

Análisis de los 25 retailers más grandes de Estados Unidos



Francisco Ecatl Melo Valle / Proyecto Datacademy

Published at Feb 27, 2022

```
from IPython.display import display
```

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

```
# Importar datos aquí
df = pd.read_csv("largest_us_retailers.csv")
df.head()
```

	Company object	Sales int64	Stores float64	Sales/Avg. St...	Store Count G...	Cate
0	Walmart US	658119	4574	65.64972497	0.01284322409	Super
1	Kroger	115037	3931	30.03184963	0.05388739946	Groce
2	Costco	90048	495	187.7956204	0.06681034483	Ware
3	Home Depot	83976	1965	42.73587786	0	Home Impr
4	Walgreen Boots	78924	8002	9.81092672	-0.01051069618	Drug

```
df.describe()
```

	Sales float64	Stores float64	Sales/Avg. St...	Store Count G...	
count	25	23	23	23	
mean	71063.08	3466.5652173913045	36.87096578552174	0.04023000882013044	
std	125405.72764395573	4038.708713821636	40.62527709398271	0.06855210542240484	
min	16592	231	1.56391854	-0.05951506245	

25%	24402	1029	11.263462319999999	-0.0018599562365	
50%	34980	1828	25.99773328	0.01952035694	
75%	71687	4242	41.358955215	0.07893390251	
max	658119	14250	187.7956204	0.2556621881	

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Company               25 non-null    object
1   Sales                 25 non-null    int64
2   Stores               23 non-null    float64
3   Sales/Avg. Store     23 non-null    float64
4   Store Count Growth   23 non-null    float64
5   Category             24 non-null    object
dtypes: float64(3), int64(1), object(2)
memory usage: 1.3+ KB
```

I. Preguntas del negocio

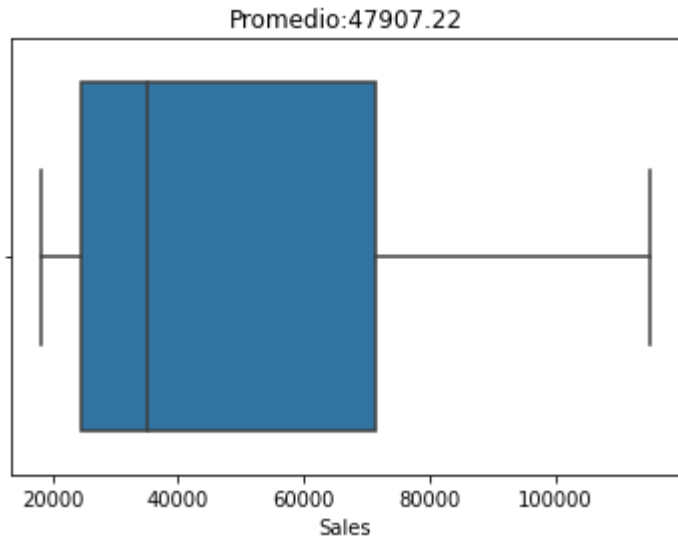
1. ¿Cuál es el promedio de ventas sin contar a la compañía dominante?

```
sns.boxplot(df["Sales"][1:-1])
promedio = str(df["Sales"][1:-1].mean().round(2))
```

```
plt.title("Promedio:" + promedio)
```

```
/shared-libs/python3.7/py/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the fol
FutureWarning
```

```
Text(0.5, 1.0, 'Promedio:47907.22')
```



```
# sns.histplot(df[1:-1], x="Sales")
# plt.title("La mayoría de la empresas sin contar la numero uno gano entre: 18,201 y 34,340.3")
```

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

```
f, (ax_box, ax_hist) = plt.subplots(2, sharex=True, gridspec_kw={"height_ratios": (.15, .85)})

