Data engineering bootcamp

Wizeline

Sharon Camacho

Monitored products in Mexico

In this presentation we will have an initial approach to product monitoring in Mexico

Monitored products in Mexico

Universe

705	139	33	41	1107
comercial chains	municipalies	states	categories	products

The highest monitores products
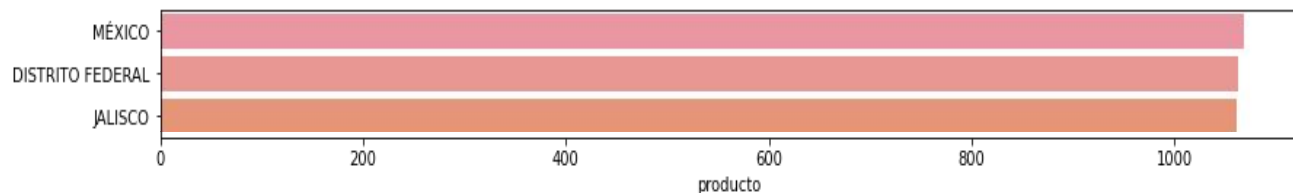
SORIANA

Top 3 monitoring
REFRESCO
DETERGENTE P/ROPA
FUD

top 10 monitored products by State?

Top 3 States by monitored products

México

product_name	#products
REFRESCO	287463
FUD	207569
LECHE ULTRAPASTEURIZADA	175640
DETERGENTE P/ROPA	173452
YOGHURT	136720
CERVEZA	136686
MAYONESA	131103
CHILES EN LATA	130598
JABON DE TOCADOR	129889
SHAMPOO	125603

Distrito Federal

product_name	#products
REFRESCO	194939
FUD	149141
DETERGENTE P/ROPA	132862
LECHE ULTRAPASTEURIZADA	116522
JABON DE TOCADOR	97330
YOGHURT	94852
MAYONESA	94286
CHILES EN LATA	92539
SHAMPOO	92307
CERVEZA	91747

Jalisco

index	producto
REFRESCO	81718
LECHE ULTRAPASTEURIZADA	81149
DETERGENTE P/ROPA	73854
FUD	61680
CHILES EN LATA	55783
JABON DE TOCADOR	55196
SHAMPOO	54409
MAYONESA	53886
DESODORANTE	46335
TOALLA FEMENINA	46019

