

# T3\_201709282

March 4, 2025

- Tarea 3
- Carlos Javier Martinez Polanco
- 201709282

## 0.1 Importar las librerías necesarias

```
[3]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

## 1 Cargar los datos

```
[10]: df = pd.read_csv('winequality-red.csv', sep=";") # Cambia la ruta al CSV
```

## 2 Mostrar las primeras filas para inspección

```
[11]: print(df.head())
print(df.dtypes)
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	\
0	7.4	0.70	0.00	1.9	0.076	
1	7.8	0.88	0.00	2.6	0.098	
2	7.8	0.76	0.04	2.3	0.092	
3	11.2	0.28	0.56	1.9	0.075	
4	7.4	0.70	0.00	1.9	0.076	

	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	\
0	11.0	34.0	0.9978	3.51	0.56	
1	25.0	67.0	0.9968	3.20	0.68	
2	15.0	54.0	0.9970	3.26	0.65	
3	17.0	60.0	0.9980	3.16	0.58	
4	11.0	34.0	0.9978	3.51	0.56	

	alcohol	quality
0	9.4	5
1	9.8	5
2	9.8	5

```

3      9.8      6
4      9.4      5
fixed acidity      float64
volatile acidity   float64
citric acid        float64
residual sugar     float64
chlorides          float64
free sulfur dioxide float64
total sulfur dioxide float64
density            float64
pH                float64
sulphates          float64
alcohol            float64
quality            int64
dtype: object

```

### 3 Revisar información general

```
[12]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   fixed acidity          1599 non-null   float64
1   volatile acidity       1599 non-null   float64
2   citric acid            1599 non-null   float64
3   residual sugar         1599 non-null   float64
4   chlorides              1599 non-null   float64
5   free sulfur dioxide     1599 non-null   float64
6   total sulfur dioxide    1599 non-null   float64
7   density                1599 non-null   float64
8   pH                    1599 non-null   float64
9   sulphates              1599 non-null   float64
10  alcohol                1599 non-null   float64
11  quality                1599 non-null   int64
dtypes: float64(11), int64(1)
memory usage: 150.0 KB

```

### 4 Resumen estadístico

```
[13]: describe_df = df.describe()
print(describe_df)
```

```

      fixed acidity  volatile acidity  citric acid  residual sugar  \
count      1599.000000      1599.000000  1599.000000      1599.000000

