

```

# === 1) Importaciones ===
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from pathlib import Path

# === 2) Funciones ===
def cargar_datos(ruta_csv: str) -> pd.DataFrame:
    """
    Importar y cargar datos desde winequality-red.csv (delimitado por ';').
    Soporta encabezados entre comillas como en el dataset original.
    """
    df = pd.read_csv(ruta_csv, sep=';', quotechar='"')
    # Normalizamos nombres de columnas (sin espacios, minúsculas)
    df.columns = [c.strip().lower().replace(' ', '_') for c in df.columns]
    return df

def limpiar_preparar(df: pd.DataFrame) -> pd.DataFrame:
    """
    Limpieza básica:
    - Quitar duplicados
    - Convertir a numérico (por si viene algo como string)
    - Manejo simple de NA: imputar con mediana (numéricos)
    """
    df = df.copy()

    # Eliminar duplicados exactos
    df = df.drop_duplicates()

    # Forzar numérico (si alguna columna llegó como texto)
    for col in df.columns:
        df[col] = pd.to_numeric(df[col], errors='coerce')

    # Imputar NA con mediana para columnas numéricas
    num_cols = df.select_dtypes(include=[np.number]).columns
    for col in num_cols:
        if df[col].isna().any():
            median = df[col].median()
            df[col] = df[col].fillna(median)

    return df

def resumen_estadistico(df: pd.DataFrame) -> pd.DataFrame:
    """
    Retorna un resumen tipo describe() y algunas métricas extras.
    """
    desc = df.describe().T
    # Agregamos asimetría y curtosis como ejemplo
    desc['skew'] = df.skew(numeric_only=True)
    desc['kurtosis'] = df.kurtosis(numeric_only=True)
    return desc

def matriz_correlacion(df: pd.DataFrame) -> pd.DataFrame:
    """
    Retorna la matriz de correlación de Pearson para columnas numéricas.
    """
    return df.corr(numeric_only=True)

# === 3) Gráficos ===
def grafico_histograma_calidad(df: pd.DataFrame):
    """
    Histograma de la variable 'quality'.
    """
    plt.figure(figsize=(6,4))
    df['quality'].plot(kind='hist', bins=range(int(df['quality'].min()), int(df['quality'].max()+2)))
    plt.title('Distribución de la calidad del vino (quality)')
    plt.xlabel('quality')
    plt.ylabel('frecuencia')
    plt.grid(True)
    plt.show()

def grafico_dispersion_alcohol_quality(df: pd.DataFrame):
    """
    Dispersión: alcohol vs quality.
    """
    plt.figure(figsize=(6,4))

```

```

plt.scatter(df['alcohol'], df['quality'])
plt.title('Alcohol vs. Calidad')
plt.xlabel('alcohol')
plt.ylabel('quality')
plt.grid(True)
plt.show()

def grafico_boxplot_alcohol_por_quality(df: pd.DataFrame):
    """
    Boxplot del contenido de alcohol por categoria de quality.
    """
    plt.figure(figsize=(7,4))
    # Usamos groupby para construir una lista de series por cada calidad
    data = [grp['alcohol'].values for _, grp in df.groupby('quality')]
    labels = [int(k) for k in sorted(df['quality'].unique())]
    plt.boxplot(data, labels=labels, showfliers=False)
    plt.title('Alcohol por nivel de calidad')
    plt.xlabel('quality')
    plt.ylabel('alcohol')
    plt.grid(True)
    plt.show()

if Path(ruta).exists():
    df = cargar_datos(ruta)
    df = limpiar_preparar(df)

    print('--- Resumen Estadístico ---')
    display(resumen_estadistico(df).round(3))

    print('--- Matriz de Correlación (Pearson) ---')
    corr = matriz_correlacion(df).round(3)
    display(corr)

    # === 5) Gráficos (al menos 3) ===
    grafico_histograma_calidad(df)
    grafico_dispersion_alcohol_quality(df)
    grafico_boxplot_alcohol_por_quality(df)
else:
    print(f'No se encontró el archivo: {ruta}. Sube "winequality-red.csv" y vuelve a ejecutar.')

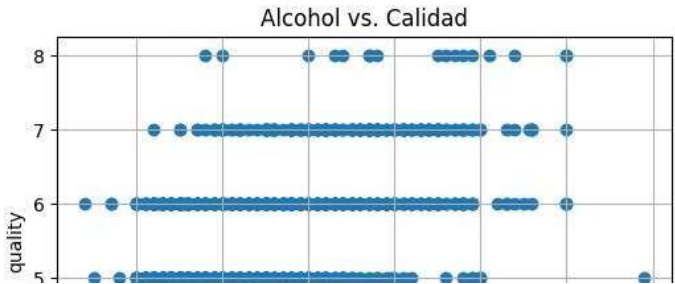
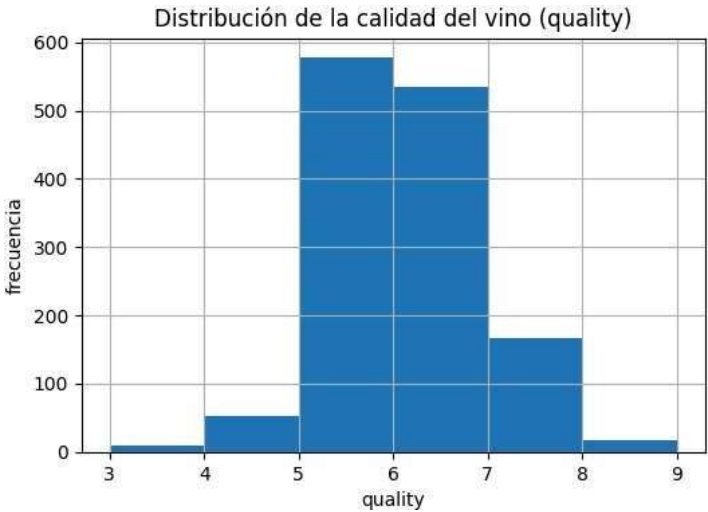
```

