



LA CALIDAD: EL ATRIBUTO MAS IMPORTANTE EN UN VINO

Grupo de trabajo:

Daniela GODOY MIR &
Ezequiel BARLETTA PAZ

MOTIVO Y AUDIENCIA

Motiva el estudio del siguiente Dataset poder aportar información y herramientas a aquellos productores de vinos que quieran producir dicha bebida de la mejor manera posible buscando la calidad necesaria a fin de poder incrementar sus ventas considerablemente como también así enfocarse en los factores técnicos que realmente realzan la calidad aprovechando los recursos de manera mas eficiente.



MOTIVO Y AUDIENCIA

La calidad es el atributo más importante en un vino. A diferencia de otras características que involucran plenamente a la subjetividad de quien lo bebe (no solo el gusto personal sino también los recuerdos y emociones que un vino evoca), el concepto de calidad es puramente objetivo.

En el caso particular de la vitivinicultura, la calidad de un ejemplar involucra un sinfín de factores relacionados con la elaboración: desde la finca hasta la guarda.



PREGUNTAS Y OBJETIVOS DE LA INVESTIGACION



¿CUÁLES SON LAS VARIABLES QUE PODRÍAN AFECTAR A LA CALIDAD DEL VINO?

¿CUÁL ES LA VARIABLE QUE MÁS PODRÍA AFECTAR A LA CALIDAD DEL VINO?

¿ES NECESARIO AUMENTAR O DISMINUIR LA CANTIDAD DE DICHAS VARIABLES PARA AUMENTAR LA CALIDAD?

