

ACTIVIDAD 1

Dataset Elegida:	2
Página web:	2
Apartado 1:	3
4 columnas:	4
Agregar columna:	4
Filtrar filas por condición (price_euros > 50):	5
Columna nula:	5
Rellenar valores nulos:	5
Agrupar por tallas:	6
Precio promedio agrupado por tallas:	7
ANEXO (CÓDIGO)	8

Dataset Elegida:

Hemos elegido un dataset con información sobre los precios de los zapatos ya que está relacionado con un tema cotidiano y comprensible. Además, el dataset tiene una estructura clara facilitando la manipulación y comprensión de los datos. El dataset, con 5.680 filas, es suficientemente grande para realizar análisis significativos, pero no tan extenso como para dificultar el procesamiento.

Finalmente, este dataset permite explorar diversas posibilidades analíticas. Se pueden realizar estudios como la predicción del precio con descuento, análisis de popularidad por marca o análisis de factores que influyen en el precio.

Página web:

Kaggle Datasets

<https://www.kaggle.com/datasets/ashutosh598/shoes-price-for-various-brands?resource=download>

Apartado 1:

```
Esta es la información que nos proporciona el shape
(5679, 5)
5679
5

Esta es la información que nos proporciona el head
  brand color size price offer_price
0  Tresmode Black 6.5  4647      4153
1   Lavie   Gold  4   2446      1172
2    FILA   Red   6   2346      1898
3   Crocs   Red   5   5942      5868
4   Crocs   Blue  4   2942      2118

Esta es la información que nos proporciona el info
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5679 entries, 0 to 5678
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   brand           5679 non-null   object
1   color           5679 non-null   object
2   size           5679 non-null   object
3   price           5679 non-null   int64
4   offer_price     5679 non-null   int64
dtypes: int64(2), object(3)
memory usage: 222.0+ KB
None

Esta es la información que nos proporciona el describe
      price  offer_price
count  5679.000000  5679.000000
mean    4945.126255  4007.745026
std     5344.881308  4761.543884
min     1346.000000   773.000000
25%     2942.000000  1968.000000
50%     3742.000000  2868.000000
75%     4646.000000  4048.000000
max     37947.000000 34123.000000
```

Registros totales (count):

- Hay 5679 precios registrados

Media (mean):

- El valor medio de la columna price es **4945.13**.
- Esto indica que, en promedio, los precios se sitúan alrededor de **4945 unidades monetarias**.
- Sin embargo, debemos considerar también la dispersión de los datos para confirmar si este promedio representa adecuadamente la columna.

Desviación estándar (std):

- La desviación estándar es **5344.88**, un valor bastante elevado.
- Esto significa que los datos de price están **muy dispersos** respecto a la media.

- Un valor de desviación estándar tan alto sugiere que hay precios muy alejados del promedio, es decir, existen **precios extremadamente altos o bajos** que aumentan esta dispersión.

Valores mínimo (min) y máximo (max):

- El precio mínimo es **1346** y el máximo es **37947**.
- La diferencia entre estos dos valores es significativa (36,601 unidades).
- Esto confirma la gran dispersión y la presencia de **precios atípicos** o "outliers" en la columna.

Percentiles (25%, 50%, 75%):

- **25% (Percentil 25): 2942**
El 25% de los precios son menores o iguales a **2942**, lo cual sugiere que una cuarta parte de los productos tienen precios bajos.
- **50% (Mediana): 3742**
La mediana es **3742**, menor que la media (**4945.13**). Esto indica que la distribución de los precios está **sesgada hacia la derecha** (muchos valores bajos, pero algunos muy altos).
- **75% (Percentil 75): 4646**
El 75% de los precios son menores o iguales a **4646**, lo que muestra que la mayoría de los productos tienen un precio cercano a este valor.

4 columnas:

Columnas mas importantes:

	brand	size	price	offer_price
0	Tresmode	6.5	4647	4153
1	Lavie	4	2446	1172
2	FILA	6	2346	1898
3	Crocs	5	5942	5868
4	Crocs	4	2942	2118

Agregar columna:

Nueva columna:

	brand	color	size	price	offer_price	price_euros
0	Tresmode	Black	6.5	4647	4153	56.2287
1	Lavie	Gold	4	2446	1172	29.5966
2	FILA	Red	6	2346	1898	28.3866
3	Crocs	Red	5	5942	5868	71.8982
4	Crocs	Blue	4	2942	2118	35.5982

Filtrar filas por condición (price_euros > 50):

```
Filtrar filas por condicion
```

	brand	color	size	price	offer_price	price_euros
0	Tresmode	Black	6.5	4647	4153	56.2287
3	Crocs	Red	5	5942	5868	71.8982
9	Crocs	Brown	11	5742	5668	69.4782
11	LaBriza	Brown	5	4746	2522	57.4266
13	Franco Leone	Bordo	7	4242	1893	51.3282