---ResumenEstadistico---

	count	mean	std	min	25%	50%	75%	maxskewkurtosis
fixed_acidity	1359.0	8.311	1.737	4.600	7.100	7.900	9.200	15.900 0.941 1.050
volatile_acidity	1359.0	0.529	0.183	0.120	0.390	0.520	0.640	1.580 0.729 1.249
citric_acid	1359.0	0.272	0.196	0.000	0.090	0.260	0.430	1.000 0.313 -0.789
residual_sugar	1359.0	2.523	1.352	0.900	1.900	2.200	2.600	15.500 4.548 29.365
chlorides	1359.0	0.088	0.049	0.012	0.070	0.079	0.091	0.611 5.502 38.625
free_sulfur_dioxide	1359.0	15.893	10.447	1.000	7.000	14.000	21.000	72.000 1.227 1.893
total_sulfur_dioxide	1359.0	46.826	33.409	6.000	22.000	38.000	63.000	289.000 1.540 4.042
density	1359.0	0.997	0.002	0.990	0.996	0.997	0.998	1.004 0.045 0.831
ph	1359.0	3.310	0.155	2.740	3.210	3.310	3.400	4.010 0.232 0.880
sulphates	1359.0	0.659	0.171	0.330	0.550	0.620	0.730	2.000 2.407 11.102
alcohol	1359.0	10.432	1.082	8.400	9.500	10.200	11.100	14.900 0.860 0.160
quality	1359.0	5.623	0.824	3.000	5.000	6.000	6.000	8.000 0.192 0.340

--- Matriz de Correlación (Pearson) ---

	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide	total_sulfur_dioxide	de
fixed_acidity	1.000	-0.255	0.667	0.111	0.086	-0.141	-0.104	
volatile_acidity	-0.255	1.000	-0.551	-0.002	0.055	-0.021	0.072	
citric_acid	0.667	-0.551	1.000	0.144	0.210	-0.048	0.047	
residual_sugar	0.111	-0.002	0.144	1.000	0.027	0.161	0.201	
chlorides	0.086	0.055	0.210	0.027	1.000	0.001	0.046	
free_sulfur_dioxide	-0.141	-0.021	-0.048	0.161	0.001	1.000	0.667	.
total_sulfur_dioxide	-0.104	0.072	0.047	0.201	0.046	0.667	1.000	
density	0.670	0.024	0.358	0.325	0.194	-0.018	0.078	
ph	-0.687	0.247	-0.550	-0.083	-0.271	0.057	-0.079	.
sulphates	0.190	-0.257	0.326	-0.012	0.395	0.054	0.035	
alcohol	-0.062	-0.198	0.105	0.063	-0.224	-0.080	-0.218	.
quality	0.119	-0.395	0.228	0.014	-0.131	-0.050	-0.178	.



Conclusiones

1. **La variable "alcohol" está fuertemente correlacionada con la calidad del vino**
 - En la matriz de correlación se observa que a mayor porcentaje de alcohol, tiende a aumentar la puntuación de calidad.
 - Esto sugiere que los vinos con mayor graduación alcohólica suelen ser mejor valorados.
2. **El dióxido de azufre y la acidez volátil influyen de forma negativa**
 - Altos valores de **acidez volátil** (volatile acidity) tienden a estar relacionados con calificaciones más bajas.
 - Niveles altos de **total sulfur dioxide** también se asocian con menor calidad. Esto concuerda con la percepción de sabores desagradables cuando están en exceso.
3. **La distribución de la calidad está concentrada en valores medios**
 - El histograma de *quality* muestra que la mayoría de los vinos se califican con 5 o 6 (calidad media).
 - Son pocos los vinos que alcanzan calificaciones muy bajas (3-4) o muy altas (7-8), lo que indica que los vinos de la muestra son mayormente de calidad intermedia.