

Análisis de los 25 retailers más grandes de Estados Unidos

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [ ]: df_path = '/work/dataset_datacademy.csv'
df = pd.read_csv(df_path)
```

```
In [ ]: df.head()
```

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category
0	Walmart US	658119	45740	65.649725	0.012843	Supercenters
1	Kroger	115037	39310	30.031850	-0.053887	Grocery
2	Costco	90048	4950	187.795620	0.066810	Warehouse Club
3	Home Depot	83976	19650	42.735878	0.000000	Home Improvement
4	Walgreen Boots	78924	80020	9.810927	-0.010511	Drug Stores

```
In [ ]: df.describe()
```

	Sales	Stores	Sales/Avg. Store	Store Count Growth
count	25.00000	23.00000	23.00000	23.00000
mean	71063.080000	3468.565217	36.870966	0.040230
std	125405.72644	4038.708714	40.625277	-0.068552
min	16592.000000	231.000000	1.563919	-0.059515
25%	24402.000000	1029.000000	11.263462	-0.001860
50%	34980.000000	1828.000000	25.997733	0.019520
75%	71687.000000	4242.000000	41.359595	0.078934
max	658119.000000	14250.000000	187.795620	0.255662

A primera vista, la mediana de ventas (71063.08) está muy por debajo de la mediana o Q2 (34980) lo que nos habla de una muestra sesgada, donde algunos de sus integrantes se despegan claramente del resto en cuanto a ventas. Mas adelante veremos de quien se trata y cuanto se despegan.

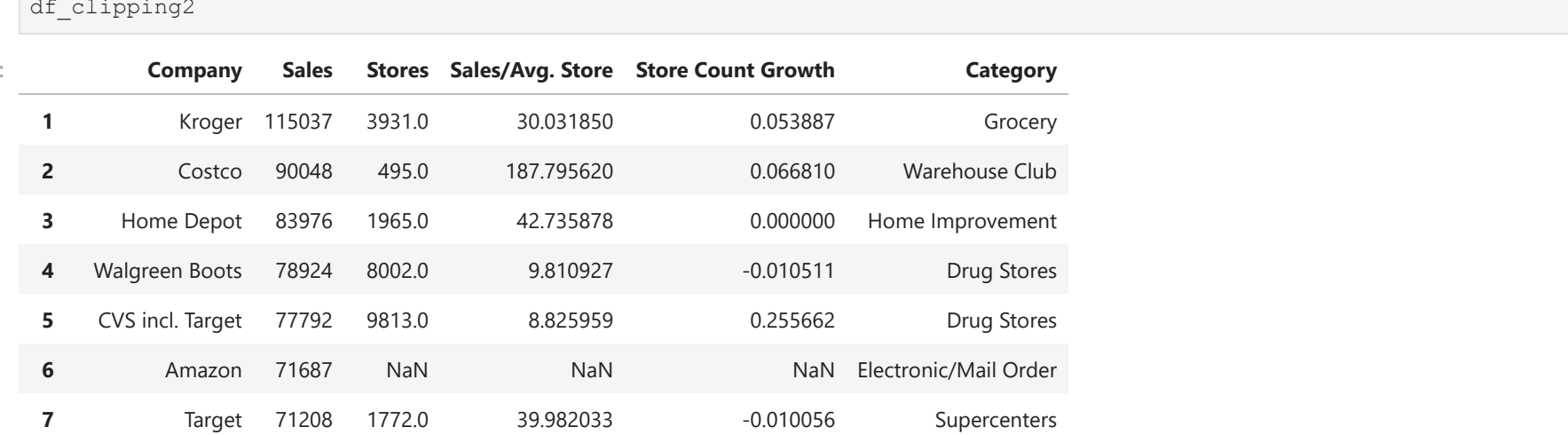
```
In [ ]: df_sorted = df.sort_values('Sales', ascending=False)
df_clipping = df_sorted
```

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category
0	Walmart US	658119	45740	65.649725	0.012843	Supercenters
1	Kroger	115037	39310	30.031850	-0.053887	Grocery
2	Costco	90048	4950	187.795620	0.066810	Warehouse Club
3	Home Depot	83976	19650	42.735878	0.000000	Home Improvement
4	Walgreen Boots	78924	80020	9.810927	-0.010511	Drug Stores
5	CVS incl. Target	77792	98130	8.825959	0.255662	Drug Stores
6	Amazon	71687	NaN	NaN	NaN	Electronic/Mail Order
7	Target	71208	17720	39.982033	-0.010056	Supercenters
8	Lowe's	60311	18280	33.311792	0.019520	Home Improvement
9	Albertsons	56829	23260	23.877731	-0.043771	Grocery
10	Sam's Club	56828	6550	87.293395	0.012365	Warehouse Clubs
11	Apple incl. Online	37664	NaN	NaN	NaN	NaN
12	Best Buy	34980	13890	24.685956	-0.038754	Electronics
13	Publix	34408	13510	25.997733	0.042438	Grocery
14	Rite Aid	27486	45530	6.025649	-0.003720	Drug Stores
15	Ahold	26903	7940	34.446863	0.033854	Grocery
16	Macy's	26028	8890	30.442105	0.082826	Department Stores
17	TIx	25012	27700	9.369545	0.078241	Apparel
18	Aldi	24402	20210	12.715998	0.112273	Grocery
19	Dollar General	22234	133500	1.768885	0.132412	Dollar Stores
20	Dollar Tree	21464	142500	1.563919	0.079627	Dollar Stores
21	HEB	21384	3230	66.825000	0.018927	Grocery
22	Kohl's	19060	11690	16.353496	0.006024	Department Stores
23	Delhaize	18201	12800	13.783415	-0.059515	Grocery
24	Meijer	16592	2310	74.738739	0.084507	Supercenters