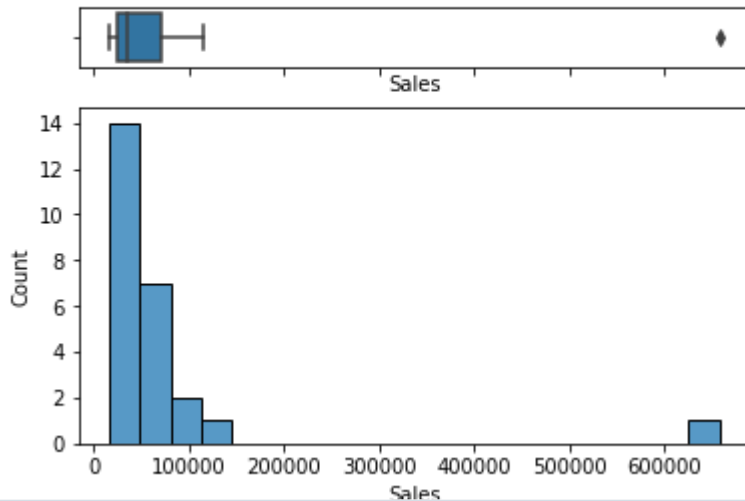
sns.histplot(df, x="Sales", ax=ax_hist)
sns.boxplot(df["Sales"], ax=ax_box)

txt_1 = "La mayoría de las compañías gano entre: 16,592 y 48,668.3"
plt.figtext(0.5, 1, txt_1, wrap=True, horizontalalignment='center', fontsize=12)
```

```
/shared-libs/python3.7/py/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the fol
FutureWarning
```

```
Text(0.5, 1, 'La mayoría de las compañías gano entre: 16,592 y 48,668.3')
```

La mayoría de las compañías gana entre: 16,592 y 48,668.3



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

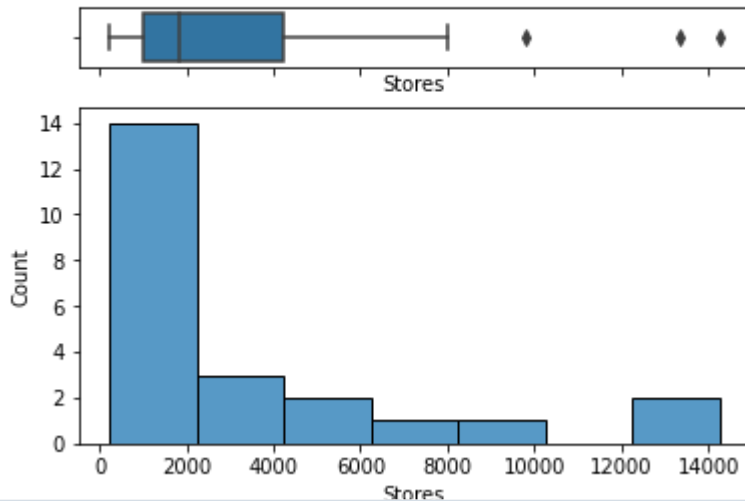
```
f, (ax_box, ax_hist) = plt.subplots(2, sharex=True, gridspec_kw={"height_ratios": (.15, .85)})
con = sns.histplot(df, x="Stores", ax=ax_hist)
sns.boxplot(df["Stores"], ax=ax_box)

txt_1 = "La mayoría de las compañías tiene entre: 231 y 2,233 tiendas"
plt.figtext(0.5, 1, txt_1, wrap=True, horizontalalignment='center', fontsize=12)
```

```
/shared-libs/python3.7/py/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the fol
FutureWarning
```

```
Text(0.5, 1, 'La mayoría de las compañías tiene entre: 231 y 2,233 tiendas')
```

La mayoría de las compañías tiene entre: 231 y 2,233 tiendas

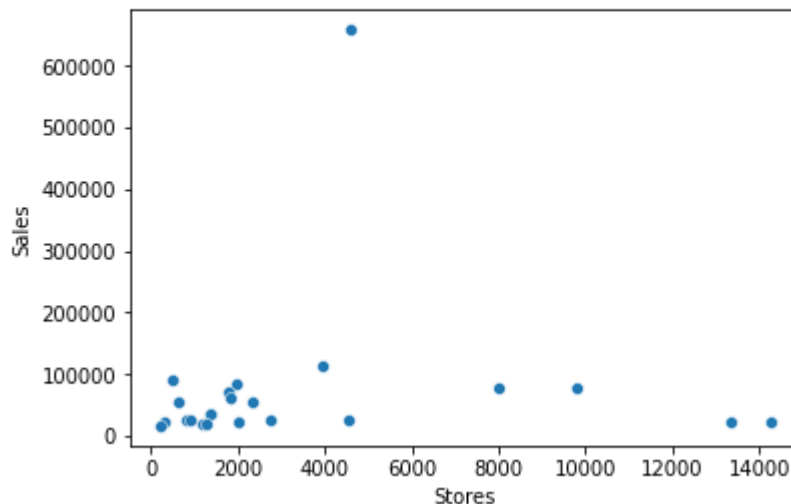


4. ¿La cantidad de tiendas está relacionada con la cantidad de ventas? Es decir, ¿si una empresa tiene más tiendas tendrá más dinero de ventas?

