un poco mayor por lo que ese puede ser el motivo de la diferencia entre densidades, además de la madurez de la uva.

Otra relación interesante para analizar es la diferenciación en las características de los vinos según su calidad. En particular, observamos que a medida que aumenta la calidad del vino, también incrementa su porcentaje de alcohol. Esto sugiere una posible asociación entre el contenido alcohólico y la clasificación de calidad, lo cual se detalla en el Gráfico 7.

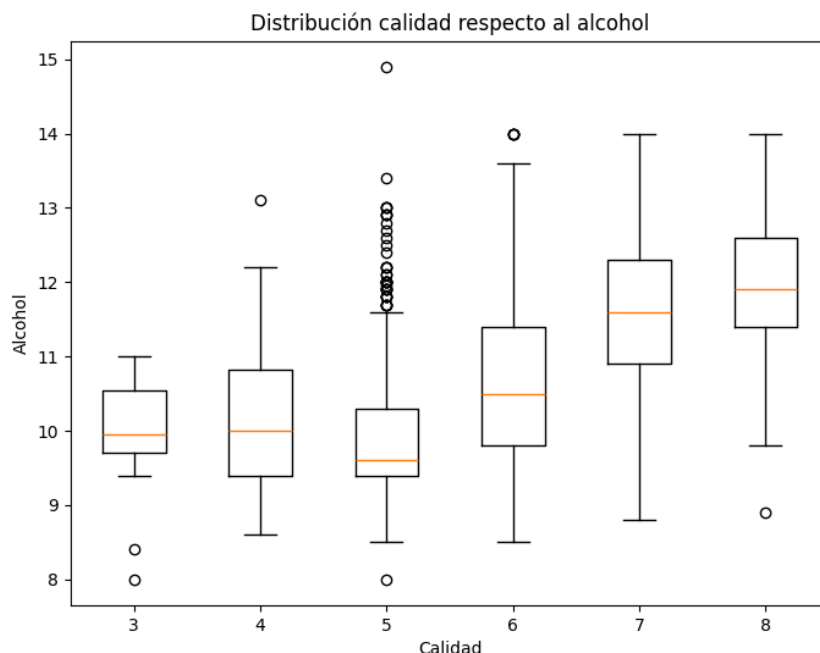


Gráfico 7: Box-Plot de la variable “Alcohol” respecto a las calidades de vinos.

Este gráfico permite observar una tendencia ascendente en el contenido de alcohol a medida que aumenta la calidad del vino, lo cual sugiere una asociación positiva entre estas dos variables. A medida que la calidad del vino aumenta, el contenido de alcohol también tiende a incrementarse. Los vinos de mayor calidad (calidades 6, 7 y 8) presentan niveles de alcohol más altos en comparación con aquellos de calidad inferior (calidades 3, 4 y 5). Además las medianas del contenido de alcohol muestran un incremento progresivo con la calidad del vino, especialmente notorio entre las calidades 5 y 7, lo cual sugiere que los vinos de calidad superior suelen tener una mayor concentración de alcohol en promedio.

En las calidades intermedias, particularmente en calidad 5, se observa la presencia de varios valores atípicos hacia niveles de alcohol más altos. Esto indica que algunos vinos de calidad media pueden alcanzar contenidos de alcohol similares a los de calidad superior, aunque posiblemente carezcan de otras características necesarias para clasificar en categorías de mayor calidad.

Otro factor interesante a observar en relación con la calidad del vino es la densidad. Al analizar cómo varía esta característica, encontramos que, a medida que aumenta la calidad del vino, la densidad tiende a disminuir. Este comportamiento puede estar relacionado con la composición química y los procesos de fermentación, donde un menor contenido de sólidos y una mayor concentración de alcohol

suelen resultar en una densidad más baja. A continuación, se presenta un gráfico que ilustra esta relación (Gráfico 8).

