

▼ Actividad Evaluable

Obtención de estadísticas descriptivas en Pandas

Equipo 6

- A01245418 Andrés Sarellano Acevedo
- A00829837 Axel Amós Hernández Cárdenas
- A01281371 Izel María Ávila Rodríguez
- A01383422 Meliss Elvia Salazar Carrillo
- A00573134 Macías Romero Jorge Humberto

Instrucciones de uso:

- I. Cargar el archivo wine-clustering.csv.
- II. Presionar Ctrl + F10 para correr todo

```
#from google.colab import drive  
#drive.mount('/content/drive')
```

Por si se requiere, se incluye un código de copia de archivo para drive.

▼ Carga de Datos

```
# Leer los datos  
import pandas as pd  
  
#df = dataframe  
#df = pd.read_excel('Futbol2021.xlsx') #pd para leer  
  
df = pd.read_csv('wine-clustering.csv')
```

Se importa la librería de pandas, con el fin de poder traspasar los datos del .csv a un dataframe manipulable por Python3.

▼ DataFrame Creado

df

	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	Flavanoids	Non-
0	14.23	1.71	2.43	15.6	127	2.80	3.06	
1	13.20	1.78	2.14	11.2	100	2.65	2.76	
2	13.16	2.36	2.67	18.6	101	2.80	3.24	
3	14.37	1.95	2.50	16.8	113	3.85	3.49	
4	13.24	2.59	2.87	21.0	118	2.80	2.69	
...
173	13.71	5.65	2.45	20.5	95	1.68	0.61	
174	13.40	3.91	2.48	23.0	102	1.80	0.75	
175	13.27	4.28	2.26	20.0	120	1.59	0.69	
176	13.17	2.59	2.37	20.0	120	1.65	0.68	
177	14.13	4.10	2.74	24.5	96	2.05	0.76	

178 rows × 13 columns



▼ Verificando cantidad de filas x columnas

df.shape

(178, 13)

Utilizando la función *shape*, se encuentra que el dataframe creado a partir del archivo "wine-clustering.csv" cuenta con un total de 178 filas y 13 columnas.

▼ Tipos de datos en cada columna

Verificación de la cantidad de datos en las variables (columnas) del dataframe.

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
```

```
Data columns (total 13 columns):
#      Column              Non-Null Count  Dtype
---  -
0      Alcohol              178 non-null  float64
1      Malic_Acid              178 non-null  float64
2      Ash                     178 non-null  float64
3      Ash_Alcanity            178 non-null  float64
4      Magnesium               178 non-null  int64
5      Total_Phenols           178 non-null  float64
6      Flavanoids              178 non-null  float64
7      Nonflavanoid_Phenols    178 non-null  float64
8      Proanthocyanins         178 non-null  float64
9      Color_Intensity         178 non-null  float64
10     Hue                     178 non-null  float64
11     OD280                   178 non-null  float64
12     Proline                 178 non-null  int64
dtypes: float64(11), int64(2)
memory usage: 18.2 KB
```

Utilizando la función de *info*, se encuentra que todas las columnas cuentan con valores, es decir, ninguna columna cuenta con un valor nulo, por lo que se puede confiar en la distribución de los datos.

Comprobación del tipo de dato de cada variable (columna)

```
df.dtypes
```

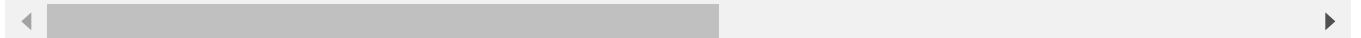
```
Alcohol              float64
Malic_Acid           float64
Ash                  float64
Ash_Alcanity         float64
Magnesium            int64
Total_Phenols        float64
Flavanoids           float64
Nonflavanoid_Phenols float64
Proanthocyanins      float64
Color_Intensity      float64
Hue                  float64
OD280                float64
Proline              int64
dtype: object
```

Se puede observar una mezcla entre datos tipo flotantes (decimales) y enteros (int) en el contenido del dataframe. En su mayoría, los datos son de tipo flotante.

▼ Estadística descriptiva del dataframe

