

Informe Trabajo Práctico Especial

Grupo 18

Calidad de vino - Bodega del Sol

Integrantes:

- Gonzalez, Simón
- Tocco, Martiniano
- Turri, Pablo

Análisis de datos sobre el dataset: *wine quality*

- Análisis Exploratorio

Con el objetivo de realizar el análisis de datos sobre el dataset propuesto, primero se realizará una preparación y posterior exploración de los datos.

Para esto, realizaremos un análisis exploratorio de los datos para así describir el comportamiento de las variables. Analizar distribuciones. Comprender patrones, relaciones y tendencias en los datos. Identificar problemas que puedan afectar el análisis, tanto como valores nulos, repetidos, valores atípicos o outliers.

También, informar el análisis y descubrir posibles hipótesis.

Iniciamos realizando el análisis univariado por cada variable del dataset.

En el caso del **alcohol** que se mide en porcentaje de volumen, hemos observado que ésta columna es de tipo objeto.

Optamos por convertirla a tipo float, insertando NaNs en los registros donde no se podía convertir a un valor numérico. (Esto sucedía en 31 registros).

A la vez, decidimos transformar estos NaNs y cambiarlos por la mediana de la columna, para que no interrumpan en el análisis de los datos, considerando que éstos solo representan el 1.2% de la muestra total.

- Limpieza de Datos

Para realizar una apropiada limpieza de los datos tendremos en cuenta la estructura de los datos. El encoding y el formato de las variables. Los valores faltantes, la conversión de unidades.

Si encontramos datos faltantes, duplicados o valores atípicos durante el análisis exploratorio, detallaremos las acciones correctivas que aplicaremos. Esto puede incluir:

- Imputación de valores nulos.
- Eliminación de duplicados.
- Transformación de variables.

Con el `head()` observamos los primeros 5 registros del dataset, así podemos ver como se reflejan los datos y que no hayan ocurrido errores en la carga de estos.

Realizamos el método `info()` para ver la cantidad de registros que tenemos, el número de columnas, la cantidad de valores nulos en ellas, y el tipo de dato (`dtype`) asociado a cada una.

También realizamos el método `describe()`, que sirve para obtener las estadísticas principales de las variables (cargadas numéricamente).

Luego de realizar estos métodos vemos que el dataset contiene un total de 3231 filas, las cuales 588 de ellas se repiten (el 18.2%).

Las eliminaremos del dataset para trabajar con datos consistentes y no con varias muestras repetidas. Además observamos que en las columnas no hay valores nulos ni NAN.

Luego de realizar la eliminación de filas repetidas podemos ver que nos quedan 2643 filas en el dataset.

A partir de ahora realizaremos el **análisis univariado**:

Análisis univariado

Primero analizamos la variable **type** que es el tipo de uva con la que se elabora el vino. Es un tipo de dato categórico nominal. Almacenado como *object* en el dataset.

Aquí podemos ver que solo tenemos dos tipos de uvas distintas con las que se elaboraron los vinos en el dataset. Estas son *moscatel* y *syrah*.

El primer tipo mencionado, **moscatel**, es una variedad de uva blanca o morada de grano redondo muy dulce, caracterizada por su aroma intenso y su elevado contenido de azúcar. Presenta un tamaño de medio a grande, con piel gruesa y una amplia gama de colores que van del verde claro al amarillo dorado, el rosa y el negro.

La maduración de esta uva suele ser temprana a media. Se caracteriza principalmente por su necesidad de climas calurosos y de humedad, es decir, de la influencia del mar para su mejor cultivo.

Se utiliza para elaborar vinos secos: ideales para acompañar aperitivos y platos ligeros. (La uva moscatel se suele mezclar con otras variedades blancas).

También para vinos dulces: perfectos para acompañar postres. (Aquí se adiciona alcohol vínico para interrumpir la fermentación).

Para vinos de licor, obtenidos mezclando vinos o mostos con alcohol.

También para vinos espumosos; ligeros, frescos y afrutados.

Y para los de coupage (realizados mezclando distintas variedades de uva o distintos tipos de vinos ya fermentados).

El otro tipo de uva presente, **syrah**, también conocida como shiraz, es una uva tinta con bayas pequeñas y oscuras, de color negro-azulado y de forma elipsoidal.