Products by State

In general, States have a similar diversity of product monitored. Approximately, 30% of state monitor fewer products
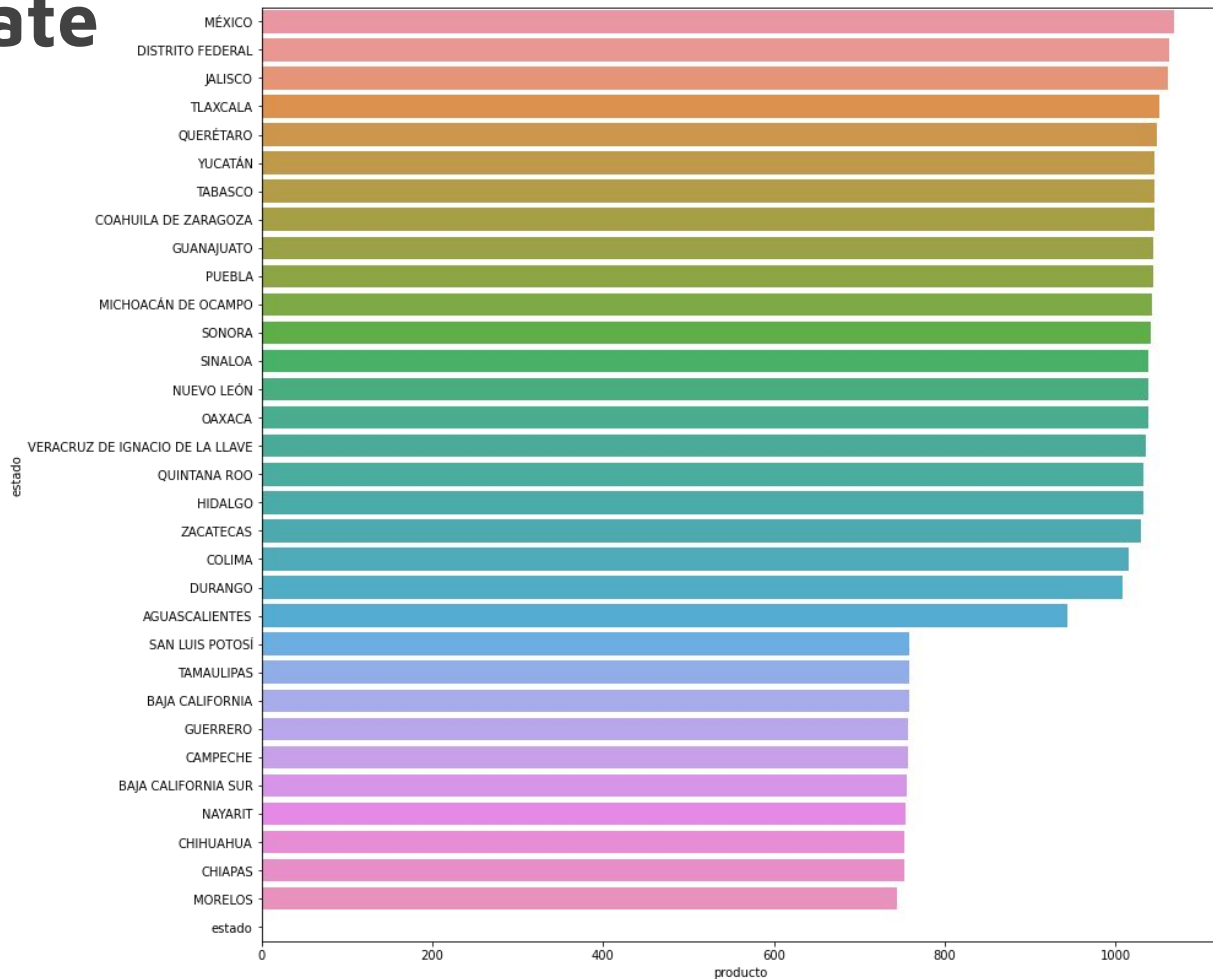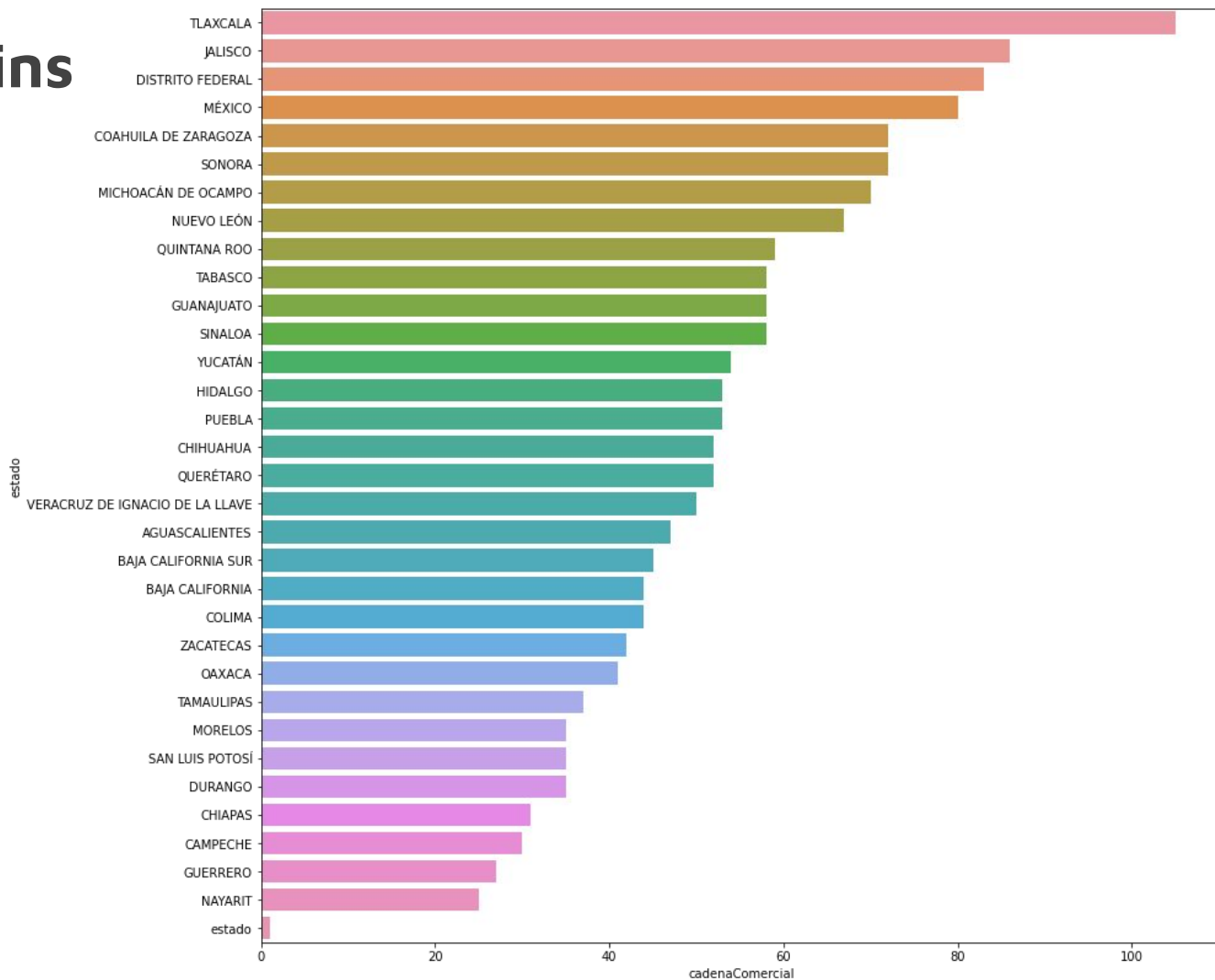
Commercial chains by State

