## **Preguntas para que respondan con Wine Quality**

Aquí tienes unas **15 preguntas mezcladas** de limpieza, preprocesamiento y EDA básico:

1. ¿Hay valores duplicados en el dataset? ¿Cuántos?
2. ¿Qué columnas tienen valores nulos?
3. ¿Qué variable muestra más variabilidad? (usa describe()).
4. ¿Cuál es la media y mediana de *alcohol*? ¿Hay diferencias notables?
5. ¿Qué valores de *residual sugar* podrían considerarse outliers (usando percentiles 1% y 99%)?
6. Agrupa los vinos en 3 categorías de calidad: bajo (≤4), medio (5–6), alto (≥7). ¿Cuántos hay en cada grupo?
7. ¿Cuál es la correlación entre *alcohol* y *quality*?
8. ¿Y entre *density* y *alcohol*? ¿Qué interpretación tiene?
9. Crea una nueva columna alcohol\_sugar\_ratio = alcohol / residual sugar. ¿Qué vinos tienen el ratio más alto?
10. ¿Qué variable parece más importante para predecir la calidad del vino? (según correlaciones).
11. ¿Qué proporción de vinos tiene pH menor que 3.2?
12. ¿Hay vinos con valores de *fixed acidity* por debajo de 5 o por encima de 15? ¿Qué harías con ellos?
13. Si tuvieras que imputar nulos en *citric acid*, ¿qué estrategia usarías: media, mediana o moda? ¿Por qué?
14. Convierte la variable *quality* en binaria: “bueno” si ≥7, “no bueno” si <7. ¿Qué proporción de vinos “buenos” hay?
15. Si construyeras un modelo, ¿qué variables descartarías y cuáles transformarías?