

Introducción

Este informe explora un conjunto de datos de vinos tintos. El objetivo es entender qué características se relacionan con la calidad del vino según una calificación numérica. Para ello se hace una limpieza básica, un resumen de los datos, una revisión de relaciones entre variables y varias gráficas que ayuden a “ver” los patrones.

Metodología

Trabajé en un cuaderno de Jupyter con Python.

Pasos: cargar el archivo, limpiar filas repetidas, ver un resumen de los números, revisar relaciones entre variables y crear gráficos para interpretar mejor los resultados.

Carga de datos

Se cargó el archivo winequality-red.csv, que contiene 1599 vinos evaluados en 12 características (por ejemplo: alcohol, acidez, azúcares, densidad, etc.) y una calificación de calidad. La calificación es un número entero.

Limpieza de datos

- Se detectaron y eliminaron 240 filas duplicadas (registros repetidos).
- No se encontraron valores vacíos (NA).
- La columna de calidad se dejó como número entero para representar mejor las notas.
- Resultado: 1359 registros limpios para el análisis.

Por qué importa: si dejamos duplicados o datos sucios, los resultados pueden inclinarse o confundir las conclusiones.

Resumen de los datos

La tabla de resumen muestra, para cada variable, valores como promedio, mínimo, máximo y puntos intermedios.

```
•[3]: #RESUMEN ESTADISTICO
df_describe = df.describe()
print(df_describe)
```

	fixed acidity	volatile acidity	citric acid	residual sugar	\
count	1599.000000	1599.000000	1599.000000	1599.000000	
mean	8.319637	0.527821	0.270976	2.538806	
std	1.741096	0.179060	0.194801	1.409928	
min	4.600000	0.120000	0.000000	0.900000	
25%	7.100000	0.390000	0.090000	1.900000	
50%	7.900000	0.520000	0.260000	2.200000	
75%	9.200000	0.640000	0.420000	2.600000	
max	15.900000	1.580000	1.000000	15.500000	

	chlorides	free sulfur dioxide	total sulfur dioxide	density	\
count	1599.000000	1599.000000	1599.000000	1599.000000	
mean	0.087467	15.874922	46.467792	0.996747	
std	0.047065	10.460157	32.895324	0.001887	
min	0.012000	1.000000	6.000000	0.990070	
25%	0.070000	7.000000	22.000000	0.995600	
50%	0.079000	14.000000	38.000000	0.996750	
75%	0.090000	21.000000	62.000000	0.997835	
max	0.611000	72.000000	289.000000	1.003690	

	pH	sulphates	alcohol	quality
count	1599.000000	1599.000000	1599.000000	1599.000000
mean	3.311113	0.658149	10.422983	5.636023
std	0.154386	0.169507	1.065668	0.807569
min	2.740000	0.330000	8.400000	3.000000
25%	3.210000	0.550000	9.500000	5.000000
50%	3.310000	0.620000	10.200000	6.000000
75%	3.400000	0.730000	11.100000	6.000000
max	4.010000	2.000000	14.900000	8.000000

	fixed acidity	volatile acidity	citric acid	\
fixed acidity	1.000000	-0.256131	0.671703	
volatile acidity	-0.256131	1.000000	-0.552496	
citric acid	0.671703	-0.552496	1.000000	
residual sugar	0.114777	0.001918	0.143577	
chlorides	0.093705	0.061298	0.203823	
free sulfur dioxide	-0.153794	-0.010504	-0.060978	
total sulfur dioxide	-0.113181	0.076470	0.035533	
density	0.668047	0.022026	0.364947	
pH	-0.682978	0.234937	-0.541904	
sulphates	0.183006	-0.260987	0.312770	
alcohol	-0.061668	-0.202288	0.109903	
quality	0.124052	-0.390558	0.226373	

	residual sugar	chlorides	free sulfur dioxide	\
fixed acidity	0.114777	0.093705	-0.153794	
volatile acidity	0.001918	0.061298	-0.010504	
citric acid	0.143577	0.203823	-0.060978	
residual sugar	1.000000	0.055610	0.187049	
chlorides	0.055610	1.000000	0.005562	
free sulfur dioxide	0.187049	0.005562	1.000000	
total sulfur dioxide	0.203028	0.047400	0.667666	
density	0.355283	0.200632	-0.021946	
pH	-0.085652	-0.265026	0.070377	
sulphates	0.005527	0.371260	0.051658	
alcohol	0.042075	-0.221141	-0.069408	
quality	0.013732	-0.128907	-0.050656	