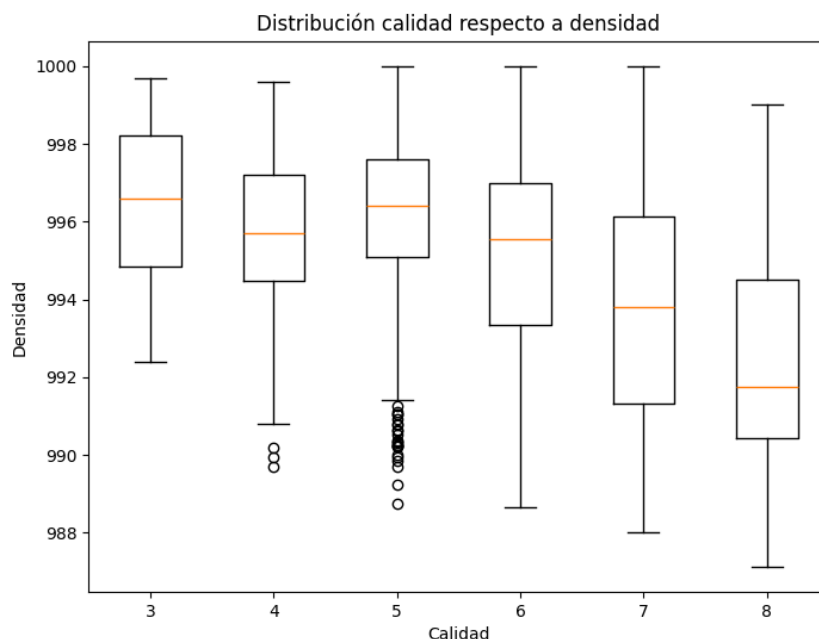


Gráfico 8: Box-Plot de la variable "Density" respecto a las calidades de vinos.

Se observa que, a medida que aumenta la calidad del vino, la densidad tiende a disminuir. Este patrón sugiere que los vinos de mayor calidad presentan densidades más bajas en promedio, posiblemente debido a un mayor contenido de alcohol y menor cantidad de azúcares residuales. Además, se observa una considerable variabilidad en la densidad dentro de cada categoría de calidad, especialmente en los niveles intermedios, lo cual indica heterogeneidad en esta característica. Los valores atípicos presentes en calidades medias reflejan que algunos vinos de calidad intermedia tienen densidades similares a los de alta calidad, lo que podría estar relacionado con diferencias en el proceso de producción.

Esta observación es coherente con el análisis previo, en el cual se identificó una relación inversa entre la densidad y el contenido de alcohol en el vino. Dado que la densidad tiende a disminuir a medida que aumenta la calidad, y que los vinos de mayor calidad también presentan niveles más altos de alcohol, es razonable suponer que estas dos características están inversamente relacionadas: un incremento en el contenido de alcohol se asocia con una disminución en la densidad del vino, y viceversa.

Reducción de dimensionalidad

A continuación, se procederá a realizar reducción de dimensionalidad sobre las variables del dataset. Este enfoque permitirá visualizar en el plano cómo se agrupan las distintas muestras de vino en función de sus características, facilitando la identificación de patrones y posibles agrupaciones según la calidad u otras propiedades relevantes. Esta técnica ayudará a simplificar la complejidad del conjunto de datos, preservando la mayor cantidad posible de información significativa en dos dimensiones.

Para añadir peso a la hipótesis de que los vinos “Viogner” tienen menor pH que los vinos “Merlot” podemos ver el gráfico 9 el cual muestra como las agrupaciones de mayor pH se encuentran dentro del conjunto de las muestras pertenecientes los vinos “Merlot”.