```
In [ ]: df_clipping = df_sorted.iloc[: 5, : ]
df_clipping
```

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category
0	Walmart US	658119	45740	65.649725	0.012843	Supercenters
1	Kroger	115037	39310	30.031850	-0.053887	Grocery
2	Costco	90048	4950	187.795620	0.066810	Warehouse Club
3	Home Depot	83976	19650	42.735878	0.000000	Home Improvement
4	Walgreen Boots	78924	80020	9.810927	-0.010511	Drug Stores

```
In [ ]: chart = sns.barplot(data = df_clipping, x = 'Company', y = 'Sales')
chart.set_title('Sales of the main 5 retailers in the USA')
chart.set_xlabel('Top 5 retailers in sales')
chart.set_ylabel('Sales in millions of dollars')
sns.set(rc = {'figure.figsize':(18,8)})
```



Efectivamente, podemos observar que Walmart US le saca una distancia de casi 6X al segundo (Kroger) en la lista de ventas. Lo cual confirma la primera impresión que teníamos al ver las medidas de tendencia central.

I. Preguntas del negocio

1. ¿Cuál es el promedio de ventas sin contar a la compañía dominante?

```
In [ ]: df_clipping2 = df_sorted.iloc[ 1 : , : ]
df_clipping2
```

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category
1	Kroger	115037	39310	30.031850	-0.053887	Grocery
2	Costco	90048	4950	187.795620	0.066810	Warehouse Club
3	Home Depot	83976	19650	42.735878	0.000000	Home Improvement
4	Walgreen Boots	78924	80020	9.810927	-0.010511	Drug Stores
5	CVS incl. Target	77792	98130	8.825959	0.255662	Drug Stores
6	Amazon	71687	NaN	NaN	NaN	Electronic/Mail Order
7	Target	71208	17720	39.982033	-0.010056	Supercenters
8	Lowe's	60311	18280	33.311792	0.019520	Home Improvement
9	Albertsons	56829	23260	23.877731	-0.043771	Grocery
10	Sam's Club	56828	6550	87.293395	0.012365	Warehouse Clubs
11	Apple incl. Online	37664	NaN	NaN	NaN	NaN
12	Best Buy	34980	13890	24.685956	-0.038754	Electronics
13	Publix	34408	13510	25.997733	0.042438	Grocery
14	Rite Aid	27486	45530	6.025649	-0.003720	Drug Stores
15	Ahold	26903	7940	34.446863	0.033854	Grocery
16	Macy's	26028	8890	30.442105	0.082826	Department Stores
17	TIx	25012	27700	9.369545	0.078241	Apparel
18	Aldi	24402	20210	12.715998	0.112273	Grocery
19	Dollar General	22234	133500	1.768885	0.132412	Dollar Stores
20	Dollar Tree	21464	142500	1.563919	0.079627	Dollar Stores
21	HEB	21384	3230	66.825000	0.018927	Grocery
22	Kohl's	19060	11690	16.353496	0.006024	Department Stores
23	Delhaize	18201	12800	13.783415	-0.059515	Grocery
24	Meijer	16592	2310	74.738739	0.084507	Supercenters