	total sulfur dioxide	density	pH	sulphates	\
fixed acidity	-0.113181	0.668047	-0.682978	0.183006	
volatile acidity	0.076470	0.022026	0.234937	-0.260987	
citric acid	0.035533	0.364947	-0.541904	0.312770	
residual sugar	0.203028	0.355283	-0.085652	0.005527	
chlorides	0.047400	0.200632	-0.265026	0.371260	
free sulfur dioxide	0.667666	-0.021946	0.070377	0.051658	
total sulfur dioxide	1.000000	0.071269	-0.066495	0.042947	
density	0.071269	1.000000	-0.341699	0.148506	
pH	-0.066495	-0.341699	1.000000	-0.196648	
sulphates	0.042947	0.148506	-0.196648	1.000000	
alcohol	-0.205654	-0.496180	0.205633	0.093595	
quality	-0.185100	-0.174919	-0.057731	0.251397	

	alcohol	quality
fixed acidity	-0.061668	0.124052
volatile acidity	-0.202288	-0.390558
citric acid	0.109903	0.226373
residual sugar	0.042075	0.013732
chlorides	-0.221141	-0.128907
free sulfur dioxide	-0.069408	-0.050656

Sirve para tener una idea del rango típico de cada característica. Por ejemplo, permite ver cuál es el rango usual del alcohol o qué tan variable es la acidez entre los vinos.

Qué mirar: si una variable tiene un rango muy amplio, puede influir de manera distinta en la calidad según el caso.

Matriz de correlación

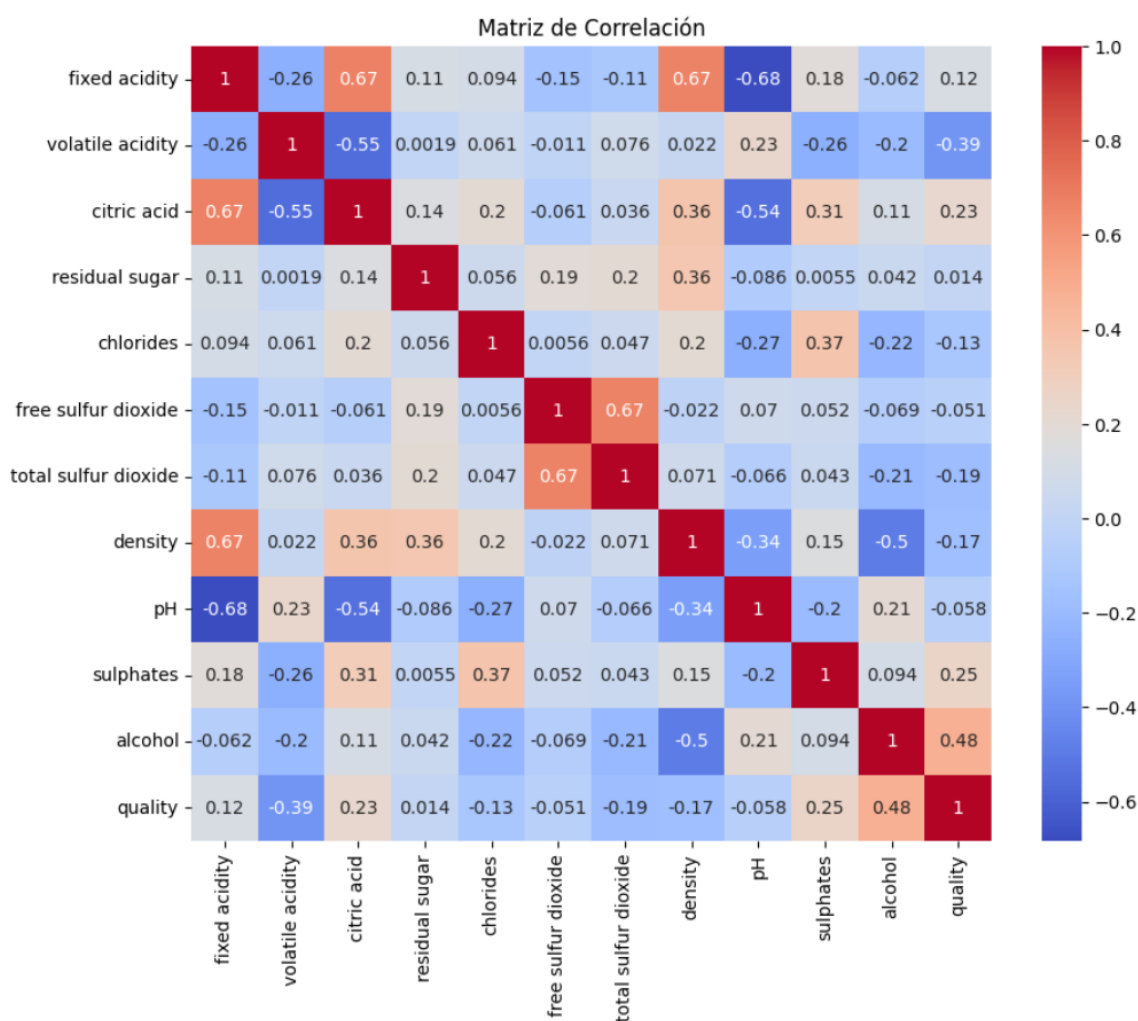
Este cuadro coloreado muestra qué tan relacionada está una variable con otra.