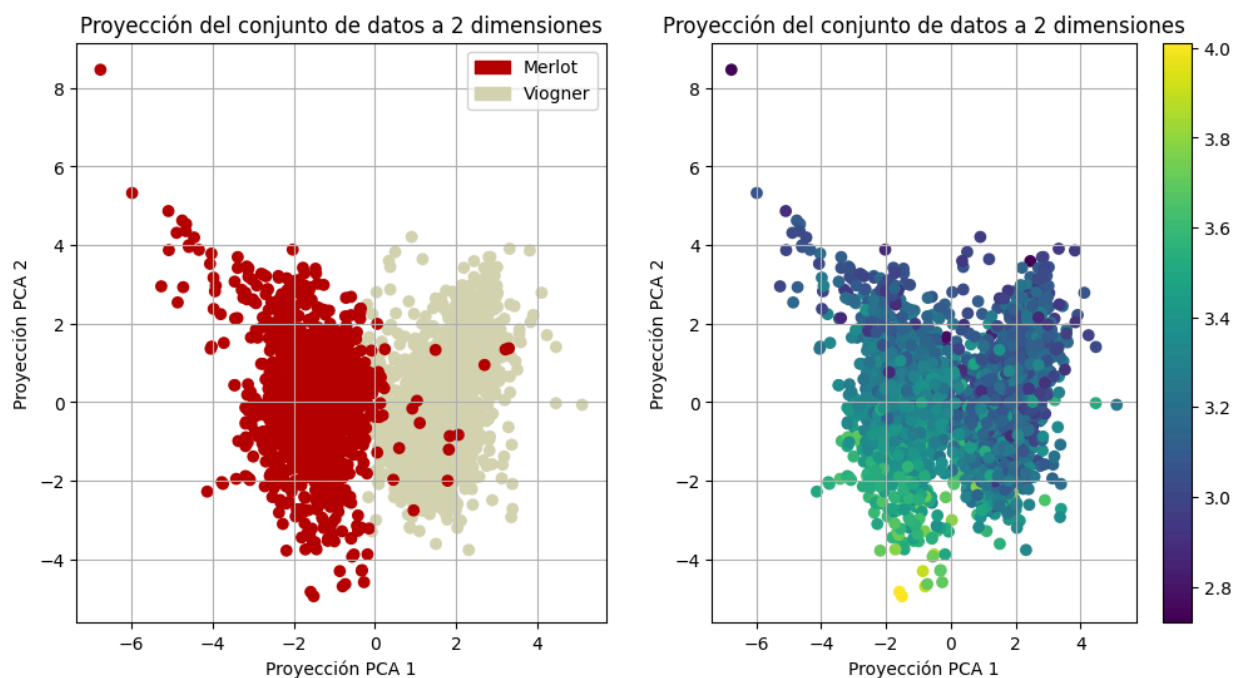


Gráfico 9: Gráfico de dispersión del pH y de las muestras luego de reducción por PCA.

Cabe destacar que la varianza explicada por ambas componentes en total es de un 50% lo cual es bastante moderado, si bien no podemos dar mucha relevancia a esta forma de visualización nos da una idea de cómo se ubican los datos en el plano.

Para continuar utilizamos un modelo de reducción no lineal para comprobar si hay diferencia entre las visualizaciones de las muestras. En este caso comparamos el azúcar residual que hay entre los datos y como vemos en el gráfico 10, queda muy bien representado como las muestras de mayor azúcar residual son los vinos “Viogner”, esto apoya la hipótesis de que los vinos blancos son más dulces.

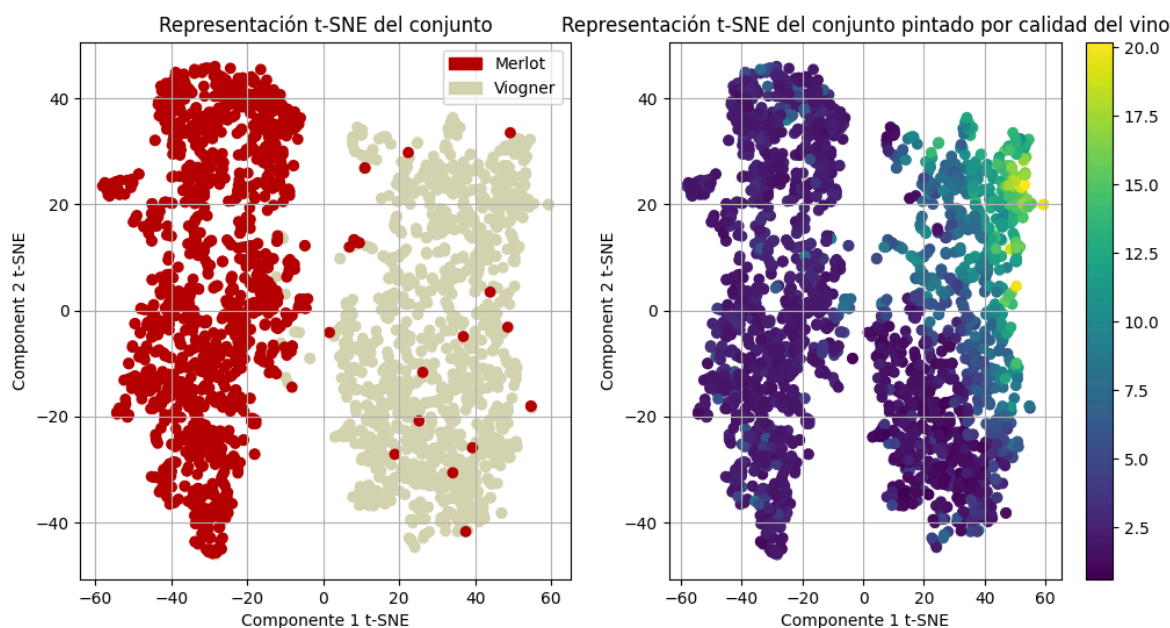


Gráfico 10: Gráfico de dispersión de las muestras luego de reducción por t-SNE.