```
sns.scatterplot(data=df, x = 'Stores', y = "Sales")
txt = "La compañía con mas tiendas no es la numero uno en ventas \n La numero uno en ventas t
plt.figtext(0.5, 1, txt, wrap=True, horizontalalignment='center', fontsize=12)
```

Text(0.5, 1, 'La compañía con mas tiendas no es la numero uno en ventas \n La numero uno en ventas

La compañía con mas tiendas no es la numero uno en ventas
La numero uno en ventas tiene menos de un tercio en comparacion de la que tiene mas tiendas



5. ¿Cuál es el rango que existe entre las ventas?

```
max_sales = df["Sales"].max()
min_sales = df["Sales"].min()
range_sales = max_sales - min_sales

print("Ventas maximas:", max_sales)
print("Ventas minimas:", min_sales)
print("Rango entre las venats:", range_sales)
```

Ventas maximas: 658119
 Ventas minimas: 16592
 Rango entre las venats: 641527

6. ¿Cuáles son las 5 empresas que más tiendas físicas tienen? ¿Cuáles de ellas están dentro de las 5 empresas que más ventas tuvieron?

```
more_stores = df.nlargest(5, "Stores")
more_sales = df.nlargest(5, "Sales")
print("Compañías con mas tiendas físicas")
display(more_stores)
print("De las cinco compañías con mas tiendas físicas, solo dos estan entre las que mas venta")
pd.merge(more_stores, more_sales, on='Company')
```

Compañías con mas tiendas físicas

	Company	Sales	Stores	Sales/Avg. Store	Store Count Growth	
20	Dollar Tree	21464	14250.0	1.563919	0.079627	Doll:
19	Dollar General	22234	13350.0	1.768885	0.132412	Doll:
5	CVS incl. Target	77792	9813.0	8.825959	0.255662	Drug
4	Walgreen Boots	78924	8002.0	9.810927	-0.010511	Drug
0	Walmart US	658119	4574.0	65.649725	0.012843	Supe

De las cinco compañías con mas tiendas físicas, solo dos estan entre las que mas ventas tienen

	Company object	Sales_x int64	Stores_x float64	Sales/Avg. St...	Store Count G...	Cate
0	Walgreen Boots	78924	8002	9.81092672	-0.01051069618	Drug

1	Walmart US	658119	4574	65.64972497	0.01284322409	Super
---	------------	--------	------	-------------	---------------	-------

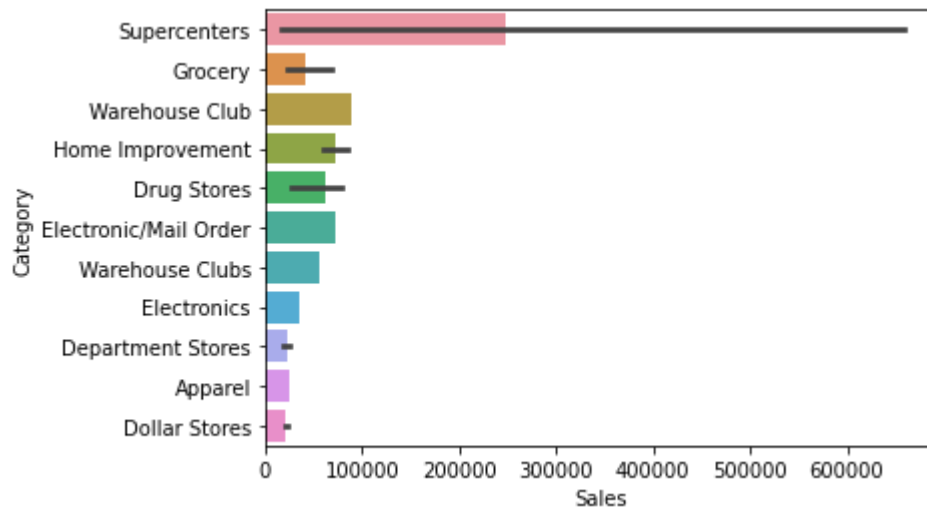
II. Preguntas opcionales

7. ¿Qué categoría de compañía generó más ventas?