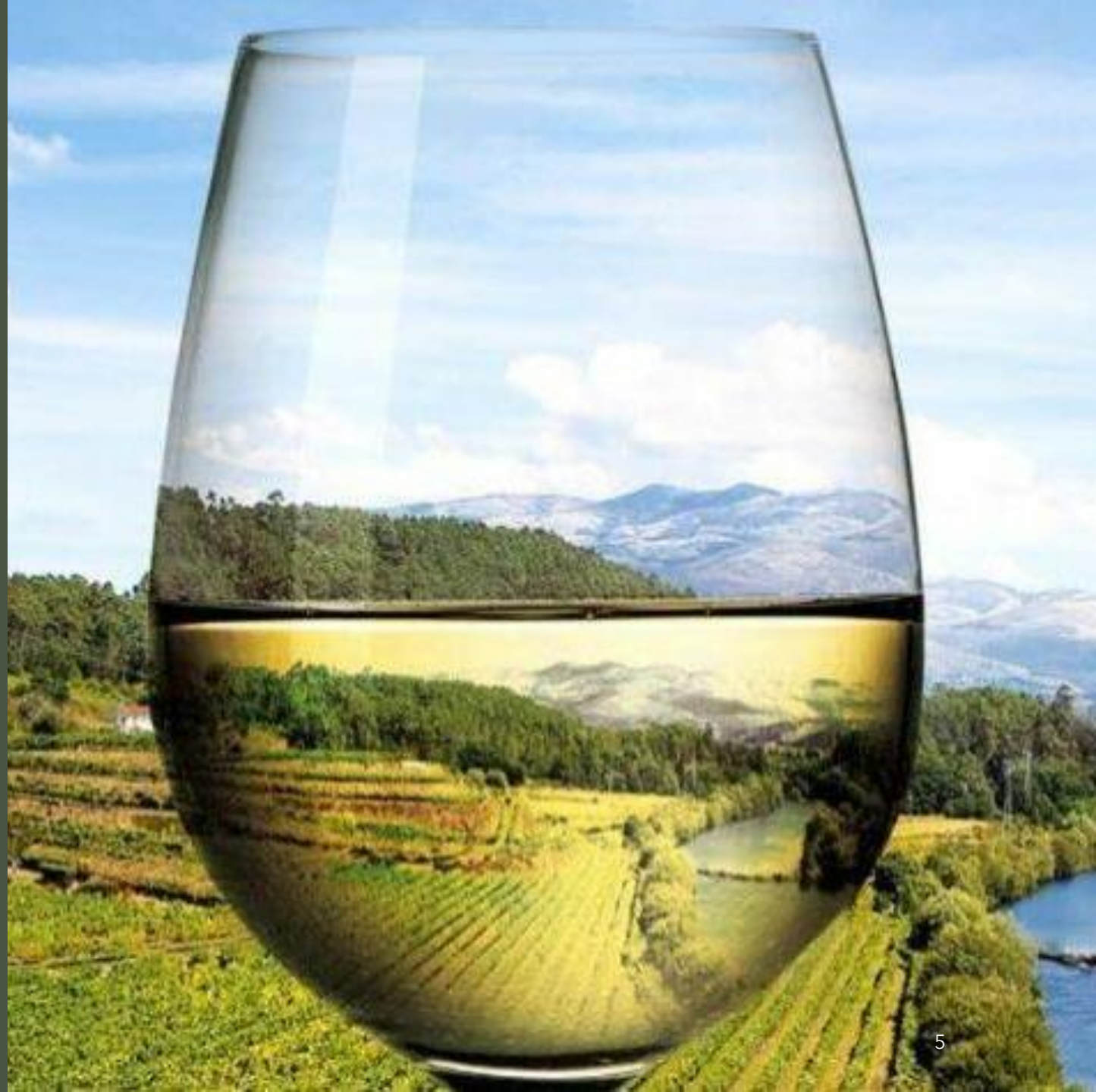
OBJETIVO: DETERMINAR AQUELLAS VARIABLES QUE HACEN A LA CALIDAD DE UN VINO

DATA ACQUISITION.

Este es un conjunto de datos relativos al vino tinto variante "Vinho Verde" portugués creado por Paulo Cortez de la Universidad de Minho, Guimaraes, Portugal para la Comisión de Vitivinicultura de la Región de Vinho Verde en Porto, Portugal cargado en el sitio UCI Machine Learning Repository.

El criterio utilizado para usar este dataset es que :

- Es un tema que nos resulta interesante.
- Es información de interés, ya que las variables allí incluidas pueden servir como análisis y asesoría a otras empresas del rubro.
- A simple vista, pareciera que los datos están completos.



DESCRIPCIÓN DE LAS VARIABLES QUE SE ENCUENTRAN EN ESTE DATASET:

1 - fixed acidity / acidez fija La mayoría de los ácidos involucrados con el vino o fijos o no volátiles (no se evaporan fácilmente).

2 - volatile acidity / acidez volátil La cantidad de ácido acético en el vino, que en niveles demasiado altos puede provocar un sabor desagradable a vinagre.

3 - citric acid / ácido cítrico Encontrado en pequeñas cantidades el ácido cítrico puede agregar 'frescura' y sabor a los vinos.

4 - residual sugar / azúcar residual La cantidad de azúcar que queda después de que se detiene la fermentación. Es raro encontrar vinos con menos de 1 gramo/litro y vinos con más de 45 gramos/litro se consideran dulces.

5 - chlorides / cloruros La cantidad de sal en el vino.

6 - free sulfur dioxide / dióxido de azufre libre La forma libre de SO₂ existe en equilibrio entre el SO₂ molecular (como gas disuelto) y el ion bisulfito; previene el crecimiento microbiano y la oxidación del vino.

7 - total sulfur dioxide / dióxido de azufre total Cantidad de formas libres y ligadas de SO₂; en bajas concentraciones, el SO₂ es mayormente indetectable en el vino, pero en concentraciones de SO₂ libres superiores a 50 ppm, el SO₂ se vuelve evidente en la nariz y el sabor del vino.

8 - density / densidad La densidad del agua es cercana a la del agua dependiendo del porcentaje de alcohol y azúcar contenido.

9 - pH / pH Describe qué tan ácido o básico es un vino en una escala de 0 (muy ácido) a 14 (muy básico); la mayoría de los vinos están entre 3 y 4 en la escala de pH

10 - sulphates / sulfatos Un aditivo del vino que puede contribuir a los niveles de dióxido de azufre (SO₂), que actúa como antimicrobiano y antioxidante

11 - alcohol / alcohol El porcentaje de contenido de alcohol del vino Output variable (based on sensory data):

12 - quality / calidad Variable de salida o target (basada en datos sensoriales, puntuación entre 0 y 10). Indica qué tan bueno es el vino en este estándar de calidad.