Analizando ambos gráficos de reducción podemos encontrar que las distribuciones se ven bastante similares, si con PCA los grupos de tipos de uvas se encontraban más conjuntos y entrelazados, en ambos se pueden encontrar los mismos grupos de forma alargada.

Para finalizar utilizaremos este método mediante t-SNE para argumentar cómo afecta el alcohol a la calidad, ya que como se ve en el gráfico 11, es claro como las zonas de mayor calidad en su mayoría también son las de mayor concentración alcohólica.

Representación t-SNE del conjunto pintado por calidad del vino

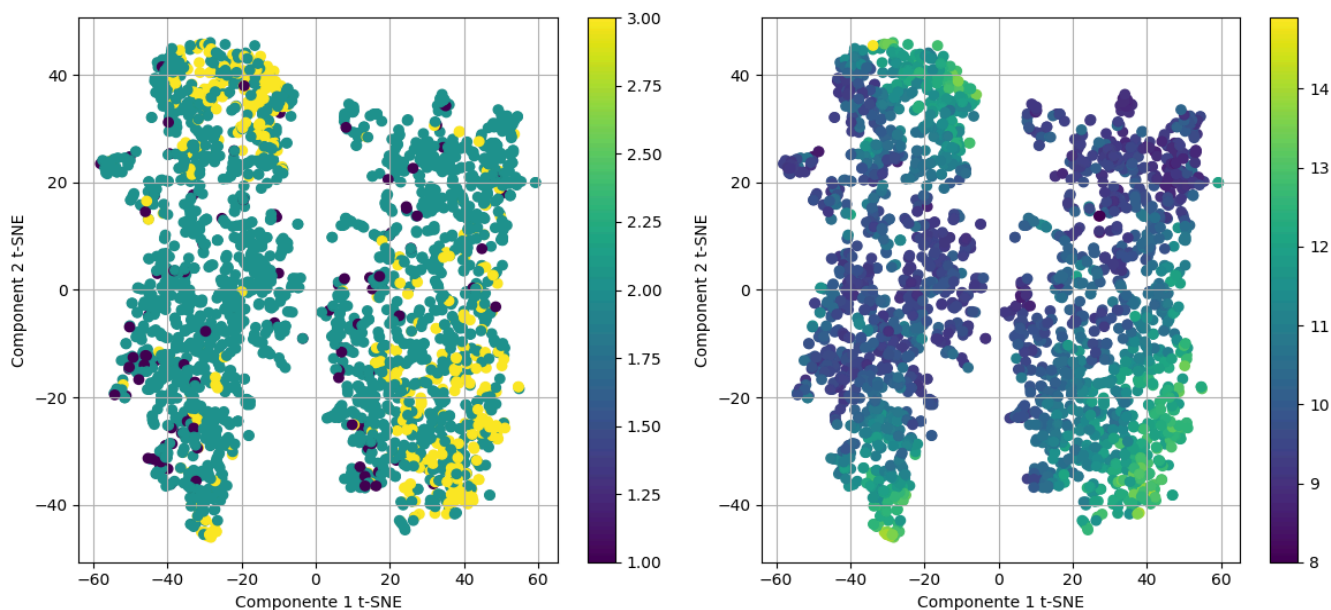


Gráfico 11: Gráfico de dispersión de las muestras luego de reducción por t-SNE.

Para este análisis dividimos las clasificaciones de vino en 3 grupos, la gama baja, media y alta y le asignamos números del 1 al 3 respectivamente. Observando el gráfico es evidente que los vinos de más calidad son los de mayor volumen alcohólico, exceptuando la parte media de la distribución derecha.

Además, si solo vemos la representación de la calidad de las muestras en el plano, se ve claramente como la mayoría de las muestras de calidad alta se encuentran en el grupo de los vinos “Viogner” lo cual concuerda con el análisis de distribución de tipo de uva hecho anteriormente.