Los vinos de Syrah se caracterizan por un color rojo profundo, acidez equilibrada, y una alta carga de taninos. Estos taninos aportan estructura y longevidad al vino, permitiendo que los vinos producidos con este tipo de uva envejezcan y evolucionen durante muchos años en botella. Pueden envejecer entre 5 a 25 años, y tienden a tener alta graduación alcohólica y mucho cuerpo.

Este tipo de uva es muy resistente a enfermedades y tiene una gran capacidad de prosperar en variedad de climas diferentes.

Con la uva Syrah suelen producirse vinos varietales (vino elaborado principalmente a partir de una sola variedad de uva), vinos tintos secos, rosados y espumosos.

La próxima variable a analizar es **fixed acidity (acidez fija)**, que es la cantidad de ácidos no volátiles presentes en el vino, medida en gramos por litro.

Este es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

La acidez fija es la suma de los ácidos naturales procedentes de la uva (tartárico, málico, cítrico y succínico) o formados en la fermentación (láctico) en el vino, que son poco volátiles, es decir, no se pueden separar del vino por evaporación durante la fermentación y el envejecimiento. Actúan como un preservante natural del vino, inhibiendo el desarrollo de microorganismos y contribuyendo a su longevidad.

El ácido fijo afecta el sabor y la estructura del vino, contribuyendo a la estabilidad y conservación de este.

Los niveles de estos ácidos en el vino pueden variar mucho. La acidez del vino es fundamental para su sabor, aspecto y elaboración. Un vino con una acidez más alta tendrá un sabor más crujiente y alquitranado, mientras que un vino con una acidez más baja tendrá un sabor más rico y redondo.

El rango que puede tener un vino de acidez fija es muy amplio, suele estar entre los 1.7 y 17.5 gramos por litro.

Volatile acidity (acidez volátil) es otra de las variables a analizar. Esta representa la cantidad de ácidos volátiles presentes en el vino, medida en gramos por litro.

Se compone principalmente de ácido acético, y suele tener valores normales entre 0.3 y 0.7g/L.

Estos ácidos, en determinadas concentraciones, provocan defectos en los vinos por lo que se procura mantenerlos en los niveles más bajos que sea posible durante la fermentación.

Valores superiores a 1 g/L pueden indicar que el vino está picado o sufrió una alteración, y tener aromas que recuerdan al vinagre y al barniz.

Los valores acéticos superiores a 1,2-1,5 g/L ya no son legales para el vino (el límite específico depende de la legislación local), a excepción de los vinos botirizados, donde puede alcanzar hasta 2,1 g/L.

Se dice que la calidad de un vino es mucho mayor cuando su acidez volátil es menor.

Este es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

Citric acid (ácido cítrico) es el contenido de ácido cítrico en el vino, en gramos por litro.

Es un tipo de ácido que se encuentra naturalmente en la uva, pero desaparece lentamente al ser fermentado por las bacterias, por lo que está presente en el vino pero en cantidades relativamente pequeñas comparadas con otros ácidos como el tartárico y el málico.

La concentración de éste suele estar entre 0.1 y 1 g/L.

Este es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

Residual sugar (azúcar residual), es la cantidad de azúcar de la uva natural que no se convierte en alcohol durante la fermentación.

Se compone principalmente del tipo de azúcar fructosa (azúcar de fruta).

El rango normal de su cantidad es muy amplio, ya que depende mucho de qué tipo de vino es producido.

El azúcar residual se mide en gramos por litro (g/L) y puede variar significativamente dependiendo del estilo del vino. Los valores pueden oscilar entre 0 y más de 120 g/L..

En vinos secos: menos de 4 gramos de azúcar por litro. En Vinos semisecos: entre 12 y 18 gramos de azúcar por litro. En vinos semidulces: entre 18 y 45 gramos de azúcar por litro.

En vinos dulces: más de 45 gramos de azúcar por litro. Vinos de cosecha tardía: alrededor

de 120g/L

Este es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

Chlorides (cloruros) representa la concentración de cloruros (sales) en el vino, medida en gramos por litro.

Estos pueden estar presentes naturalmente en el vino debido a la composición del suelo y el agua utilizada en la viticultura. También pueden ser introducidos durante el proceso de vinificación, para ajustar el pH del vino.