```
In [ ]: df_clipping2['Sales'].mean()
```

```
46602.416666666664
```

```
In [ ]: plt.bar(df_clipping2['Company'], height=df_clipping2['Sales'], width = 0.9, color = ["green","blue","grey", , ]
plt.grid()
plt.xlabel(df_clipping2['Sales'].mean(), color = "r", linestyle = '--', linewidth = 5)
plt.xticks(rotation='vertical')
plt.title('Sales of the main companies excluding Walmart with average sales')
plt.ylabel('Main retailers excluding the dominant one')
plt.ylabel('Sales in millions of dollars')
plt.text(x = 'Rite Aid',y = df_clipping2['Sales'].mean()+3000, s= 'Average sales', fontfamily = 'serif', size = 12, color = 'r')
plt.show()
```



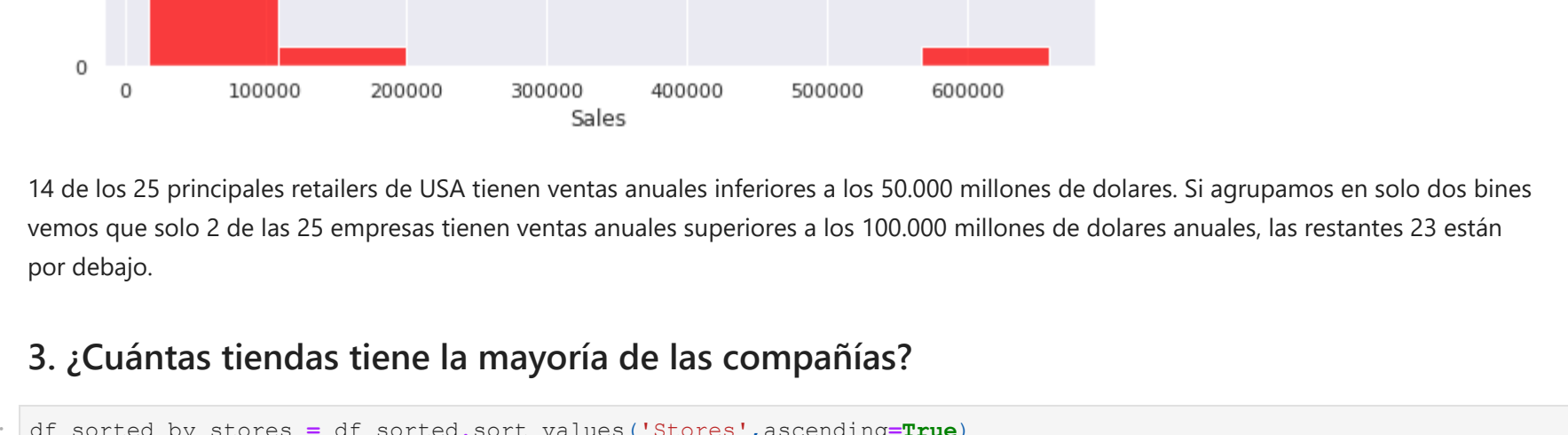
Como podemos observar en el gráfico, el promedio de ventas excluyendo a la compañía dominante (Walmart) es de 46602.42.

2. ¿Cuánto dinero en ventas generó la mayoría de las compañías?

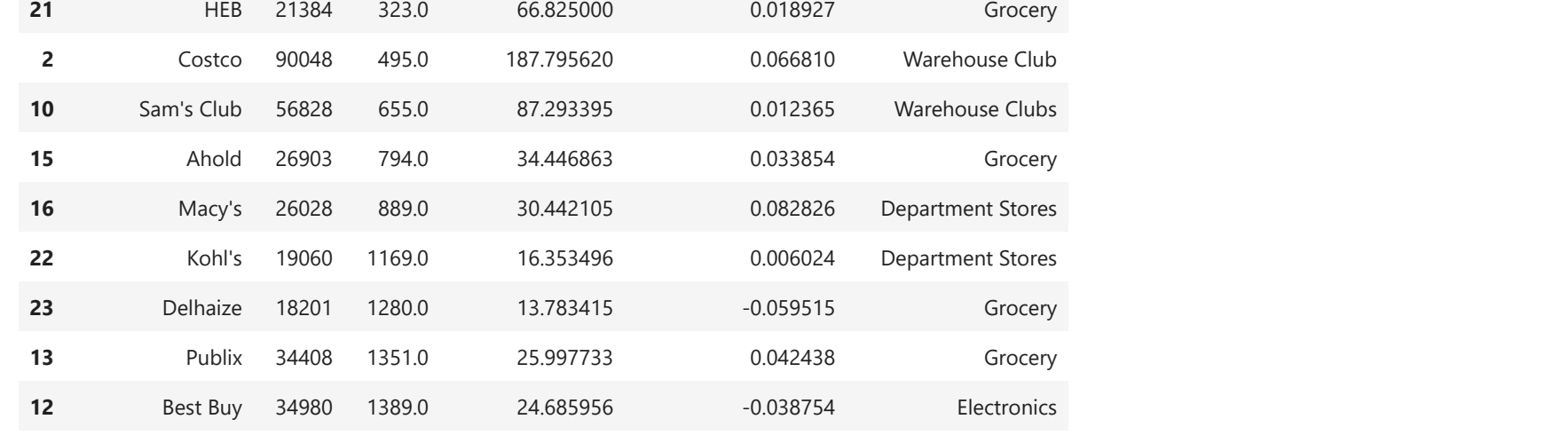
```
In [ ]: df_sorted_by_sales = df_sorted.sort_values('Sales',ascending=True)
df_sorted_by_sales
```

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category
24	Meijer	16592	2310	74.738739	0.084507	Supercenters
23	Delhaize	18201	12800	13.783415	-0.059515	Grocery
22	Kohl's	19060	11690	16.353496	0.006024	Department Stores
21	HEB	21384	3230	66.825000	0.018927	Grocery
20	Dollar Tree	21464	142500	1.563919	0.079627	Dollar Stores
19	Dollar General	22234	133500	1.768885	0.132412	Dollar Stores
18	Aldi	24402	20210	12.715998	0.112273	Grocery
17	TIx	25012	27700	9.369545	0.078241	Apparel
16	Macy's	26028	8890	30.442105	0.082826	Department Stores
15	Ahold	26903	7940	34.446863	0.033854	Grocery
14	Rite Aid	27486	45530	6.025649	-0.003720	Drug Stores
13	Publix	34408	13510	25.997733	0.042438	Grocery
12	Best Buy	34980	13890	24.685956	-0.038754	Electronics
11	Apple incl. Online	37664	NaN	NaN	NaN	NaN
10	Sam's Club	56828	6550	87.293395	0.012365	Warehouse Clubs
9	Albertsons	56829	23260	23.877731	-0.043771	Grocery
8	Lowe's	60311	18280	33.311792	0.019520	Home Improvement
7	Target	71208	17720	39.982033	-0.010056	Supercenters
6	Amazon	71687	NaN	NaN	NaN	Electronic/Mail Order
5	CVS incl. Target	77792	98130	8.825959	0.255662	Drug Stores
4	Walgreen Boots	78924	80020	9.810927	-0.010511	Drug Stores
3	Home Depot	83976	19650	42.735878	0.000000	Home Improvement
2	Costco	90048	4950	187.795620	0.066810	Warehouse Club
1	Kroger	115037	39310	30.031850	-0.053887	Grocery
0	Walmart US	658119	45740	65.649725	0.012843	Supercenters