Tlaxcala, who is the 4th in most monitored products, is the first with the most commercial chains.

Here we saw a significant difference in the number of commercial chains by state.
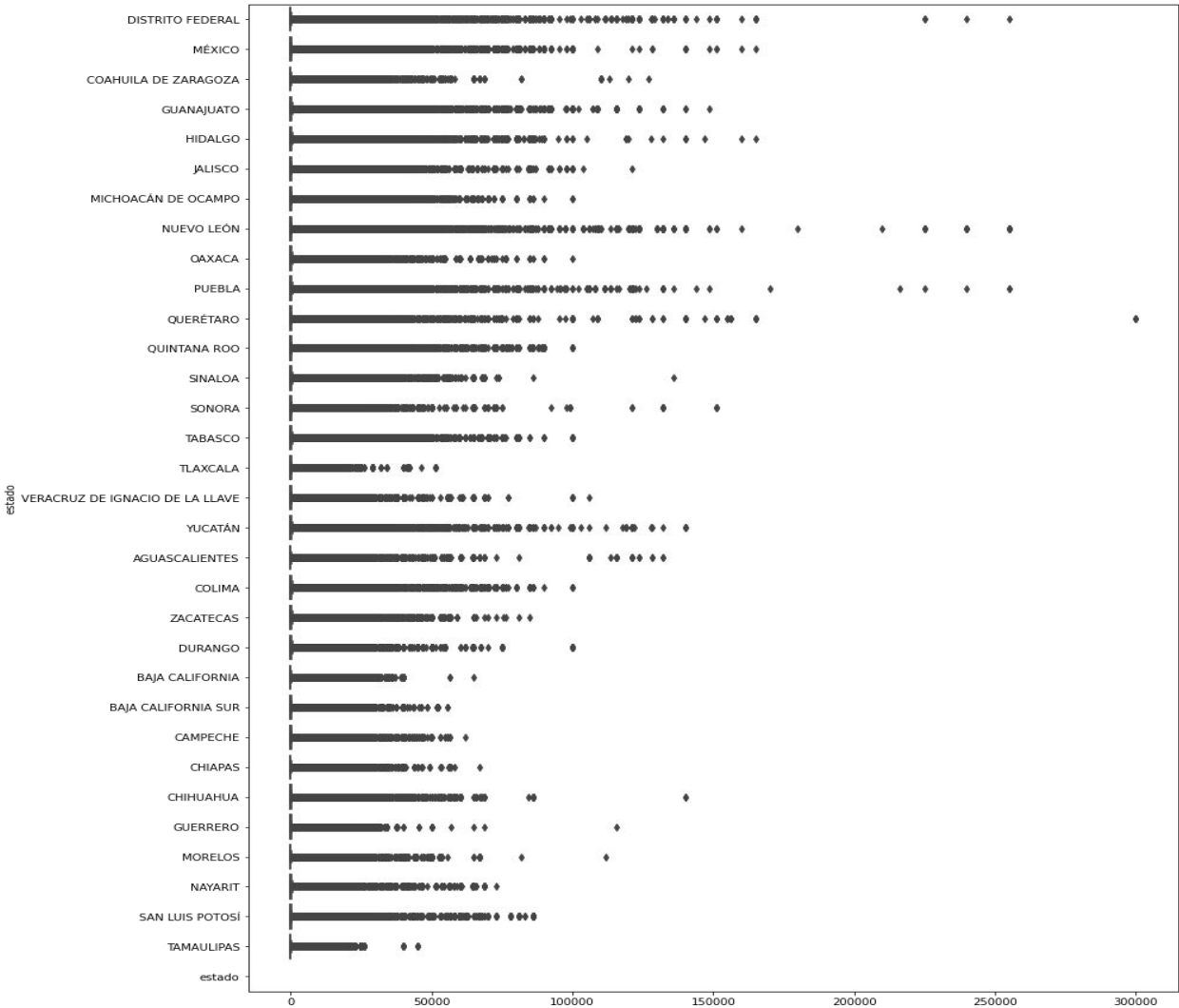
Prices by state