Regresión Logística Múltiple

Para nuestro conjunto de datos usamos un modelo de regresión logística múltiple, porque necesitamos analizar la variable calidad como dependiente y esta es cualitativa, asimismo aplicamos todas las demás features como variables dependientes, debido a esto es que utilizamos un modelo múltiple. El objetivo de aplicar este modelo es adquirir información sobre cuáles son las características que más influyen en la calidad de los vinos, para esto observamos los coeficientes asociados a cada feature, de manera que, los más grandes (en su módulo) son los que más afectan a la variable objetivo.

Como resultado de este experimento obtuvimos que una de las variables que más afecta es la acidez volátil y lo hace de forma negativa (ver Gráfico 12). Esto tiene sentido ya que como mencionamos previamente este químico produce un olor avinagrado en los vinos, por lo tanto mientras más concentración haya menos placentero será la fase olfativa de la degustación.

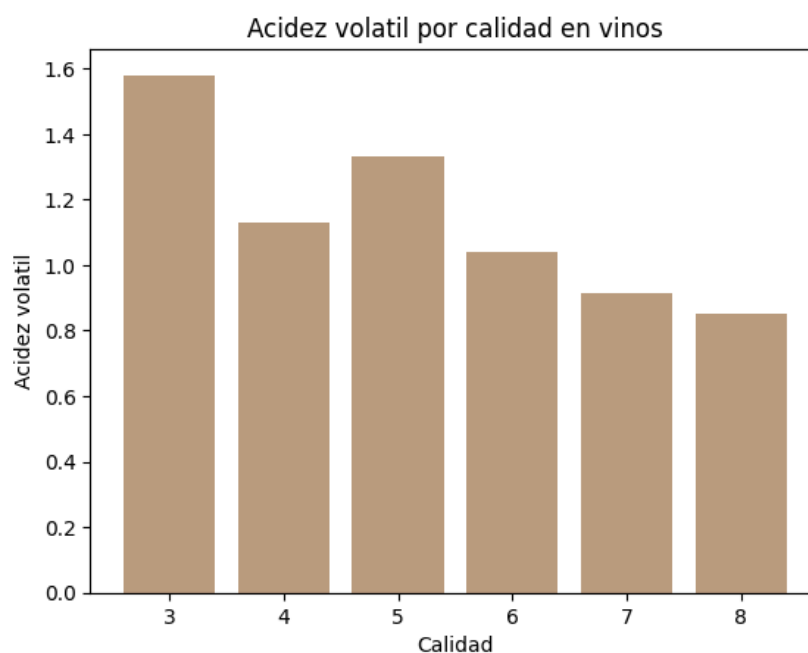


Gráfico 12: Gráfico de barras de la variable “Volatile acidity”.

Como se puede observar claramente hay una tendencia decreciente. Aunque en la calidad 5, es decir, los vinos gama “media”, se ve una diferenciación con respecto a las demás calidades, esto puede deberse a que es una de las calidades con mayor cantidad de muestras junto con la 6 como examinamos previamente. Además esta categoría alberga varios outliers que pueden afectar al análisis.

Conjuntamente, la característica que más ayuda a definir la calidad del vino según el regresor es el alcohol, lo cuál concuerda con todo el análisis hecho anteriormente.

Desarrollo de Hipótesis

Los vinos de uvas “Viogner” son más ácidos que los vinos de uvas “Merlot”.

Hipótesis: Los vinos blancos presentan un pH significativamente menor en comparación con los vinos tintos.

Para validar esta hipótesis, hemos llevado a cabo un análisis que incluye la división del conjunto de datos, la verificación de supuestos fundamentales y la selección del test de hipótesis adecuado.

División del Dataset Original: Inicialmente, se dividió el conjunto de datos en dos subconjuntos: uno que contiene el pH de los vinos Viognier (representando los vinos blancos) y otro con el pH de los vinos Merlot (representando los vinos tintos). Esta segmentación permite un análisis comparativo claro entre las dos categorías de vino.

Verificación de Supuestos: Para determinar el tipo de test de hipótesis más adecuado, es crucial comprobar si los datos cumplen con los supuestos de normalidad y homocedasticidad.

Normalidad: La normalidad de ambos conjuntos de datos se evaluó mediante el test de Shapiro-Wilk. El resultado fue un p-valor de 0.0 para ambos subconjuntos, lo que indica que los datos no siguen una distribución normal.