```
In [ ]: sns.histplot(data = df_sorted_by_sales, x = "Sales", multiple="layer", color = "green")
sns.set(rc = {'figure.figsize':(15,8)})
```



```
In [ ]: sns.histplot(data = df_sorted_by_sales, x = "Sales", bins = 7, multiple = "layer", color = "red")
sns.set(rc = {'figure.figsize':(15,8)})
```



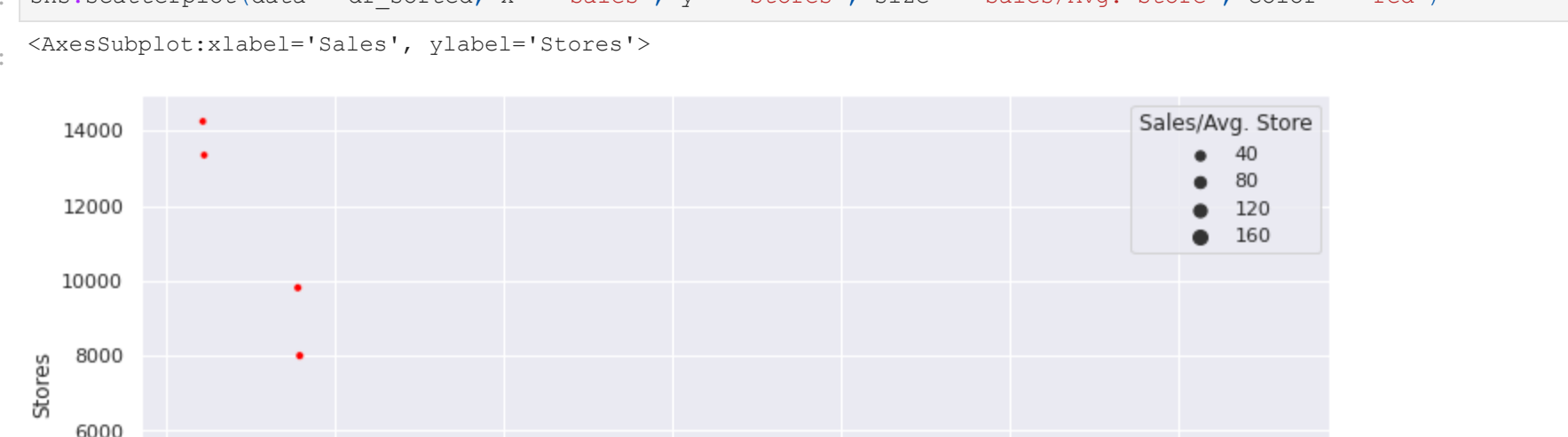
14 de los 25 principales retailers de USA tienen ventas anuales inferiores a los 50.000 millones de dólares. Si agrupamos en solo dos bins vemos que solo 2 de las 25 empresas tienen ventas anuales superiores a los 100.000 millones de dólares anuales, las restantes 23 están por debajo.

3. ¿Cuántas tiendas tiene la mayoría de las compañías?