ANALISIS EXPLORATORIO

VARIABLE 'QUALITY'

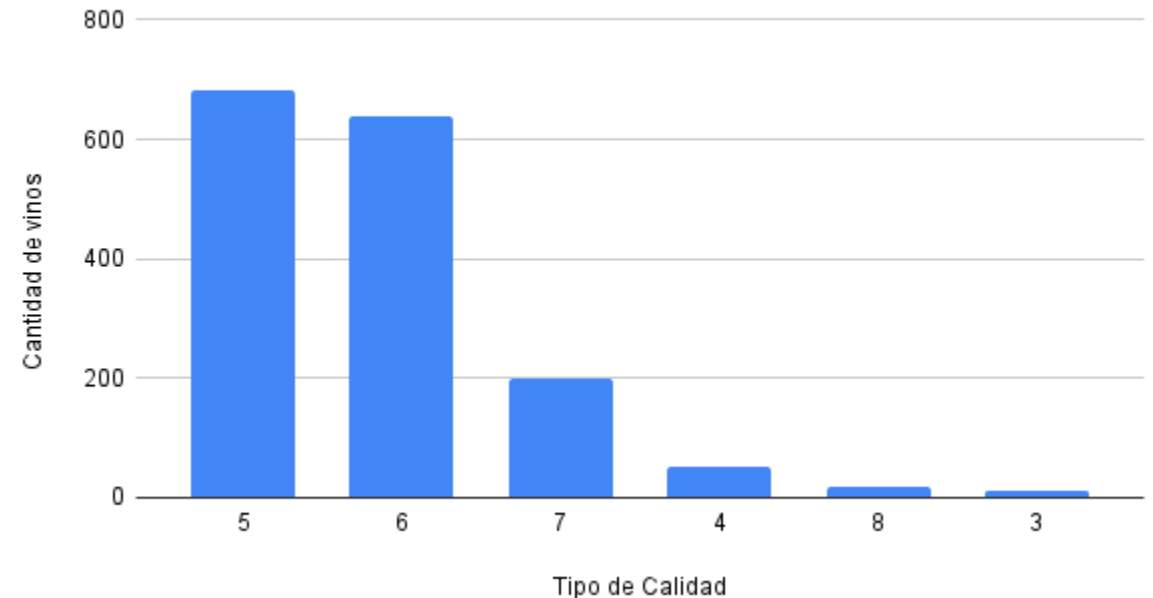
INVESTIGACIÓN

La variable "quality" es de tipo discreta y nos indica la calidad del vino en una escala de 0 al 10 donde este último valor es el de mayor calidad. En nuestro dataset, los valores analizados de calidad se encuentran en un rango que va de 3 a 8 y el valor que se repite mas es el 5.

CONCLUSIÓN:

Para evitar ruidos en las predicciones debido a la diferencia en la cantidad de datos que hacen referencia a calidad distinta a 5 y 6, no se tendrán en cuenta valores con calificaciones diferentes a 5 y 6.

Cantidad de vinos frente a Tipo de Calidad



RELACIONES MATEMATICAS

COMO NUESTRO TRABAJO SE CENTRA EN EL ESTUDIO DE LOS FACTORES QUE AFECTAN A LA CALIDAD, NOS CENTRAREMOS EN ESTA VARIABLE, ES POR ESO QUE SE OBSERVA QUE LAS VARIABLES QUE MAYOR RELACIÓN TIENEN CON LA CALIDAD SON:

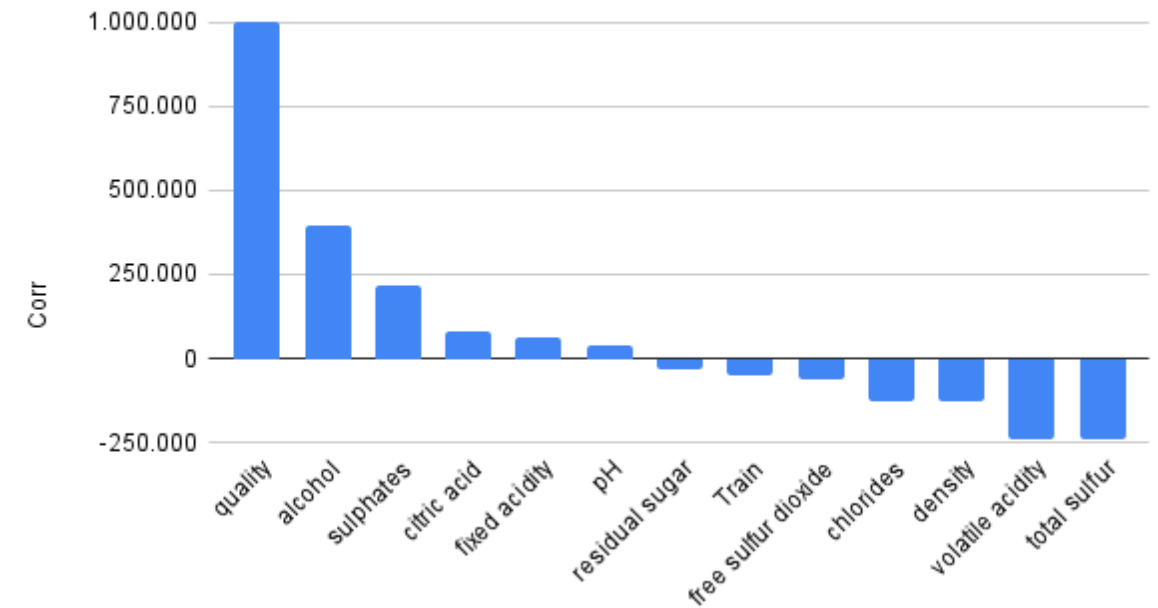
(Positivamente)