Homocedasticidad: Para evaluar si las varianzas de los datos eran homogéneas, se aplicó el test de Levene. El p-valor obtenido fue de 0.9, lo que sugiere que los datos son homocedásticos, es decir, las varianzas entre los grupos son aproximadamente iguales.

Selección del Test de Hipótesis: Dado que los datos no se distribuyen normalmente pero cumplen con el supuesto de homocedasticidad, se decidió utilizar el test de Mann-Whitney U para evaluar la hipótesis. Este test no paramétrico es adecuado para comparar medianas entre dos grupos independientes cuando no se puede asumir la normalidad de los datos.

Evaluación de la Hipótesis: El test de Mann-Whitney U nos da como resultado un p valor = 0.

Por ende se rechaza la hipótesis nula y podemos decir que existe una diferencia significativa en pH entre los vinos viognier y los merlot.

En la prueba de esta hipótesis mediante el test U, encontramos diferencias significativas entre las muestras, con un valor p muy cercano a 0, lo que indica una fuerte evidencia estadística para rechazar la hipótesis nula.

No obstante, aunque el test confirma la significancia de la diferencia, no proporciona información sobre la dirección de esta diferencia; es decir, no especifica si el pH de los vinos Viognier es efectivamente menor que el de los vinos Merlot, como se planteaba en nuestra hipótesis inicial, o si ocurre lo contrario.

Para apoyar la interpretación a favor de nuestra hipótesis, hemos recurrido a los gráficos analizados previamente, es decir, el gráfico 6 y el 9 que muestran cómo efectivamente los vinos “Merlot” son más básicos y los “Viogner” más ácidos. Es por esto que concluimos que la hipótesis es verdadera.

Los vinos de uvas “Viogner” son más dulces en comparación a los vinos de uvas “Merlot”.

Hipótesis: Los vinos blancos presentan una cantidad significativamente mayor de azúcar residual en comparación con los vinos “merlot”.

El azúcar residual en el vino son los azúcares no fermentados como la glucosa y la fructosa, es crucial para determinar el nivel de dulzor del vino. Generalmente, los vinos con más azúcar residual se perciben como más dulces. Por ello, centraremos nuestro análisis en esta variable para obtener resultados que validen nuestra hipótesis.

Inicialmente, se dividió el conjunto de datos en dos subconjuntos: uno que contiene el azúcar residual de los vinos Viognier y otro con el azúcar residual de los vinos Merlot.

Verificación de Supuestos: Para determinar el tipo de test de hipótesis más adecuado, es crucial comprobar si los datos cumplen con los supuestos de normalidad y homocedasticidad.

Selección del Test de Hipótesis: Dado que los resultados fueron que no se los datos no se distribuyen normalmente y que no se cumple el supuesto de homocedasticidad, se decidió utilizar el test de test Kruskal-Wallis para evaluar la hipótesis, Este es un test no paramétrico utilizado para comparar las medianas de más de dos grupos independientes.

Evaluación de la Hipótesis: El test Kruskal-Wallis nos da como resultado un p valor = 0.

Por lo que se rechaza la hipótesis nula y podemos decir que existe una diferencia significativa en el azúcar residual entre los vinos viognier y los merlot. Conjuntamente con el análisis previo, podemos afirmar que esta hipótesis es verdadera.

A mayor calidad los vinos tienen mayor cantidad de alcohol

Hipótesis: Los vinos presentan una cantidad significativamente mayor de alcohol a medida que su calidad aumenta.

Como se presentó anteriormente, se dividió el conjunto de datos en cinco subconjuntos, cada uno correspondiente a una calidad de vino. Esta elección se basó en el rango de calidades presentes en el conjunto de datos, que va del tres al ocho.

Al comprobar la normalidad mediante Shapiro-Wilk, el resultado fue un p-valor de 0,025 por lo que los datos no siguen una distribución normal. Para evaluar si las varianzas de los datos eran homogéneas, se aplicó el test de Levene. El p-valor obtenido fue de 0.0, lo que sugiere que los datos no son homocedásticos. Debido a esto se optó por utilizar el test de la mediana para evaluar la hipótesis. Este test no paramétrico es adecuado para comparar múltiples grupos independientes, como las distintas calidades de vinos en este estudio, y determina si las medianas de estos grupos difieren significativamente.