```
In [ ]: df_sorted_by_stores = df_sorted.sort_values('Stores',ascending=True)
df_sorted_by_stores
```

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category
24	Meijer	16592	2310	74.738739	0.084507	Supercenters
23	HEB	21384	3230	66.825000	0.018927	Grocery
22	Kohl's	19060	11690	16.353496	0.006024	Department Stores
21	HEB	21384	3230	66.825000	0.018927	Grocery
20	Dollar Tree	21464	142500	1.563919	0.079627	Dollar Stores
19	Dollar General	22234	133500	1.768885	0.132412	Dollar Stores
18	Aldi	24402	20210	12.715998	0.112273	Grocery
17	TIx	25012	27700	9.369545	0.078241	Apparel
16	Macy's	26028	8890	30.442105	0.082826	Department Stores
15	Ahold	26903	7940	34.446863	0.033854	Grocery
14	Rite Aid	27486	45530	6.025649	-0.003720	Drug Stores
13	Publix	34408	13510	25.997733	0.042438	Grocery
12	Best Buy	34980	13890	24.685956	-0.038754	Electronics
11	Apple incl. Online	37664	NaN	NaN	NaN	NaN
10	Sam's Club	56828	6550	87.293395	0.012365	Warehouse Clubs
9	Albertsons	56829	23260	23.877731	-0.043771	Grocery
8	Lowe's	60311	18280	33.311792	0.019520	Home Improvement
7	Target	71208	17720	39.982033	-0.010056	Supercenters
6	Amazon	71687	NaN	NaN	NaN	Electronic/Mail Order
5	CVS incl. Target	77792	98130	8.825959	0.255662	Drug Stores
4	Walgreen Boots	78924	80020	9.810927	-0.010511	Drug Stores
3	Home Depot	83976	19650	42.735878	0.000000	Home Improvement
2	Costco	90048	4950	187.795620	0.066810	Warehouse Club
1	Kroger	115037	39310	30.031850	-0.053887	Grocery
0	Walmart US	658119	45740	65.649725	0.012843	Supercenters

```
In [ ]: sns.histplot(data = df_sorted_by_stores, x="Stores", bins=8, multiple="layer", color = "green")
sns.set(rc = {'figure.figsize':(12,6)})
```



13 de los 25 principales retailers en USA (la mayoría), tienen menos de 2000 tiendas.

4.

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category	Sales/Avg. Store 2	Only online
2	Costco	90048	495.0	187.795620	0.066810	Warehouse Club	181.915152	No
0	Walmart US	658119	4574.0	65.69725	0.013843	Supercenters	143.882597	No
10	Sam's Club	56828	655.0	87.293595	0.012365	Warehouse Clubs	86.760305	No
24	Mejor	16592	231.0	74.73879	0.084507	Supercenters	71.826840	No
21	HEB	21384	323.0	66.825000	0.018927	Grocery	66.204334	No
3	Home Depot	83976	1965.0	42.735878	0.000000	Home Improvement	42.735878	No
7	Target	71208	1772.0	39.582033	-0.010056	Supercenters	40.181502	No
15	Ahold	26903	794.0	34.468663	0.033887	Grocery	33.882872	Yes
8	Lowe's	60311	1828.0	33.311792	0.019520	Home Improvement	32.992888	No
16	Macy's	26028	889.0	30.442105	0.082826	Department Stores	29.277840	No
1	Kroger	115037	3931.0	30.031850	0.053887	Grocery	29.264055	Yes
13	Publix	34408	1351.0	25.997733	0.042438	Grocery	25.468542	Yes
12	Best Buy	34980	1389.0	24.685956	-0.038754	Electronics	25.183585	No
9	Albertsons	56829	2326.0	23.877731	-0.043371	Grocery	24.432072	No
22	Kohl's	19060	1169.0	16.353496	0.006024	Department Stores	16.304534	No
23	Delhaize	18201	1280.0	13.783415	-0.059515	Grocery	14.219531	No
18	Aldi	24402	2021.0	12.715998	0.112273	Grocery	12.074221	No
4	Walgreen Boots	78924	8002.0	9.810927	-0.010511	Drug Stores	9.863034	No
17	TIJX	25012	2770.0	9.369545	0.078241	Apparel	9.029603	Yes
5	CVS incl Target	77792	9813.0	8.825959	0.255662	Drug Stores	7.927443	No
14	Rite Aid	27486	4553.0	6.025649	-0.003720	Drug Stores	6.036899	No
19	Dollar General	22234	13350.0	1.768885	0.132412	Dollar Stores	1.665468	No
20	Dollar Tree	21464	14250.0	1.563919	0.079627	Dollar Stores	1.506246	No
6	Amazon	71687	NaN	NaN	NaN	Electronic/Mail Order	NaN	Yes
11	Apple incl. Online	37664	NaN	NaN	NaN	NaN	NaN	Yes