Colores hacia el rojo: cuando una sube, la otra tiende a subir también (relación positiva).

Colores hacia el azul: cuando una sube, la otra tiende a bajar (relación negativa).

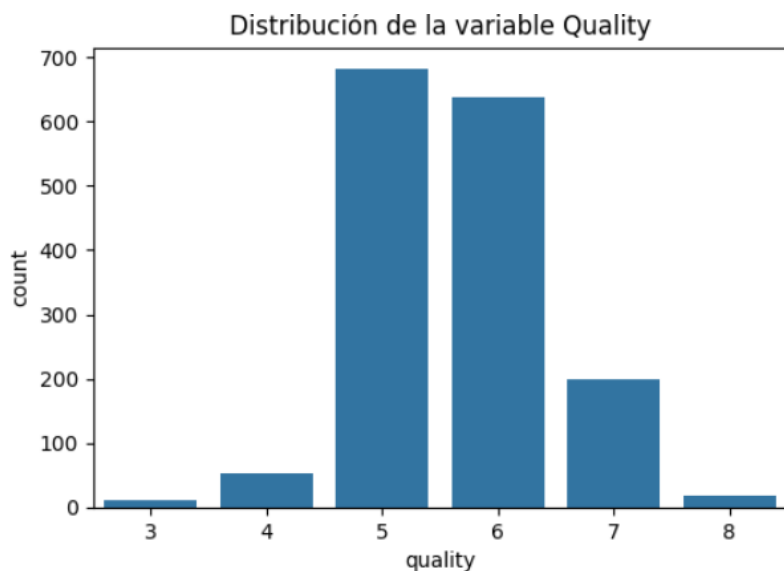
Cerca de cero: no hay una relación clara en línea recta.

La línea diagonal es 1 porque cada variable se compara consigo misma.



Lectura rápida: el alcohol aparece más ligado a la calidad (positivo); la acidez volátil, en cambio, se asocia a calidades más bajas (negativo).

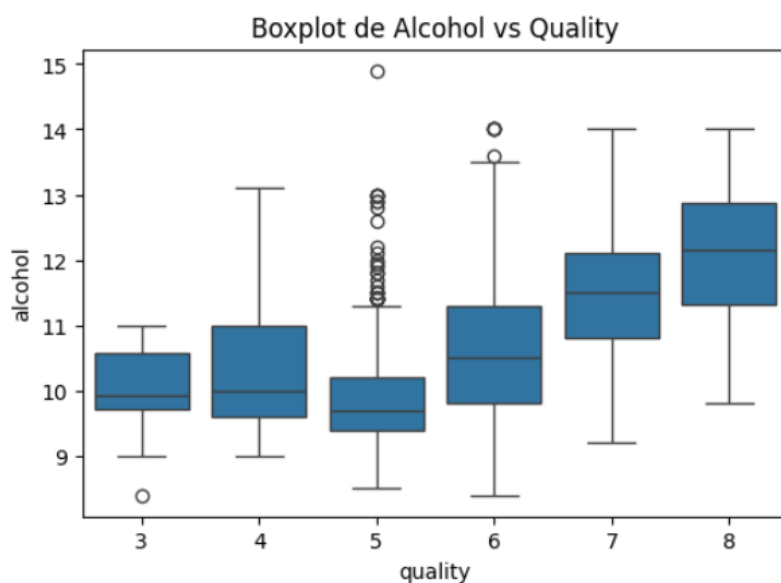
Gráfica 1 – Histograma de la calidad



“La mayoría de vinos se concentran en calidades intermedias (alrededor de 5–6). Los extremos, tanto muy bajos como muy altos, son menos frecuentes.”