Evaluación de la Hipótesis: El test de la mediana nos dio como resultado un p valor = 0, lo que permite rechazar la hipótesis nula y concluir que existe una diferencia significativa de alcohol entre los distintos conjuntos. Por lo que consecuentemente y luego de lo planteado en los gráficos, podemos comprobar la veracidad de la hipótesis.

La densidad del vino es menor a mayor calidad

Hipótesis: Los vinos presentan una cantidad significativamente menor de densidad a medida que su calidad aumenta.

Al igual que en la hipótesis anterior, comenzamos dividiendo el conjunto de datos en cinco subconjuntos, cada uno correspondiente a una calidad de vino.

En esta hipótesis, analizamos cómo varía la densidad del vino en cada uno de los grupos de calidad. La densidad es un factor importante en la evaluación del vino, ya que está relacionada con la concentración de sólidos disueltos en el líquido. Un vino con una menor densidad puede indicar un contenido más equilibrado y refinado, atributos que suelen asociarse con una mayor calidad.

Debido a que los datos no cumplen con los supuestos de normalidad ni de homocedasticidad, se decidió utilizar el test de Kruskal-Wallis para evaluar la hipótesis. El mismo arrojó un p-valor de 0.0, lo que nos lleva a rechazar la hipótesis nula. Indicando así que hay una diferencia significativa en la densidad entre los distintos grupos de calidad de vino, variando significativamente según la calidad. Como previamente estudiamos las medianas, podemos concluir que es cierto que a mayor calidad la densidad es menor.

El alcohol del vino es menor a mayor densidad (concentración de sólidos disueltos).

Esta suposición surge del análisis hecho sobre la relación entre las variables “Alcohol” y “Density”. Como vimos anteriormente, el coeficiente de correlación calculado sobre estas características están linealmente relacionadas de forma inversa, lo cual es una buena base para la hipótesis pero se podría indagar más para demostrarlo.

La densidad es menor a menor azúcar residual.

Hipótesis que resulta luego de observar que el azúcar la relación entre el azúcar residual y la densidad en los vinos “Viogner” la cual resultó bastante alta, por lo que sería pertinente separar el conjunto de datos por tipo de uvas y analizar esta correlación de forma más específica.

Los vinos con uva Viognier tienen más alcohol que los que se realizan con la uva Merlot.

A partir del análisis de los datos con relación a cómo afectaba la proporción de alcohol con respecto a la calidad, y la cantidad de vinos que presentaba cada valor, hipotetizamos que si la cantidad de vinos de uva Viogner era mayor en vinos de más calidad, y la cantidad de alcohol presente en estos vinos también lo era, entonces sería probable que los vinos Viogner tuvieran mayor alcohol que los Merlot.

Conclusiones

Durante la investigación, se corroboraron varias hipótesis acerca de las propiedades de los vinos dependiendo de la uva empleada y la calidad del producto final. Se observó que los vinos elaborados con uvas Viognier tienen una mayor acidez y dulzura en comparación con los vinos elaborados con uvas Merlot.

Además, se determinó una relación significativa entre la calidad de los vinos y su nivel de alcohol. Los vinos de alta calidad suelen presentar una concentración de alcohol más elevada, destacando así la relevancia de este elemento en la percepción de calidad. Asimismo, se encontró una correlación inversa entre la densidad y la presencia de alcohol, subrayando la relevancia de estos factores en la apreciación global del vino. Al evaluar la densidad, se concluyó que los vinos con menor densidad tienden a tener una mejor valoración de calidad. Esto sugiere que la densidad es un factor crucial que influye en la percepción sensorial del vino.

Las diferencias entre las variedades de uva y sus características particulares podría emplearse para mejorar tanto la venta como la elaboración del vino. Esto podría tomar en cuenta otros elementos como el período de conservación o los procesos de fermentación para profundizar en su influencia en la calidad del vino.

Este estudio validó las hipótesis propuestas y brindó un entendimiento más detallado de las variables críticas que afectan la calidad del vino. Estos descubrimientos no solo resultan beneficiosos para perfeccionar los productos existentes, sino también para orientar las estrategias de producción.

Bibliografía

1. [Diferenciación de vinos por su azúcar residual](#)
2. [La fermentación alcohólica. Qué es y por qué afecta la calidad de los vinos](#)
3. [¿Qué es el cuerpo del vino?](#)
4. [Sulfuroso, el antiséptico histórico del vino. Su funciones y los sulfitos](#)
5. [What is Central Tendency Bias?](#)