```

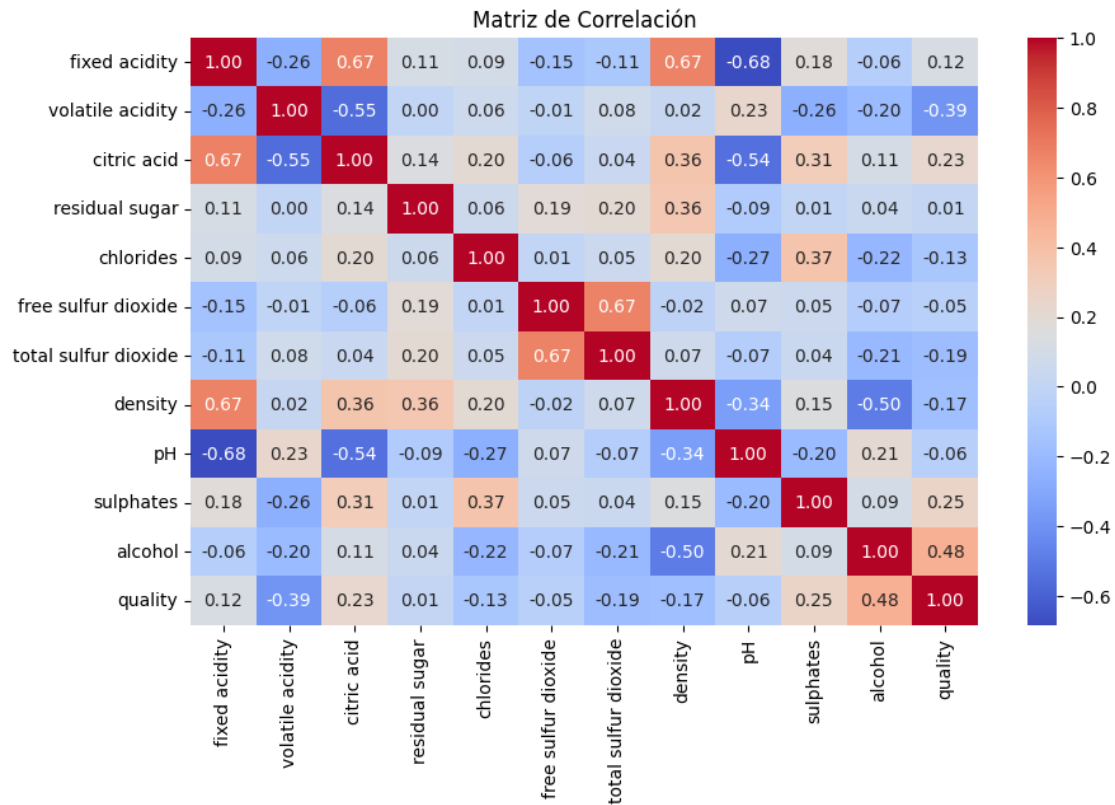
mean	8.319637	0.527821	0.270976	2.538806
std	1.741096	0.179060	0.194801	1.409928
min	4.600000	0.120000	0.000000	0.900000
25%	7.100000	0.390000	0.090000	1.900000
50%	7.900000	0.520000	0.260000	2.200000
75%	9.200000	0.640000	0.420000	2.600000
max	15.900000	1.580000	1.000000	15.500000

	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

## 5 Matriz de correlación

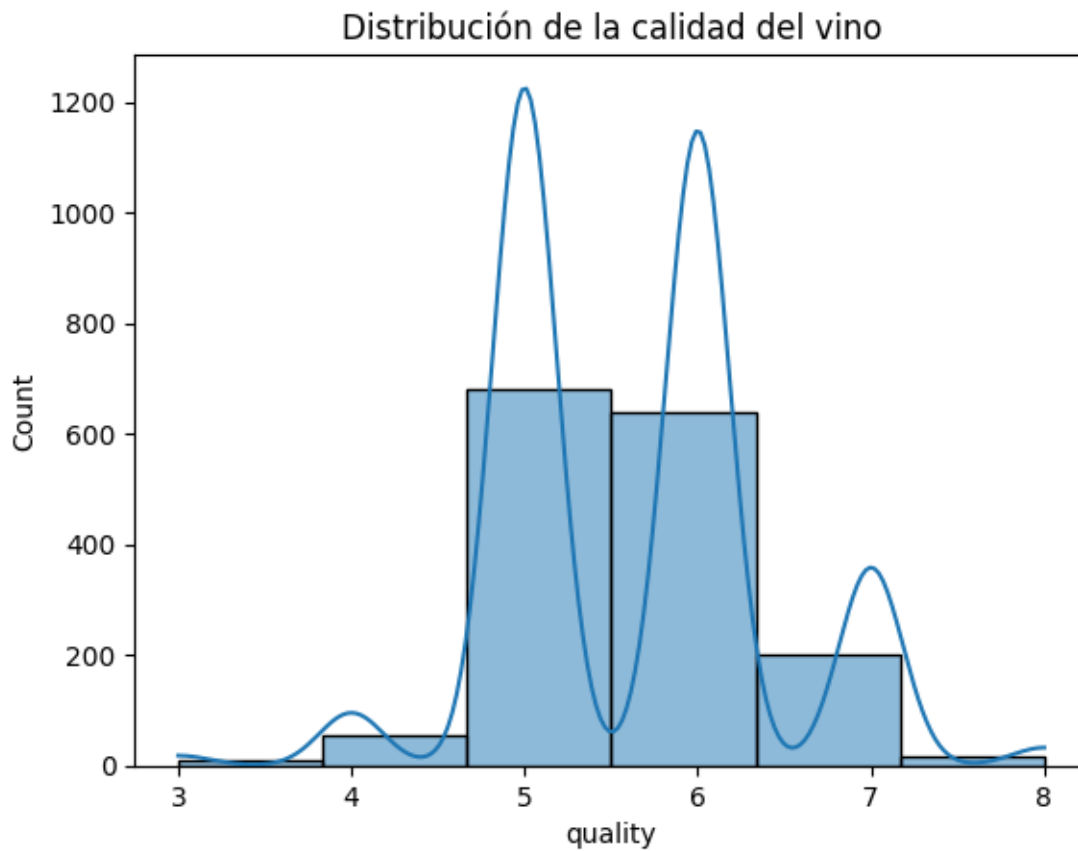
```
[14]: plt.figure(figsize=(10, 6))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Matriz de Correlación')
plt.show()
```