Columna nula:

En el csv hemos añadido una fila con estos datos seguidos de comas:

```
Limpiar la tabla de nulos
```

	brand	color	size	price	offer_price	price_euros
0	Crocs	Black	6	NaN	NaN	NaN
1	Tresmode	Black	6.5	4647.0	4153.0	56.2287
2	Lavie	Gold	4	2446.0	1172.0	29.5966
3	FILA	Red	6	2346.0	1898.0	28.3866
4	Crocs	Red	5	5942.0	5868.0	71.8982

Filas antes: 5680
Filas después: 5679

Rellenar valores nulos:

```
Rellenar valores nulos
```

Antes

	brand	color	size	price	offer_price	price_euros
0	Crocs	Black	6	NaN	NaN	NaN
1	Tresmode	Black	6.5	4647.0	4153.0	56.2287
2	Lavie	Gold	4	2446.0	1172.0	29.5966
3	FILA	Red	6	2346.0	1898.0	28.3866
4	Crocs	Red	5	5942.0	5868.0	71.8982
5	Crocs	Blue	4	2942.0	2118.0	35.5982

Despues

	brand	color	size	price	offer_price	price_euros
0	Crocs	Black	6	1935287.0	1074516.0	23416.9727
1	Tresmode	Black	6.5	4647.0	4153.0	56.2287
2	Lavie	Gold	4	2446.0	1172.0	29.5966
3	FILA	Red	6	2346.0	1898.0	28.3866
4	Crocs	Red	5	5942.0	5868.0	71.8982
5	Crocs	Blue	4	2942.0	2118.0	35.5982

Agrupar por tallas:

Agrupar por tallas y sacar promedio:

size

1 2

10 514

10.5 27

11 410

11.5 3

12 130

12.5 4

13 15

14 16

15 2

16 2

2 107

2.5 43

3 295

3.5 89

37 4

38 4

39 4

4 387

4.5 1

40 3

41 2

5 428

5.5 22

6 774

6.5 107

7 794

7.5 19

8 752

8.5 30

9 629

9.5 30

UK-06 7

UK-07 4

UK-08 7

UK-09 5

UK-10 5

UK-11 3

Precio promedio agrupado por tallas:

```
Promedio de precio por tallas:  
price_euros
```

```
size
```

```
1      41.648200  
10     58.075833  
10.5   281.108096  
11     61.876773  
11.5   99.788700  
12     53.329075  
12.5   24.345200  
13     46.575320  
14     50.647575  
15     29.548200  
16     32.573200  
2      44.158667  
2.5    56.228700  
3      43.266770  
3.5    55.990643  
37     48.272950  
38     51.116450  
39     50.148450  
4      47.714396  
4.5    344.208700  
40     51.126533  
41     38.381200  
5      58.535418  
5.5    377.786200  
6      53.560900  
6.5    83.195981  
7      53.859493  
7.5    294.006437  
8      53.798048  
8.5    248.132280  
9      55.758935  
9.5    220.926237  
UK-06  59.328029  
UK-07  58.334100  
UK-08  58.809457  
UK-09  56.942600  
UK-10  58.636600  
UK-11  59.039933
```

ANEXO (CÓDIGO)

```
vidad1.py > ...
import pandas as pd

df = pd.read_csv("preciosZapatos.csv")

print(df.shape)
print(df.shape[0])
print(df.shape[1])

print("\nHead")
print(df.head())

print("\nInfo")
print(df.info())

print("\nDescribe")
print(df.describe())

print("\nColumnas mas importantes: ")
zapatos = df[["brand", "size", "price", "offer_price"]]
print(zapatos.head())

print("\nAgregar columna:")
df["price_euros"] = df["price"] * 0.0121
print(df.head())

print("\nFiltrar filas por condicion")
mas_de_50 = df[df["price_euros"] > 50]
print(mas_de_50.head())

print("\nEliminar fila con valores nulos")
df_limpio = df.dropna()
print(f"Filas antes: {df.shape[0]}")
print(f"Filas después: {df_limpio.shape[0]}")

print("\nRellenar valores nulos")
print("\nAntes")
print(df.head(6))
df["price"] = df["price"].fillna(1935287)
df["offer_price"] = df["offer_price"].fillna(1074516)
df["price_euros"] = df["price_euros"].fillna(df["price"]*0.0121)
print("\nDespués")
print(df.head(6))

print("\nAgrupar por tallas: ")
conteo_por_tallas = df.groupby("size").size()
print(conteo_por_tallas)
promedio_precio_euros = df.groupby("size")[["price_euros"]].mean()
print("\nPromedio de precio en euros por talla: ")
print(promedio_precio_euros)
```