```
df.describe().round(0)
```

	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	Flavanoids	I
count	178.0	178.0	178.0	178.0	178.0	178.0	178.0	
mean	13.0	2.0	2.0	19.0	100.0	2.0	2.0	
std	1.0	1.0	0.0	3.0	14.0	1.0	1.0	
min	11.0	1.0	1.0	11.0	70.0	1.0	0.0	
25%	12.0	2.0	2.0	17.0	88.0	2.0	1.0	
50%	13.0	2.0	2.0	20.0	98.0	2.0	2.0	
75%	14.0	3.0	3.0	22.0	107.0	3.0	3.0	
max	15.0	6.0	3.0	30.0	162.0	4.0	5.0	



Alcohol: Es el resultado de la fermentación de alguna fruta con levadura, lo cual resulta en la descomposición del azúcar en dióxido de carbono y etanol. (Italianowine, 2018)

- Mean: 13.0
- Min: 11.0
- Q1: 12.0
- Mediana: 13.0
- Q3: 14.0
- Max: 15.0

Malic_Acid: Uno de los principales ácidos que se encuentran en la acidez de las uvas. (Italianowine, 2018)

- Mean: 2.0
- Min: 1.0
- Q1: 2.0
- Mediana: 2.0
- Q3: 3.0
- Max: 3.0

Ash: Es uno de los indicadores más importantes para la determinación de calidad del vino (Dikanović-Lučan, Palić, Hanser, 1992). Se obtiene por medio de la incineración de residuos de evaporación del vino.

- Mean: 2.0

- Min: 1.0
- Q1: 2.0
- Mediana: 2.0
- Q3: 3.0
- Max: 3.0

Ash Alcanity (Alcanidad de Ceniza): Es una medida de basicidad (alcanidad) de la ceniza obtenida de una muestra. Esto se logra agregando ácido a la ceniza hasta que el nivel de pH de la muestra es neutro.

- Mean: 19.0
- Min: 11.0
- Q1: 17.0
- Mediana: 20.0
- Q3: 22.0
- Max: 30.0

Magnesium: Nutriente que el cuerpo necesita para estar sano, regulando la función muscular y nerviosa, así como los niveles de azúcar en la sangre y la presión arterial; Estando presente en el vino por la absorción de nutrientes del suelo en el viñedo.

- Mean: 100.0
- Min: 70.0
- Q1: 88.0
- Mediana: 98.0
- Q3: 107.0
- Max: 162.0

Total_Phenols: La firma de identidad del vino porque está relacionada con la experiencia sensorial del vino gracias a su composición fenólica y los polifenoles. Afecta características como la densidad de color y tono, nivel de taninos, etc. (A Guide to Wine Phenolics, 2020)

- Mean: 2.0
- Min: 1.0
- Q1: 2.0
- Mediana: 2.0
- Q3: 3.0
- Max: 4.0

Flavonoids: Dan el color rojo al vino, además de ser recomendable que tengan una baja concentración. (The Key Chemicals in Red Wine – Colour, Flavour, and Potential Health Benefits, 2014)

- Mean: 2.0
- Min: 0.0
- Q1: 1.0
- Mediana: 2.0
- Q3: 3.0
- Max: 5.0

Nonflavonoid_Phenols: Compuesto que influye en el olor y oxidación del vino / alcohol. (A Guide to Wine Phenolics, 2020)

- Mean: 0.0
- Min: 0.0
- Q1: 0.0
- Mediana: 0.0
- Q3: 0.0
- Max: 1.0

Proanthocyanins: Responsable de unir las proteínas salivales, influyendo en la astringencia del vino. (Noah Andrew White, 2019)

- Mean: 2.0
- Min: 0.0
- Q1: 1.0
- Mediana: 2.0
- Q3: 2.0
- Max: 4.0

Color_Intensity: Se refiere a la concentración de color en el vino / alcohol.

- Mean: 5.0
- Min: 1.0
- Q1: 3.0
- Mediana: 5.0
- Q3: 6.0
- Max: 13.0

HUE: Apoya a indicar la edad del vino. (Italianowine, 2018)

- Mean: 1.0
- Min: 1.0
- Q1: 1.0
- Mediana: 1.0
- Q3: 1.0

- Max: 2.0

OD280: Método para determinar la concentración de proteínas. Entre más alto OD280, más alta pureza proteica. (Bai et al, S.F)

- Mean: 3.0
- Min: 0.0
- Q1: 1.0
- Mediana: 1.0
- Q3: 1.0
- Max: 2.0

Proline: Es el aminoácido más abundante presente en el jugo de uva y el vino. En el cuerpo ayuda a producir proteínas como el colágeno y también se encuentra en la función general de las células. (Long et al., 2012)

- Mean: 747.0
- Min: 278.0
- Q1: 500.0
- Mediana: 747.0
- Q3: 985.0
- Max: 1680.0

▼ Preguntas

Macías Romero JH

¿Cuál es la mayor cantidad de Magnesio en este estudio?