## 6 Gráficos

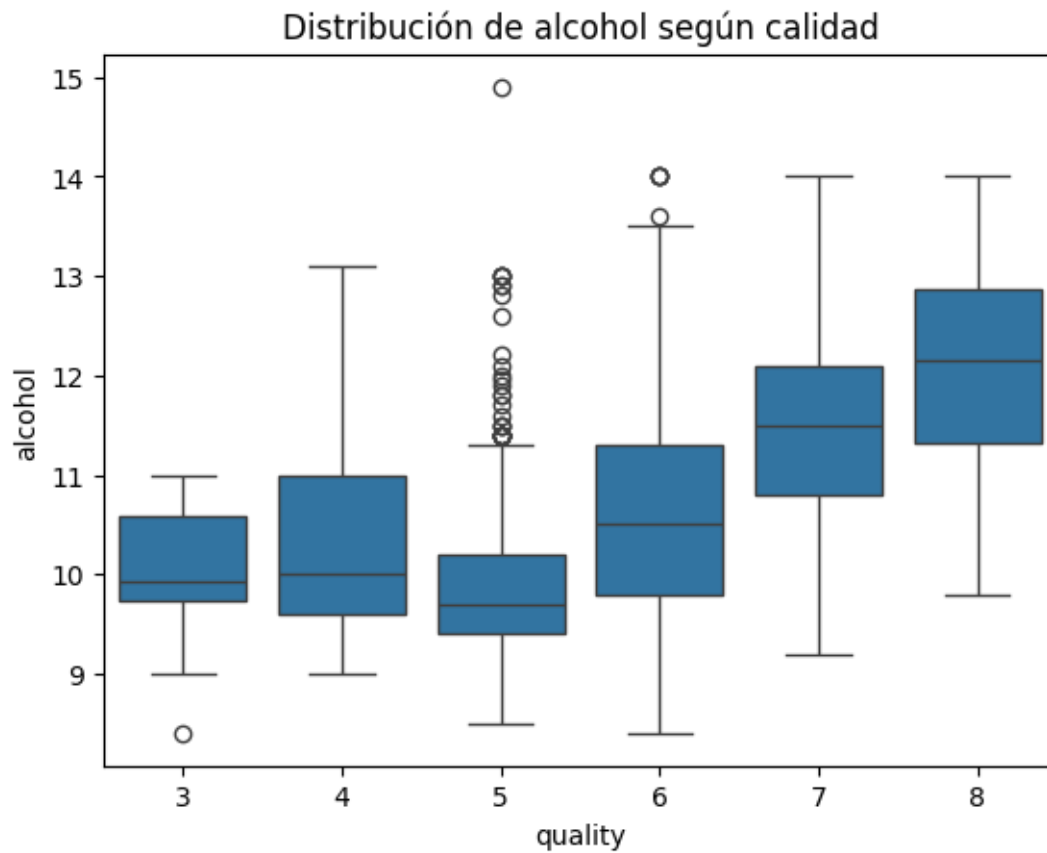
### 6.1 Histograma de la calidad del vino

```
[15]: sns.histplot(df['quality'], bins=6, kde=True)
plt.title('Distribución de la calidad del vino')
plt.show()
```