We saw a lot of outliers prices in all states, the highest are in Distrito Federal, Nuevo León and Puebla.

It would be interesting to investigate what happen there, if the movement of products increases prices, demand, or some other factor.
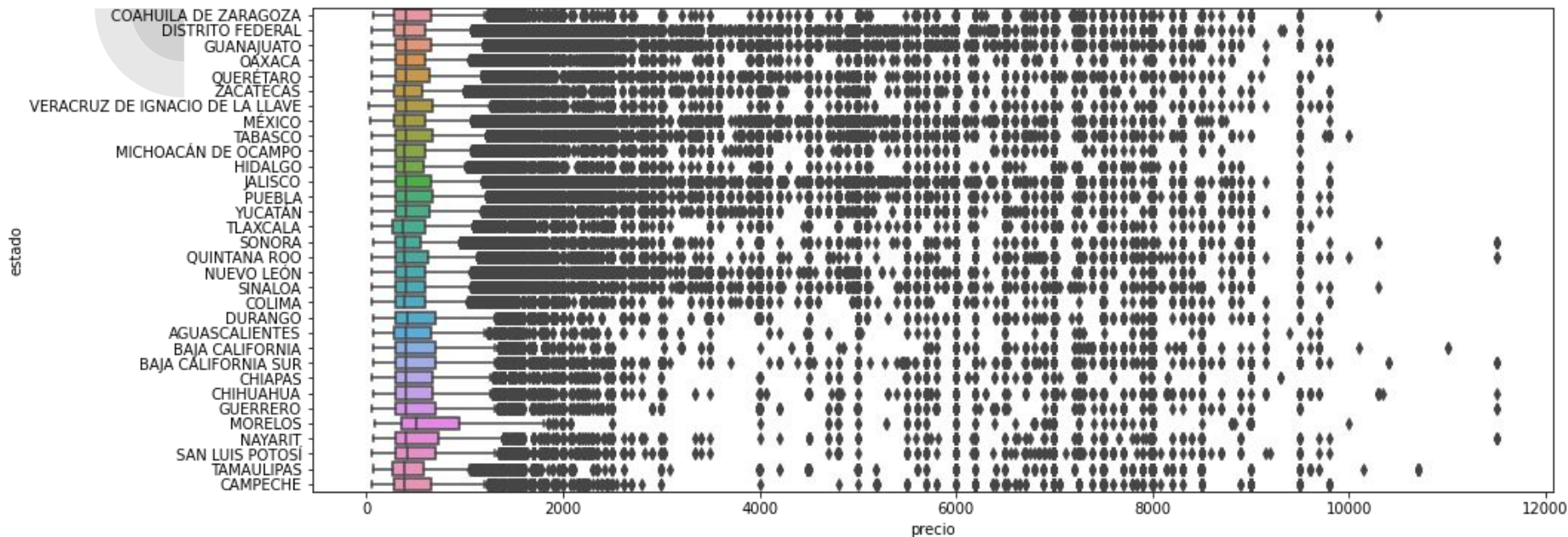
Prices by Category (½)

Here we saw where we have the outliers: Aparatos eléctricos, aparatos electrónicos, Medicamentos, Frutas y Legumbres Procesadas, Muebles de cocina, Pescados y Mariscos, Utensilios domésticos and Juguetes.