Los valores normales de cloruros en el vino varían según el tipo de vino:

- **Vinos Tintos:** 0.08 a 0.18 g/l de cloruro de sodio.
- **Vinos Blancos:** 0.14 a 0.35 g/l de cloruro de sodio

La INV establece en Argentina un límite de 1 g/L de cloruros en el vino.

Este es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

Free sulfur dioxide (cantidad de dióxido de azufre libre) es un compuesto químico importante en la elaboración del vino, utilizado principalmente por sus propiedades antimicrobianas (inhibe el crecimiento de bacterias y levaduras no deseadas, protegiendo el vino de posibles contaminaciones) y antioxidantes (previene la oxidación del vino, ayudando a conservar su color y sabor).

El dióxido de azufre libre es la fracción que no se ha combinado con otros compuestos en el vino. Es esencial para mantener la calidad y estabilidad del vino durante su almacenamiento y transporte.

Esta medida en mg/L y es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

La cantidad de SO₂ libre en el vino es cuidadosamente controlada, ya que un exceso puede afectar negativamente el sabor y aroma del vino, mientras que una cantidad insuficiente puede no proporcionar la protección necesaria.

La cantidad de este suele ir de 0.5 mg/L a 80 mg/L. Esta variación depende del pH del vino.

total sulfur dioxide (dióxido de azufre total) representa la suma del dióxido de azufre libre (free sulfur dioxide) y el combinado en el vino, en miligramos por litro.

Este es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

La regulación máxima en Europa es no sobrepasar los 50mg/L, mientras que en EEUU se permiten hasta 350mg/L en un vino.

En vinos tintos no se pueden superar los 150 mg/L. Pero si contenido de azúcar residual es mayor o igual a 5 g/L, el límite aumenta a 200 mg/L.

Vinos blancos y rosados: límite de 200 mg/L, o de 250 mg/L para aquellos con 5 g/L o más de azúcar residual.

density (densidad): medida de la masa por unidad de volumen del vino, utilizada para estimar la concentración de sólidos disueltos, medida en gramos por centímetro cúbico.

Los valores normales rondan de 0.9 a 1.07 g/cm³.

La densidad está relacionada con su contenido alcohólico. Cuanto más alcohol tenga un vino, más baja será su densidad.

Este es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

El **pH** representa la medida de la acidez o alcalinidad del vino.

Este oscila entre 2.8 y 4. Este rango se encuentra en el lado ácido de la escala de pH, que va desde 0 (muy ácido) a 14 (muy alcalino).

Depende de la concentración de ácidos, que varía según la variedad de uva, el clima, las fermentaciones y los métodos de elaboración.

Es importante tener una acidez adecuada para la calidad y estabilidad del vino.

Un pH bajo (más ácido) ayuda a mantener la frescura y la estabilidad del vino, mientras que un pH alto (menos ácido) puede aumentar el riesgo de oxidación y contaminaciones microbianas. Se intenta que los vinos no tengan un pH cercano o superior a 4 porque eleva el riesgo de oxidación y aumenta el riesgo de contaminaciones microbianas.

Los vinos blancos suelen estar entre 3 y 3,3 mientras que la mayoría de los tintos suelen estar entre 3,3 y 3,6.

Este es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

La variable **sulphates (sulfatos)** representa la concentración de sales de sulfato en el vino, en gramos por litro.

Estos cumplen funciones de ser:

- Conservantes: Evitan la oxidación del vino y mantienen sus propiedades organolépticas.
- Antioxidantes: Protegen el vino de la oxidación, preservando su color, sabor y aroma.
- Antimicrobianos: Inhiben el crecimiento de microorganismos indeseados, como bacterias y levaduras, que podrían alterar el vino.
- Antioxidásicos: Ayudan a prevenir la formación de compuestos oxidados que podrían afectar la calidad del vino.

El Instituto Nacional de Vitivinicultura (INV), establece límites para las cantidades de sulfatos en los vinos:

- Vinos secos: 1,30 g/l.
- Vinos edulcorados: 1,50 g/l.
- Vinos con más de 4 g/l de azúcares reductores: 1,50 g/l.
- Añejados mínimo 2 años en barrica: 2 g/l.

Este es un tipo de dato numérico continuo, cargado en el dataset de tipo *float64*.

La variable **alcohol** representa el contenido alcohólico del vino, medido en porcentaje de volumen (% vol).