Interpretación: la base de datos está balanceada hacia vinos “correctos” o “aceptables”, con pocos casos excelentes o muy malos.

Gráfica 2 – Dispersión: alcohol vs. Calidad

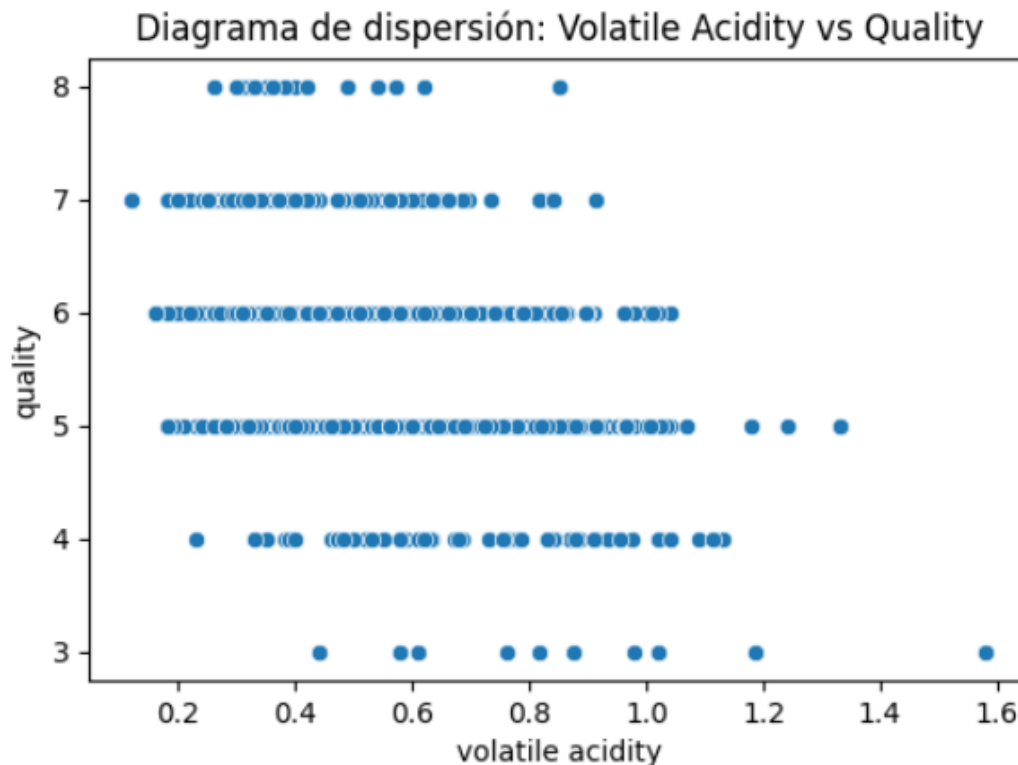


“Se observa una tendencia ascendente: a mayor alcohol, suele mejorar la calidad. No es una regla perfecta, pero el patrón es claro.”

Idea clave: el grado alcohólico podría estar actuando como un indicador indirecto de madurez del fruto o de estilo del vino, lo que favorece su calificación.

Gráfica 3 – Boxplot: acidez volátil por nivel de calidad

```
plt.show()
```



“Los vinos con mejor calidad tienden a mostrar acidez volátil más baja. La acidez volátil alta suele asociarse a aromas o sensaciones menos agradables.”

Idea clave: mantener controlada la acidez volátil parece importante para lograr vinos mejor valorados.

Conclusiones

- Más alcohol, mejor nota (en general): los vinos con mayor grado alcohólico tienden a recibir una calificación más alta.
- Acidez volátil alta, peor nota: cuando la acidez volátil sube, la calidad suele bajar.
- Calidad concentrada en valores medios: la mayoría de vinos están alrededor de la nota 6 (pocos muy malos o excelentes).
- Sulphates ayudan un poco: hay una relación positiva leve con la calidad, posiblemente por su efecto de protección/estabilidad del vino.