```
print("La categoria con mas ventas es: Supercenters")
sns.barplot(data=df, x="Sales", y="Category")
```

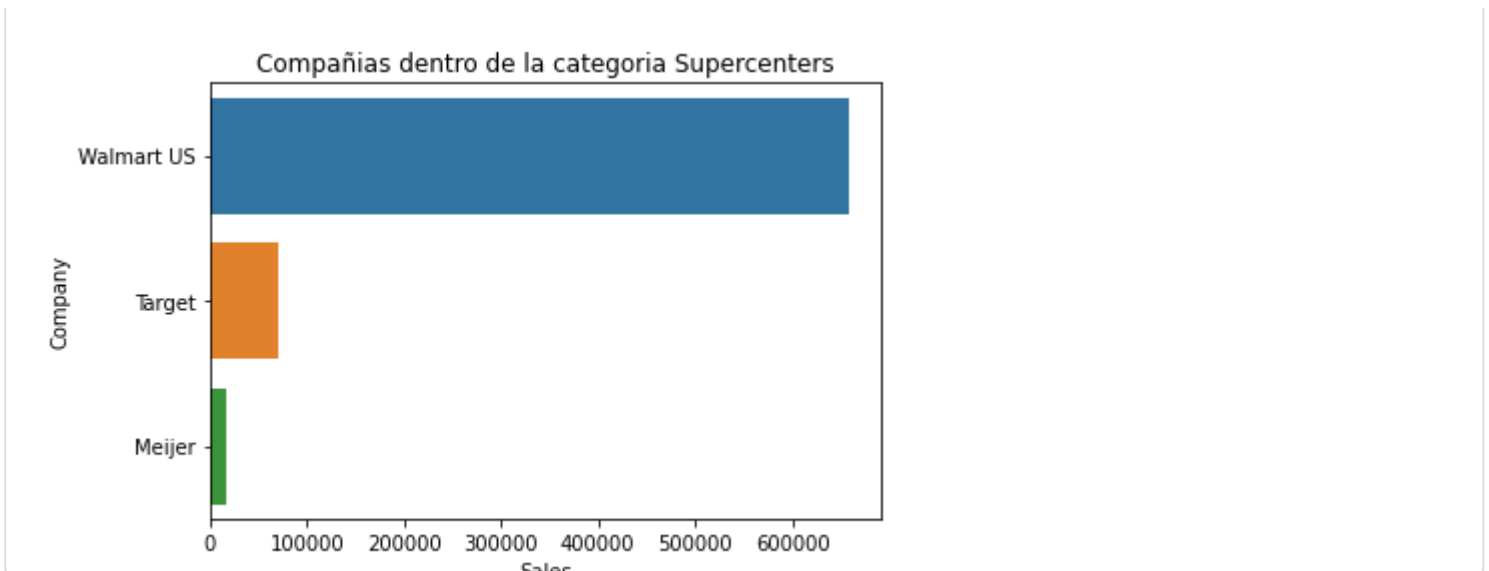
La categoria con mas ventas es: Supercenters

<AxesSubplot:xlabel='Sales', ylabel='Category'>



```
company_supercenters = df[(df["Category"]=="Supercenters")]
sns.barplot(data=company_supercenters, x="Sales", y="Company")
plt.title("Compañías dentro de la categoria Supercenters", size=12)
```

Text(0.5, 1.0, 'Compañías dentro de la categoria Supercenters')



8. ¿Cuál es la compañía que en relación con su cantidad de tiendas físicas genera más ventas?

```
sns.barplot(data=df, x = 'Sales/Avg. Store' , y="Company")
plt.title("Costco Es la compañía que genera mas ventas por tienda")
```

Text(0.5, 1.0, 'Costco Es la compañía que genera mas ventas por tienda')



9. ¿Cuáles compañías venden de forma exclusivamente online? ¿Sus ventas destacan sobre las que también tienen tiendas físicas?