Este es un tipo de dato numérico continuo pero está cargado como tipo *object*, lo que nos obligará a realizar una revisión sobre éste ya que podría ser que contenga valores no numéricos o datos inconsistentes, por lo que veremos que los datos que contenga sean adecuados para el posterior análisis.

Para que un vino se califique como tal, debe tener una graduación alcohólica mínima de 9% vol.

- **Rangos Comunes:**
 - **Vino Blanco:** Entre 9% y 12% vol.
 - **Vino Rosado:** Entre 9% y 12% vol.
 - **Vino Tinto:** Entre 12% y 15% vol.
 - **Vino Espumoso:** Entre 12% y 14% vol.

La máxima concentración alcohólica de un vino es de 15% vol.
Por lo tanto el rango va de un 9% a 15%.

La graduación alcohólica depende de varios factores: Concentración de Azúcares: La cantidad de azúcares presentes en la uva, que se convierten en alcohol durante la fermentación

Tipo de Uva: Diferentes variedades de uva tienen distintos niveles de azúcar y, por lo tanto, producen vinos con diferentes graduaciones alcohólicas

Madurez de la Uva: La madurez de la uva afecta directamente la cantidad de azúcar disponible para la fermentación

Clima y Viticultura: El clima y las prácticas de viticultura influyen en la madurez y el contenido de azúcar de la uva

Procesos Fermentativos y Tratamientos Posteriores: Los métodos de fermentación y cualquier tratamiento posterior pueden afectar la graduación final del vino

Por último, la variable **quality (calidad)** representa la puntuación del vino, con una escala que va de 0 a 10.

Este es un tipo de dato numérico discreto, cargado en el dataset de tipo int64.

Hipótesis

A partir de los gráficos de dispersión y las proyecciones realizadas con métodos de reducción de dimensionalidad generados en la notebook de Jupyter, así como de diversas fuentes de información consultadas, formulamos las siguientes hipótesis:

- La acidez fija está relacionada con el tipo de uva utilizado para elaborar el vino.
- La graduación alcohólica se relaciona con el tipo de uva.
- La densidad del vino varía según el tipo de uva utilizada para su elaboración.
- La densidad está relacionada con el contenido alcohólico.
- La calidad del vino se puede predecir mediante el contenido de alcohol.
- El tipo de uva utilizado se relaciona con la cantidad de azúcar residual.
- La acidez volátil se relaciona con el tipo de vino.
- La calidad del vino puede predecirse mediante una combinación de variables.

En este informe decidimos estudiar las siguientes hipótesis:

La acidez fija está relacionada con el tipo de uva utilizado para elaborar el vino.

La hipótesis fue formulada a través de la proyección PCA de nuestros datos. Luego de un estudio de las variables concluimos que la hipótesis es correcta, y mirando el gráfico realizado pareciera que los vinos elaborados con uva tipo Syrah suelen tener una mayor acidez fija que los elaborados con uva Moscatel, pero esto no está comprobado.

La graduación alcohólica del vino depende del tipo de uva utilizada.

Esta hipótesis surgió a partir del análisis de la proyección PCA de nuestros datos. Luego de un análisis pudimos concluir que nuestra hipótesis es acertada, y podría inferirse que los vinos elaborados con uva de tipo Moscatel tienen mayor graduación alcohólica respecto a los de tipo Syrah, pero esto no está comprobado de manera matemática.

La densidad del vino varía según el tipo de uva utilizada para su elaboración.

La hipótesis fue formulada a través de la proyección PCA de nuestros datos. Después de realizar los respectivos test pudimos concluir que la hipótesis es correcta, y los gráficos realizados sugieren que los vinos elaborados con uva de tipo Syrah poseen una densidad mayor a los elaborados con uva Moscatel, pero esto no está comprobado.

La densidad está relacionada con el contenido alcohólico.

Esta hipótesis surgió luego de estudiar y verificar las dos hipótesis anteriores, además de consultar con fuentes de internet:

[Tabla de densidad y peso específico de líquidos - TB01 \[Física - Estática de los fluidos\]](#)

[Vendimia 2015 – 6. Fermentación alcohólica | Paco & Lola.](#)

Luego de hacer los test pertinentes concluimos que la hipótesis es errónea y que no hay una relación fuerte entre las dos variables estudiadas.