```
In [ ]: df_filter = df_sorted_by_sales_stores[df_sorted_by_sales_stores['Only online'] == 'Yes']
df_filter = df_filter.sort_values('Sales', ascending=False)
df_filter
```

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category	Sales/Avg. Store 2	Only online
1	Kroger	115037	3931.0	30.031850	0.053887	Grocery	29.264055	Yes
6	Amazon	71687	NaN	NaN	NaN	Electronic/Mail Order	NaN	Yes
11	Apple incl. Online	37664	NaN	NaN	NaN	NaN	NaN	Yes
12	Best Buy	34980	1389.0	24.685956	-0.038754	Electronics	25.183585	Yes
15	Ahold	26903	794.0	34.446863	0.033854	Grocery	33.882872	Yes
17	TIJX	25012	2770.0	9.369545	0.078241	Apparel	9.029603	Yes

```
In [ ]: df_filter_slice = df_filter.iloc[1, : 2]
df_filter_slice = df_filter_slice.set_index('Company')
chart4 = sns.heatmap(data = df_filter_slice, cmap='BuBu', annot=True, fmt='d')
chart4.set_title("Sales in millions of dollars from online retailers")
chart4.set_ylabel("Companies that sell exclusively online")
chart4.set_xlabel("Sales in millions of dollars")
chart4
```

```
Out [ ]: <AxesSubplot:title="Sales in millions of dollars from online retailers", xlabel="Sales in millions of dollars", ylabel="Companies that sell exclusively online">
```



Como podemos observar, las compañías de retail que venden de forma exclusivamente online son:

Kroger

Amazon

Apple incl. Online

Best Buy

Ahold

TIJX

```
In [ ]: import matplotlib as plt
plt.figure(figsize=(10, 10))
plt.bar(df_sorted_by_sales_stores['Company'], df_sorted_by_sales_stores['Sales'])
plt.xticks(df_sorted_by_sales_stores['Company'])
plt.ylabel('Sales in millions of dollars')
plt.show()
```

```
Out [ ]: <Figure: 100x100px>
```

Como podemos observar, no hay diferencias significativas entre aquellos retailers que tienen solo venden online y aquellos que tienen tiendas físicas en cuanto a suma de ventas. La única empresa que destaca sobre el resto es Walmart como ya lo mostramos mas arriba.

III. Tus propias preguntas

```
In [ ]: df_sorted_by_sales_stores
```

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category	Sales/Avg. Store 2	Only online
2	Costco	90048	495.0	187.80	0.07	Warehouse Club	181.92	No
0	Walmart US	658119	4574.0	65.65	0.01	Supercenters	143.88	No
10	Sam's Club	56828	655.0	87.29	0.01	Warehouse Clubs	86.76	No
24	Mejor	16592	231.0	74.74	0.08	Supercenters	71.83	No
21	HEB	21384	323.0	66.83	0.02	Grocery	66.20	No
3	Home Depot	83976	1,965.0	42.74	0.00	Home Improvement	42.74	No
7	Target	71208	1,772.0	39.98	-0.01	Supercenters	40.19	No
15	Ahold	26903	794.0	34.45	0.03	Grocery	33.88	Yes
8	Lowe's	60311	1,828.0	33.31	0.02	Home Improvement	32.99	No
16	Macy's	26028	889.0	30.44	0.08	Department Stores	29.28	No
1	Kroger	115037	3,931.0	30.03	0.05	Grocery	29.26	Yes
13	Publix	34408	1,351.0	26.00	0.04	Grocery	25.47	No
12	Best Buy	34980	1,389.0	24.69	-0.04	Electronics	25.18	Yes
9	Albertsons	56829	2,326.0	23.88	-0.04	Grocery	24.43	No
22	Kohl's	19060	1,169.0	16.35	0.01	Department Stores	16.30	No
23	Delhaize	18201	1,280.0	13.78	-0.06	Grocery	14.22	No
18	Aldi	24402	2,021.0	12.72	0.11	Grocery	12.07	No
4	Walgreen Boots	78924	8,002.0	9.81	-0.01	Drug Stores	9.86	No
17	TIJX	25012	2,770.0	9.37	0.08	Apparel	9.03	Yes
5	CVS incl Target	77792	9,813.0	8.83	0.26	Drug Stores	7.93	No
14	Rite Aid	27486	4,553.0	6.03	-0.00	Drug Stores	6.04	No
19	Dollar General	22234	13,350.0	1.77	0.13	Dollar Stores	1.67	No
20	Dollar Tree	21464	14,250.0	1.56	0.08	Dollar Stores	1.51	No
6	Amazon	71687	NaN	NaN	NaN	Electronic/Mail Order	NaN	Yes
11	Apple incl. Online	37664	NaN	NaN	NaN	NaN	NaN	Yes

10. ¿Que dispersión de ventas encontramos entre los 25 principales retailers de USA?

El rango de ventas ya lo hemos encontrado mas arriba por lo que lo traemos nuevamente. Pero a este dato le vamos a agregar el calculo del rango intercuartil, la desviación estandar y graficamos la dispersión de ventas con un boxplot.