```
print("Amazon y Apple no cuentan con tiendas físicas")
print("amazon esta en el lugar numero 7 en ventas")
print("Apple esta en el lugar nuero 12 en ventas")
print("Sus ventas si destacan sobrelas que tienen tiendas físicas")
df[df['Stores'].isnull()]
```

Amazon y Apple no cuentan con tiendas físicas

amazon esta en el lugar numero 7 en ventas

Apple esta en el lugar nuero 12 en ventas

Sus ventas si destacan sobrelas que tienen tiendas físicas

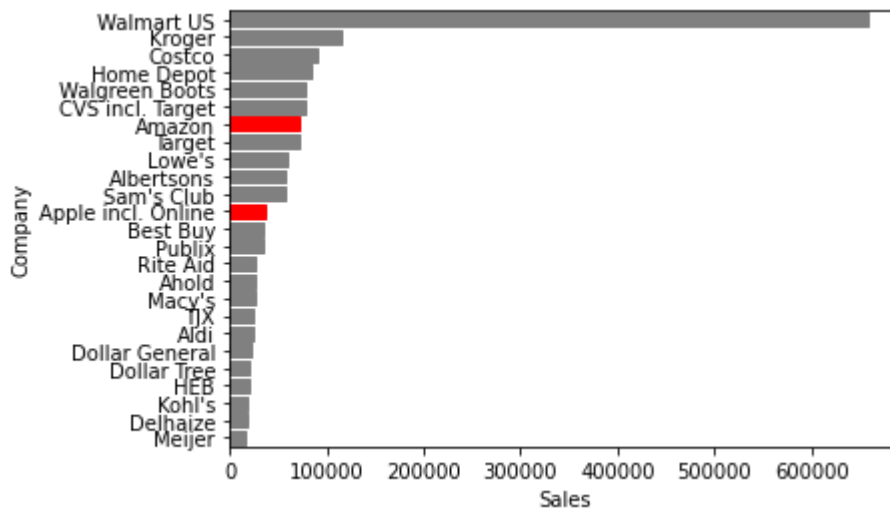
	Company object	Sales int64	Stores float64	Sales/Avg. St...	Store Count G...	Cate
6	Amazon	71687	nan	nan	nan	Elec Orde
11	Apple incl. Online	37664	nan	nan	nan	nan

```
idx_no_stores =[6, 11]
```

```
clrs = ["red" if x in idx_no_stores else "grey" for x in df["Company"] ]
```

```
ax = sns.barplot(data=df, x="Sales", y="Company")
```

```
for bar in ax.patches:
    if bar.get_y().round() in idx_no_stores:
        bar.set_color('red')
    else:
        bar.set_color('grey')
```



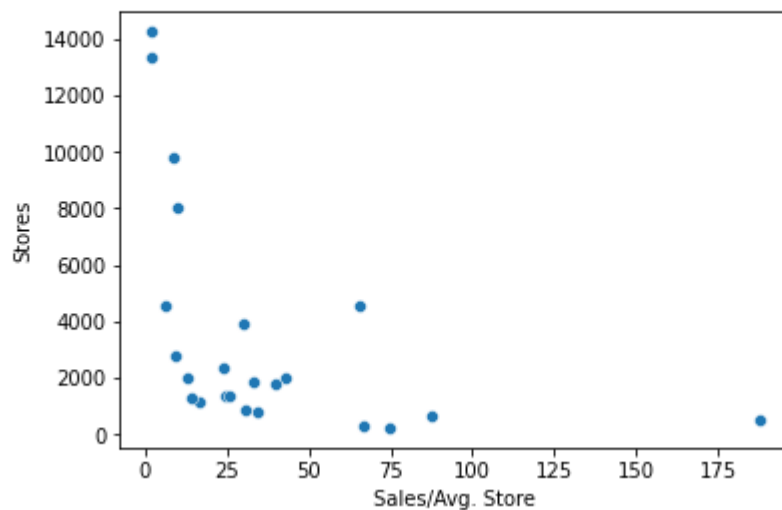
III. Tus propias preguntas

10. Existe relacion entre numero de tiendas y ventas por tienda?

