

# Estudio de clasificación de calidad del vino

Genaro Velázquez Avilés



# Importancia del vino en la historia de la humanidad

El consumo de vino data desde los miles de años antes de nuestra era.

El humano ha desarrollado un gusto por esta bebida a través de los años.

Este conocimiento milenario ha llevado a tener diversidad en tipos de vino.

---

# Beneficiarios del estudio

- ❏ Vinícolas
    - Mejores ventas
    - Aumento de precio
    - Mayor prestigio
  - ❏ Enólogos
    - Adquirir conocimiento
    - Mejorar procedimientos
-

¿Cuáles son los  
elementos y propiedades  
físico-químicas que hacen  
que un vino tenga una  
calidad más alta?





## Para un buen *Marketing*:

- ❑ Vino superior a la media
- ❑ Calidad
- ❑ Certificado



## Referencia:

Tenemos la base de datos<sup>1</sup> de las propiedades físico-químicas de vinos de origen portugueses.

<sup>1</sup>M Yasser H (2021). Wine Quality Dataset. [Data set]. <https://www.kaggle.com/datasets/yasserh/wine-quality-dataset>

# Las propiedades obtenidas en este estudio son:

- ❑ Tipo de vino
- ❑ Acidez fija
- ❑ Acidez volátil
- ❑ Ácido cítrico
- ❑ Azúcar residual
- ❑ Cloruros
- ❑ Dióxido de azufre libre
- ❑ Dióxido de azufre total
- ❑ Densidad
- ❑ pH
- ❑ Sulfatos
- ❑ Alcohol
- ❑ Calidad

# ¿Cuáles son las propiedades que más influyen en la calidad del vino?

Mediante el análisis de cada una de las propiedades encontramos una mediana correlación en las siguientes variables:

- Alcohol

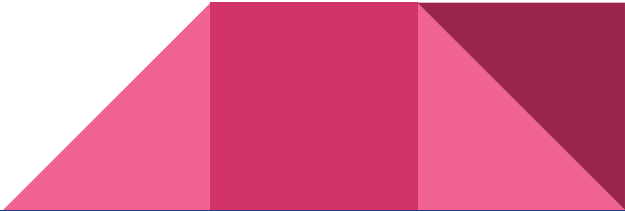
Y una baja correlación en:

- Acidez volátil

- Densidad







¿Cuál es la influencia del pH en la calidad del vino?

## pH Scale



Las características más representativas del vino son:

- ❑ Color
- ❑ Sabor

El pH es uno de los responsables más importantes de estas características ya que mejora el envejecimiento del vino.

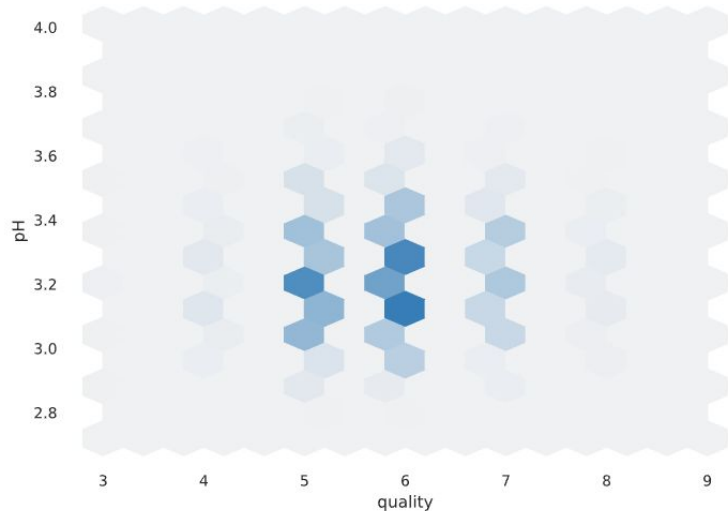
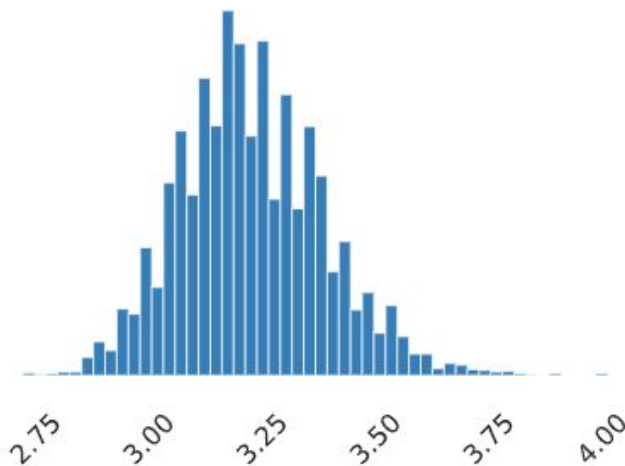
Un vino que envejece bien da colores más brillantes.



Min: 2.72

Máx: 4.01

Media  $\approx$  3.22



Podemos observar que entre más alta es la calidad del vino el valor del pH debe estar cerca de la media.

# Modelos de Machine learning

# Modelos a usar

Se utilizaron 5 modelos

- C-Support Vector Classification
- DecisionTreeClassifier
- KNeighborsClassifier
- AdaBoostClassifier
- Multi-layer Perceptron classifier.

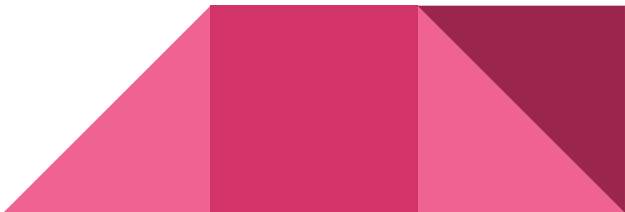


# El mejor modelo para este database

El modelo con mejor desempeño fue el Multi-layer Perceptron classifier.

	precisión	support
3	0.00	13
4	0.57	60
5	0.64	466
6	0.47	530
7	0.41	211
8	0.00	43
9	0.00	1

Destacando un 50% de aciertos con una validación cruzada de los datos.



# Conclusión

Para este conjunto de datos pudimos encontrar que efectivamente hay correlación entre los datos, sin embargo no son suficientes para los modelos vistos tener un porcentaje de precisión alto.

Es decir que con la información que tenemos un mejor porcentaje de aciertos en la calidad de los vinos del rango de 4 a 7.





Anexo

<https://github.com/GenVel/CoderHouseDesafios>