- Alcohol
- Sulphates

(Negativamente)

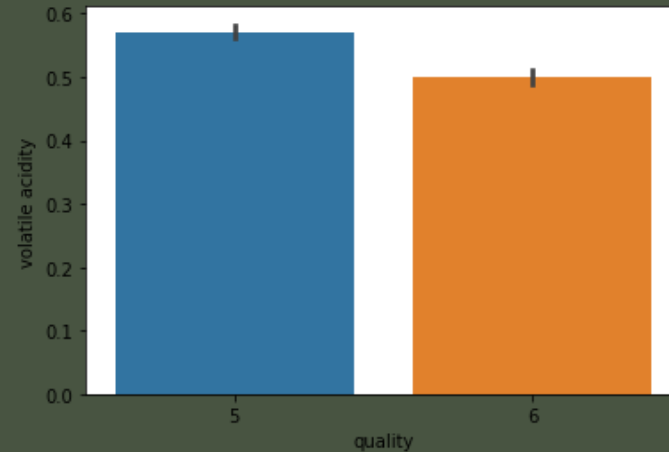
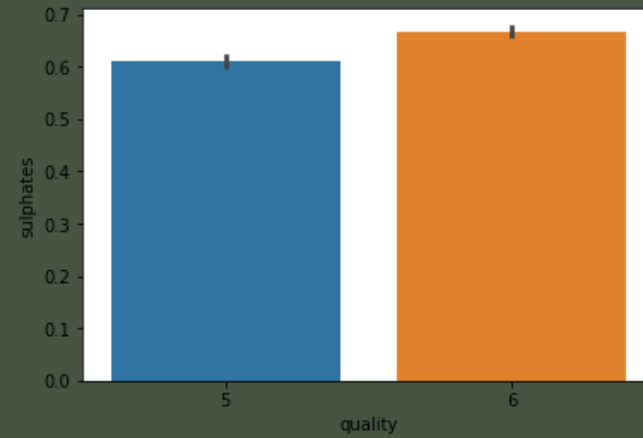
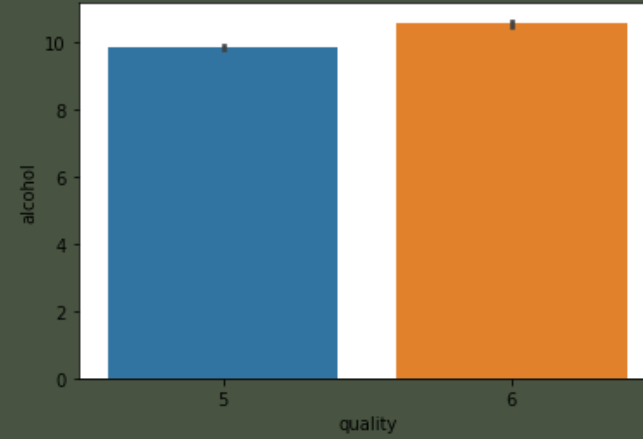
- Volatile acidity

Variables que mayor relación tienen con la calidad (quality)