```
df.Magnesium.max()
```

```
162162
```

Axel Amós Hernández Cárdenas

¿Qué alcohol o grupo de alcoholes cuentan con la mayor intensidad de color?

```
df[df['Color_Intensity'] == df['Color_Intensity'].max()]
```

Alcohol

Malic_Acid

Ash

Ash_Alcanity

Magnesium

Total_Phenols

Flavanoids

Nonf

Melissa Elvia Salazar Carrillo

¿Cuál es el promedio de alcohol en este estudio?

```
df.Alcohol.mean()


13.00061797752808313.000617977528083
```

Andrés Sarellano

¿Qué muestras tienen un nivel de alcalinidad de ceniza mayor a 25 ?

```
df[df['Ash_Alcanity'] > 25]
```

	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	Flavanoids	Nonf
73	12.99	1.67	2.60	30.0	139	3.30	2.89	
87	11.65	1.67	2.62	26.0	88	1.92	1.61	
121	11.56	2.05	3.23	28.5	119	3.18	5.08	
122	12.42	4.43	2.73	26.5	102	2.20	2.13	
127	11.79	2.13	2.78	28.5	92	2.13	2.24	
152	13.11	1.90	2.75	25.5	116	2.20	1.28	
157	12.45	3.03	2.64	27.0	97	1.90	0.58	



Izel Ávila

¿En realidad el total phenol y los flavonoids son el principal influyente en el color del vino?

Entre el total phenol y los flavonoids a pesar de ser de tener la misma estrcutura y complementarse, estos se suponen que son los principales componentes al momento de influir en la intensidad de color del vino. Pero no están en sus valores máximo para si quiera tener una alta intensidad. Lo cual me hace creer que hay otros factores.

```
df.nlargest(3, 'Color_Intensity')
```


	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	Flavanoids	Non
158	14.34	1.68	2.70	25.0	98	2.80	1.31	
159	13.48	1.67	2.64	22.5	89	2.60	1.10	
151	12.79	2.67	2.48	22.0	112	1.48	1.36	



#Con el fin de comparar, podemos ver los primeros valores para mayor percepción
df [['Color_Intensity', 'Total_Phenols', 'Flavanoids']]

	Color_Intensity	Total_Phenols	Flavanoids
0	5.64	2.80	3.06
1	4.38	2.65	2.76
2	5.68	2.80	3.24
3	7.80	3.85	3.49
4	4.32	2.80	2.69
...
173	7.70	1.68	0.61
174	7.30	1.80	0.75
175	10.20	1.59	0.69
176	9.30	1.65	0.68
177	9.20	2.05	0.76

178 rows × 3 columns

▼ Conclusiones

¿Qué conclusiones puedes tener de los datos?

Los datos del frame muestran muy poca dispersión entre ellos. Por ejemplo, la desviación estandar de casi todas las variables es pequeña a excepción de aquellas variables que cuentan con valores mas grandes Así mismo, es importante destacar que la mayoría de los boxplot muestran valores similares para los minimos, máximos y cuartiles. Finalmente, otro aspecto importante obtenido de la realización de la estadística descriptiva fue que la media \bar{X} es idéntica a

la mediana (50%) en la mayoría de los casos.

¿Crees que esta base de datos tendría buena respuesta para clustering/agrupación?

Consideramos que esta base de datos si sería una buena respuesta para realizar clustering/agrupación, debido a la baja desviación estándar que se tiene en cada variable, para la mayoría de estas.

Referencias

A Guide to Wine Phenolics. (2020). Guildsomm.com.

https://www.guildsomm.com/public_content/features/articles/b/jennifer-angelosante/posts/phenolics

Bai, X., Wang, L., & Li, H. (n.d.). Identification of red wine categories based on physicochemical properties. <https://doi.org/10.25236/etmhs.2019.309>

ETS Labs. (S, f). Alcanity of Ash. <https://www.etslabs.com/analyses/ALK>

Italianowine. (2018). Wine composition - Italianowine. Italianowine.com.

<https://www.italianowine.com/en/technology/wine-composition/>

Long, D., Wilkinson, K. L., Poole, K., Taylor, D. K., Warren, T., Astorga, A. M., & Jiranek, V. (2012). Rapid Method for Proline Determination in Grape Juice and Wine. Journal of Agricultural and Food Chemistry, 60(17), 4259–4264. <https://doi.org/10.1021/jf300403b>

Noah Andrew White. (2019, March). Proanthocyanidins. Waterhouse Lab.

<https://waterhouse.ucdavis.edu/whats-in-wine/proanthocyanidins>

✓ 0 s completado a las 12:02

● ✕