```
In [ ]: range_sales = df_sorted['Sales'].max() - df_sorted['Sales'].min()
range_sales
```

```
Out [ ]: 641527
```

```
In [ ]: Q1 = df_sorted['Sales'].quantile(0.25)
Median = df_sorted['Sales'].median()
Q2 = df_sorted['Sales'].quantile(0.5)
Q3 = df_sorted['Sales'].quantile(0.75)
Min = df_sorted['Sales'].min()
Max = df_sorted['Sales'].max()
[Min, Q1, Median, Q2, Q3, Max]
```

```
Out [ ]: [16592, 24402.0, 34980.0, 34980.0, 71687.0, 658119]
```

```
In [ ]: iqr = Q3 - Q1
iqr
```

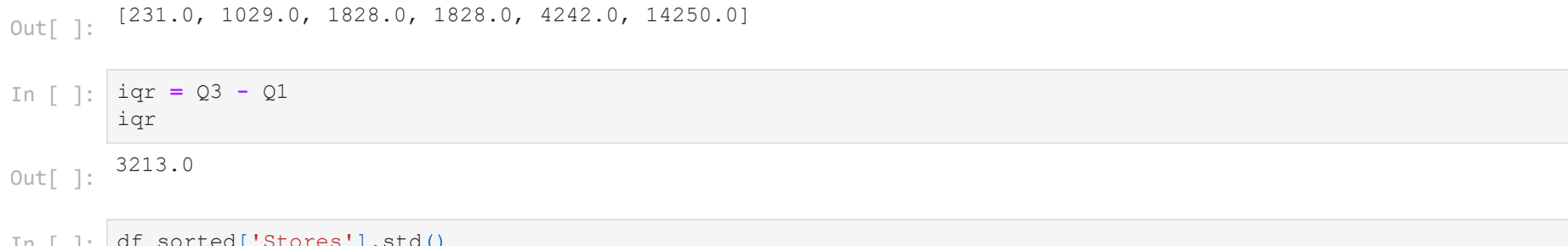
```
Out [ ]: 47285.0
```

```
In [ ]: df_sorted['Sales'].std()
```

```
Out [ ]: 125405.72764395573
```

```
In [ ]: sns.set(rc={'figure.figsize':(13,9)})
f, (ax_hist, ax_box) = plt.subplots(2, sharex=True, gridspec_kw={'height_ratios': (.6, .4)})
sns.histplot(df_sorted['Sales'], ax=ax_hist)
sns.boxplot(df_sorted['Sales'], ax=ax_box)
```

```
Out [ ]: <Figure: 13x9px>
```



Como podemos observar las ventas de los 25 principales retailers de USA presentan un rango de 641527 millones de dolares, un rango intercuartil de 47285 millones de dolares y una desviación estandar de 125405,72 millones de dolares. Podemos ubicar dentro de esta dispersión a un unico outlier, el cual es claramente Walmart US.

11. ¿Que dispersión de tiendas encontramos entre los 25 principales retailers de USA?

```
In [ ]: range_stores = df_sorted['Stores'].max() - df_sorted['Stores'].min()
range_stores
```

```
Out [ ]: 14019.0
```

```
In [ ]: Q1_2 = df_sorted['Stores'].quantile(0.25)
Median_2 = df_sorted['Stores'].median()
Q2_2 = df_sorted['Stores'].quantile(0.5)
Q3_2 = df_sorted['Stores'].quantile(0.75)
Min_2 = df_sorted['Stores'].min()
Max_2 = df_sorted['Stores'].max()
[Min_2, Q1_2, Median_2, Q2_2, Q3_2, Max_2]
```

```
Out [ ]: [231.0, 1029.0, 1828.0, 1828.0, 4242.0, 14250.0]
```

```
In [ ]: iqr = Q3 - Q1
iqr
```

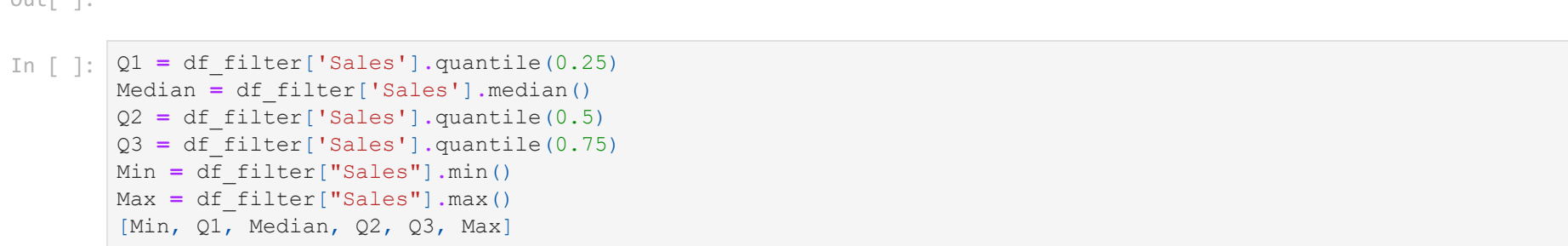
```
Out [ ]: 3213.0
```