```
sns.scatterplot(data=df, x="Sales/Avg. Store", y="Stores")
txt = "En la grafica se puede notar que a en la matoria de las compa ias a moyor tiendas tie
plt.figtext(0.5, 1, txt, wrap=True, horizontalalignment='center', fontsize=12)
```

```
Text(0.5, 1, 'En la grafica se puede notar que a en la matoria
```

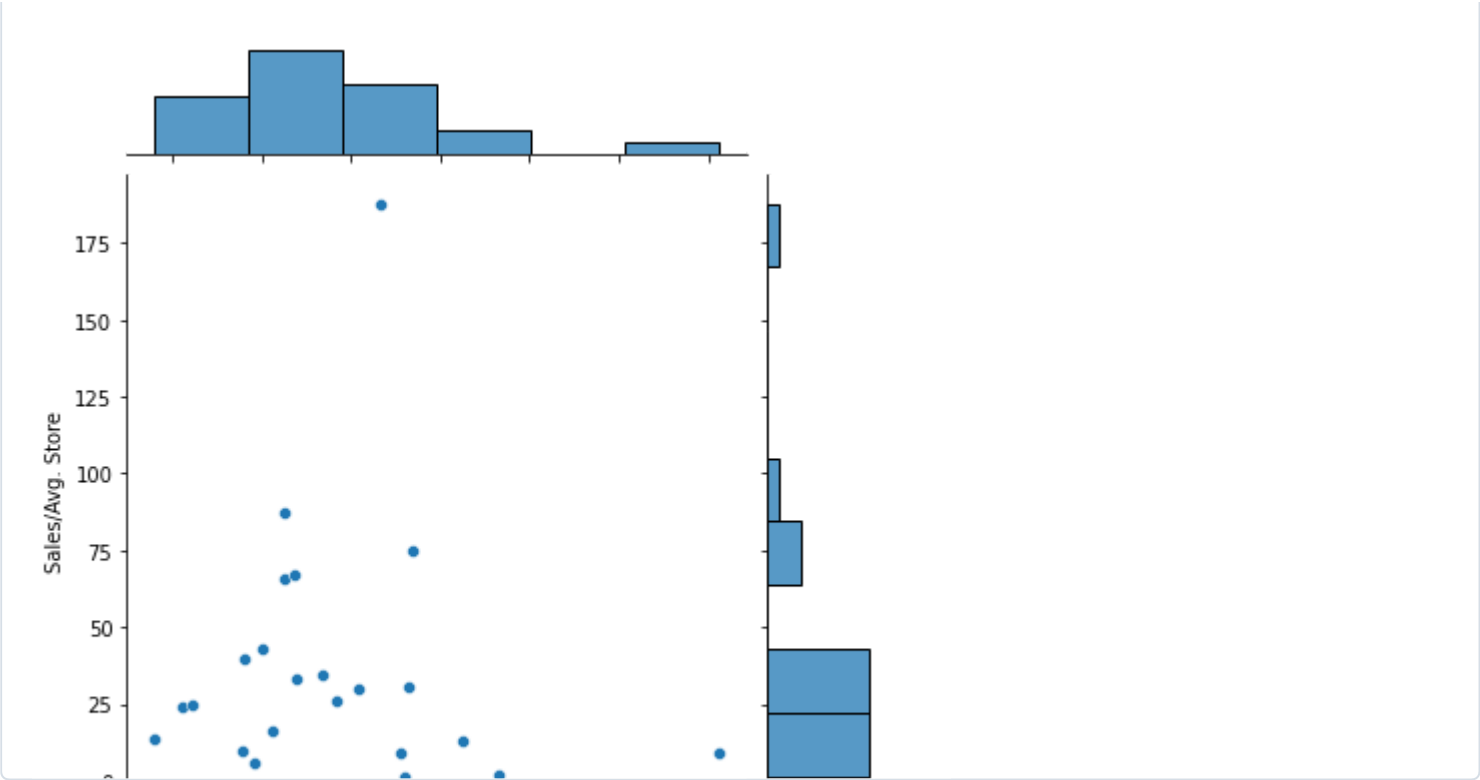
En la grafica se puede notar que a en la matoria de las compa ias a moyor tiendas tiene menores son sus ventas y



11. Pregunta

```
sns.jointplot(data=df, x="Store Count Growth", y="Sales/Avg. Store")
"Sales/Avg. Store"
```

```
'Sales/Avg. Store'
```



12. Pregunta