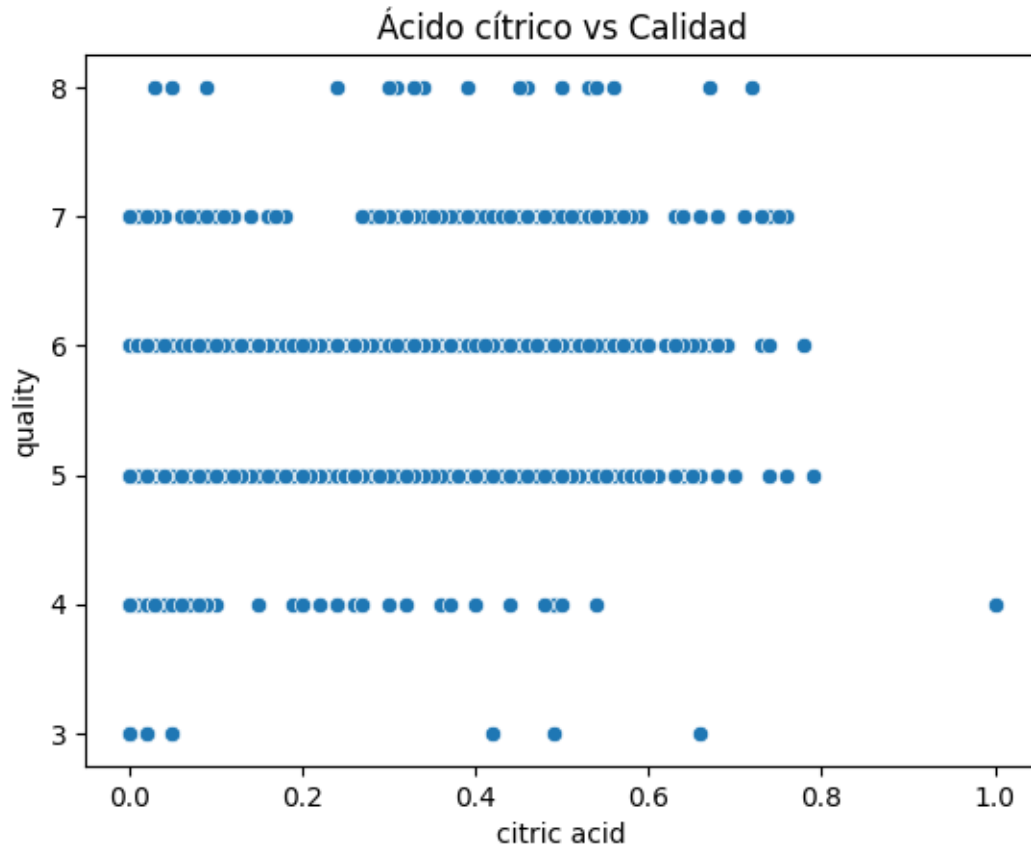
## 6.2 Boxplot de alcohol vs calidad

```
[16]: sns.boxplot(x=df['quality'], y=df['alcohol'])  
plt.title('Distribución de alcohol según calidad')  
plt.show()
```



### 6.3 Relación entre ácido cítrico y calidad

```
[17]: sns.scatterplot(x=df['citric acid'], y=df['quality'])  
plt.title('Ácido cítrico vs Calidad')  
plt.show()
```



## 7 Conclusiones

```
[18]: conclusiones = """
1. La mayoría de los vinos tienen una calidad entre 5 y 6.
2. Existe una relación positiva entre el contenido de alcohol y la calidad del
    ↪vino.
3. El ácido cítrico parece no tener una fuerte correlación con la calidad del
    ↪vino.
"""
print(conclusiones)
```

1. La mayoría de los vinos tienen una calidad entre 5 y 6.
2. Existe una relación positiva entre el contenido de alcohol y la calidad del vino.
3. El ácido cítrico parece no tener una fuerte correlación con la calidad del vino.