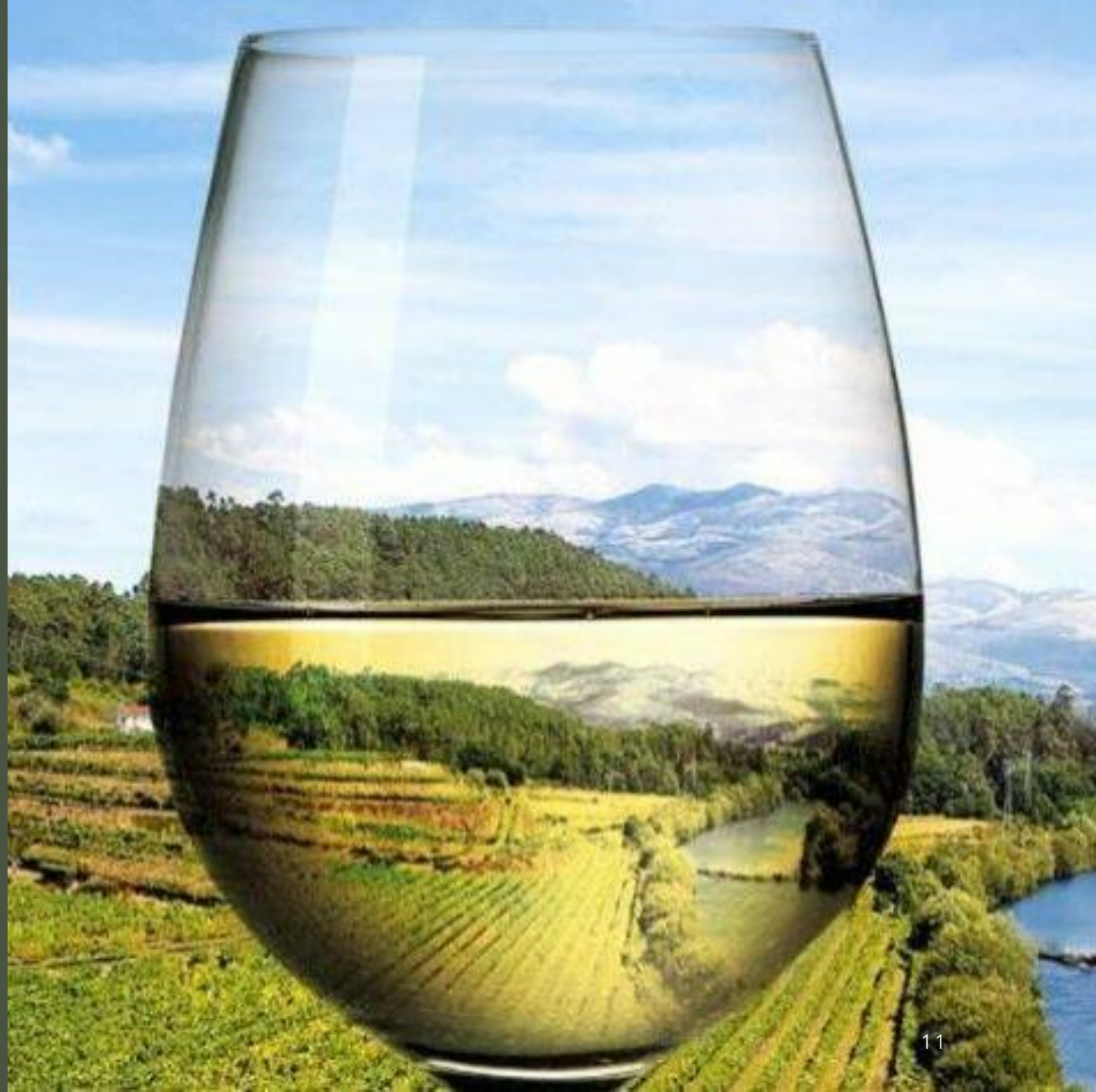
DEMOSTRACIÓN

CON AYUDA DE ESTOS
TRES GRÁFICOS PODEMOS
DECIR QUE A MAYOR
CANTIDAD DE "SULFATHES"
Y "ALCOHOL" LA CALIDAD
DEL VINO ES MEJOR,
MIENTRAS QUE LA
CANTIDAD DE LA VARIABLE
"VOLATILE ACIDITY" SOLO
AUMENTA LA CALIDAD DE
LOS VINOS INFERIORES A
LA CALIDAD 6, CUYA
CALIDAD ES LA MAS ALTA



RESUMEN

Luego de analizar todas las variables podemos ver y deducir que luego de aplicar a nuestros datos EDA llegamos a la conclusión de que las variables que condicionan en mayor proporción la calidad de un vino de estas características son la graduación alcohólica (alcohol) y los sulfitos (sulphates)





GRACIAS

Alumnos:

Daniela GODOY MIR

Ezequiel BARLETTA PAZ

Curso: Data Science

Institucion: CODER HOUSE