```
In [ ]: df_sorted['Stores'].std()
```

```
Out [ ]: 4038.708713821636
```

```
In [ ]: sns.set(rc={'figure.figsize':(13,9)})
f, (ax_hist, ax_box) = plt.subplots(2, sharex=True, gridspec_kw={'height_ratios': (.6, .4)})
sns.histplot(df_sorted['Stores'], ax=ax_hist)
sns.boxplot(df_sorted['Stores'], ax=ax_box)
```

```
Out [ ]: <Figure: 13x9px>
```



Como podemos observar la cantidad de tiendas de los 25 principales retailers de USA presentan un rango de 14019 tiendas, un rango intercuartil de 3213 tiendas y una desviación estandar de 4038,70 tiendas. Podemos ubicar dentro de esta dispersión a tres outliers: Dolar Tree con 14250 tiendas, Dolar General con 13350 tiendas y CVS incl. Target con 9813 tiendas.

12. Entre las tiendas que solo venden de forma online ¿Es mayor o menor la dispersión de ventas y tiendas?

```
In [ ]: df_filter
```

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	Category	Sales/Avg. Store 2	Only online
1	Kroger	115037	3931.0	30.031850	0.053887	Grocery	29.264055	Yes
6	Amazon	71687	NaN	NaN	NaN	Electronic/Mail Order	NaN	Yes
11	Apple incl. Online	37664	NaN	NaN	NaN	NaN	NaN	Yes
12	Best Buy	34980	1389.0	24.685956	-0.038754	Electronics	25.183585	Yes
15	Ahold	26903	794.0	34.446863	0.033854	Grocery	33.882872	Yes
17	TIJX	25012	2770.0	9.369545	0.078241	Apparel	9.029603	Yes

```
In [ ]: range_sales = df_filter['Sales'].max() - df_filter['Sales'].min()
range_sales
```

```
Out [ ]: 90025
```

```
In [ ]: range_stores = df_filter['Stores'].max() - df_filter['Stores'].min()
range_stores
```

```
Out [ ]: 3137.0
```

```
In [ ]: Q1 = df_filter['Sales'].quantile(0.25)
Median = df_filter['Sales'].median()
Q2 = df_filter['Sales'].quantile(0.5)
Q3 = df_filter['Sales'].quantile(0.75)
Min = df_filter['Sales'].min()
Max = df_filter['Sales'].max()
[Min, Q1, Median, Q2, Q3, Max]
```

```
Out [ ]: [25012, 28922.25, 36322.0, 36322.0, 63181.25, 115037]
```

```
In [ ]: Q1_2 = df_filter['Stores'].quantile(0.25)
Median_2 = df_filter['Stores'].median()
Q2_2 = df_filter['Stores'].quantile(0.5)
Q3_2 = df_filter['Stores'].quantile(0.75)
Min_2 = df_filter['Stores'].min()
Max_2 = df_filter['Stores'].max()
[Min_2, Q1_2, Median_2, Q2_2, Q3_2, Max_2]
```

```
Out [ ]: [794.0, 1240.25, 2079.5, 2079.5, 3060.25, 3931.0]
```

```
In [ ]: iqr_sales = Q3 - Q1
iqr_sales
```

```
Out [ ]: 34259.0
```

```
In [ ]: iqr_stores = Q3_2 - Q1_2
iqr_stores
```

```
Out [ ]: 1820.0
```

```
In [ ]: df_filter['Sales'].std()
```

```
Out [ ]: 35254.98122393487
```

```
In [ ]: df_filter['Stores'].std()
```

```
Out [ ]: 1408.7883683032974
```

```
In [ ]: sns.set(rc={'figure.figsize':(13,9)})
f, (ax_hist, ax_box) = plt.subplots(2, sharex=True, gridspec_kw={'height_ratios': (.6, .4)})
sns.histplot(df_filter['Sales'], ax=ax_hist)
sns.boxplot(df_filter['Sales'], ax=ax_box)
```

```
Out [ ]: <Figure: 13x9px>
```



```
In [ ]: sns.set(rc={'figure.figsize':(13,9)})
f, (ax_hist, ax_box) = plt.subplots(2, sharex=True, gridspec_kw={'height_ratios': (.6, .4)})
sns.histplot(df_filter['Stores'], ax=ax_hist)
sns.boxplot(df_filter['Stores'], ax=ax_box)
```

```
Out [ ]: <Figure: 13x9px>
```



Podemos observar que la dispersión disminuye cuando solo retailers que operan exclusivamente online al menos en cuanto a cantidad de tiendas. En este caso estamos hablando de tiendas que no están abiertas al público sino que funcionan como centros de almacenamiento y distribución desde donde se entregan los pedidos una vez realizados. En cuanto a ventas la dispersión también se ve disminuida, pero seguimos manteniendo un outlier entre nuestra muestra. En este caso se trata de Kroger.