jr.code.training.21

Wizeline Data Engineering Bootcamp Challenge

Technical Challenge Part1 - Coding

Codesignal - Completed

Technical Challenge Part2 - Practical Applications

How many commercial chains are monitored, and therefore, included in this database?

705 Commercial Chains where founded

In [6]:
```
grouped = df.groupby("cadenaComercial").agg({
    "cadenaComercial": 'count'
})

grouped.count()
```

Out[6]:
```
cadenaComercial    705
dtype: int64
```

We obtained this result by grouping the data taken as a base the field “cadenaComercial"

What are the top 10 monitored products by State?

The result is a dataset of 321 rows

Out[18]:

	estado	producto	precio
0	AGUASCALIENTES	FUD	12005
1	AGUASCALIENTES	DETERGENTE P/ROPA	10188
2	AGUASCALIENTES	LECHE ULTRAPASTEURIZADA	9824
3	AGUASCALIENTES	SHAMPOO	9654
4	AGUASCALIENTES	REFRESCO	9481
...	...	...	...
29336	ZACATECAS	SHAMPOO	15012
29337	ZACATECAS	CHILES EN LATA	14866
29338	ZACATECAS	COMPONENTES DE AUDIO	14799
29339	ZACATECAS	REFRESCO	13925
30360	estado	producto	20

321 rows × 3 columns

In [18]:
```
products_grouped = df.groupby(['estado','producto'])[['precio']].count().reset_index()

products_sorted = products_grouped.groupby(['estado']).apply(lambda x: x.sort_values(['precio'],ascending = False)).reset_index(drop = True)

products_sorted.groupby(['estado']).head(10)
```

We obtained this result by grouping the data taken as a base the fields “estado" and “producto”, counting how many products we have on each classification, using after that a sorting function to have the top products, and at the final using a function to select just the top 10 rows per group

Which is the commercial chain with the highest number of monitored products?

Wal-Mart has the most products monitored

Out[17]:

cadenaComercial	producto
WAL-MART	8643133
BODEGA AURRERA	6765453
SORIANA	6546211
MEGA COMERCIAL MEXICANA	4899509
CHEDRAUI	4221625
...	...

In [17]:

```
groupedbystate_max = df.groupby(['cadenaComercial']).agg({
    'producto': 'count'
}).sort_values(['producto'], ascending=False)

groupedbystate_max
```

We obtained this result by grouping the data taken as a base the field “cadenaComercial" and using a sort function based on the “product” field to have the top record

Use the data to find an interesting fact.

As an interesting fact, we found that the most monitored products was the “REFRESCO”

Out[23]:

producto	precio
REFRESCO	1247981
DETERGENTE P/ROPA	990122
FUD	933410
LECHE ULTRAPASTEURIZADA	886716
SHAMPOO	745467

In [23]:

```
products_monitored_grouped = df.groupby(['producto'])[['precio']].count().sort_values(['precio'], ascending=False)

products_monitored_grouped
```

What are the lessons learned from this exercise?

There are different ways to find the results, but depending on the knowledge we have with the tools we can choose how to deliver the results promptly and accurately

Can you identify other ways to approach this problem? Explain.

Since I had no previous experience working with pandas, I think I was able to deliver the same results using other tools, like excel or some database