Conclusión

El análisis realizado sobre el dataset de calidad de vinos de la Bodega del Sol nos permitió identificar diversos patrones y relaciones entre las características del vino y su calidad final. A través del proceso de limpieza y transformación de los datos, fue posible obtener un conjunto de información confiable y representativo. El análisis univariado y las hipótesis planteadas permitieron confirmar que algunas variables, como el tipo de uva, la acidez fija,

la graduación alcohólica y la densidad, presentan relaciones significativas con el perfil del vino, mientras que otras, como la relación entre la densidad y la graduación alcohólica, mostraron una independencia que contradice lo planteado previamente.

Las hipótesis sobre las diferencias de acidez y densidad según el tipo de uva sugieren que la uva Syrah tiende a producir vinos con mayor acidez y densidad que la uva Moscatel, mientras que esta última tiende a generar vinos con mayor graduación alcohólica.

En general, este trabajo nos permitió comprender mejor las características de los vinos y su relación con la calidad final.

Herramientas utilizadas:

<https://www.perplexity.ai/>

Fuentes utilizadas:

<https://www.vinoscutanda.com/blog/caracteristicas-de-la-uva-moscatel/>
<https://faustinatorivero.com/vino/uva-moscatel/>
<https://www.vinoseleccion.com/blog/vinos/vino-moscatel-uva/>
<https://faustinatorivero.com/vino/uva-shiraz/>
<https://vinedosbalmoral.com/es/caracteristicas-de-la-uva-syrah/>
<https://www.vinoseleccion.com/blog/vinos/vino-syrah/>
<https://www.gourmethunters.com/blog/es/todo-lo-que-debes-saber-acerca-del-vino-de-syrah>
<https://www.deepl.com/es/translator>
<https://terroirreview.com/2023/02/23/acidity-in-wine/>
<https://bodegagarzon.com/es/blog/acidez-vino/>
<https://www.aprenderdevino.es/acidosis-acidez-vino/>
<https://hannachile.com/2024/05/10/acidez-en-vinos/>
<https://sinatech.info/en/acetic-volatile-acidity-in-wine-cider-vinegars-and-juices/>
<https://www.aprenderdevino.es/acidosis-acidez-vino/>
<https://criollawine.com.ar/2023/04/26/acidez-en-el-vino/>
<https://bodegagarzon.com/es/blog/acidez-vino/>
<https://www.gourmethunters.com/blog/es/descubre-mas-sobre-el-vino-y-la-acidez-volatil/>
<https://www.bodegastrespiedras.com/acidez-del-vino/>
<https://altosdelitata.cl/acidez-del-vino-que-es-la-acidez-y-como-se-forma/>
<https://www.boutiqueduvin.com/es/blog/terminos-de-vino-8/azucar-residual-84>
<https://catatu.es/blog/el-azucar-en-el-vino/>
<https://www.cestalia.com/blog/azucar-en-vino-tinto-el-vino-tinto-lleva-azucar>
<https://glossaire.wein.plus/azucar-residual>
<https://foro.e-nologia.com/thread-37415-page-1.html>
<https://carlosserres.com/dioxido-azufre-funcion-vino/>
<https://www.hannabolivia.com/blog/post/687/importancia-la-cuantificacion-del-dioxido-azufre-libre-y-total-en-el-vino>
<https://t-d-i.es/quimica-del-dioxido-de-azufre-en-el-vino/>
<https://admin.lallemandwine.com/wp-content/uploads/2017/02/WE6-Argentina-1.pdf>
<https://www.usc.gal/caa/MetAnalisisStgo1/enologia.pdf>
<https://www.fisicanet.com.ar/fisica/estatica-de-fluidos/tb01-densidades-liquidos.php>
<https://www.enate.es/blog/que-es-el-ph-en-el-vino/>
<https://heredadaduna.com/el-ph-y-el-vino-que-relacion-tienen/>
<https://seguraviudas.com/es/noticias/vinos-y-cavas/ph-vino/>
<https://catatu.es/blog/ph-vinos/>
<https://raizdeguzman.com/que-es-ph-del-vino-como-influye-cualidades/>
<https://www.eufic.org/es/que-contienen-los-alimentos/articulo/que-son-los-sulfitos-del-vino-son-perjudiciales-para-la-salud/>
<https://www.enate.es/blog/los-sulfitos-del-vino/>
<https://www.conosur.com/news/por-que-hay-sulfitos-en-el-vino/>
<https://t-d-i.es/quimica-del-dioxido-de-azufre-en-el-vino/>