With this it will be easier to do the research
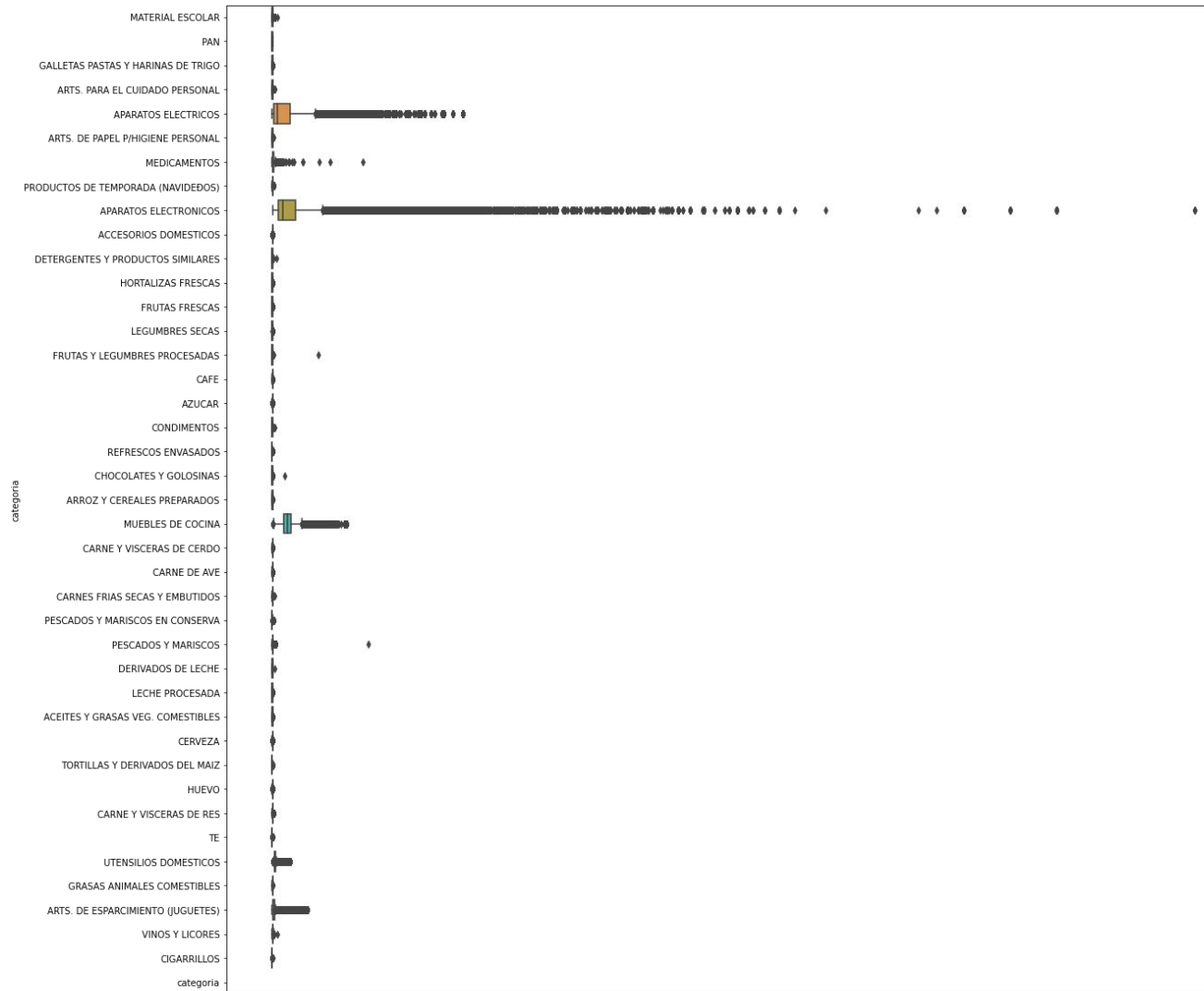
Prices by Category (2/2)

If we study the prices of Juguetes by state, we can find that the median price of Morelos is higher than the other states.

We can also identify the extreme outliers in Sonora, Quintana Roo, Baja California, Guerrero and Nayarit.

We this, we can look more closely in municipalities and products to find special situations

What are the lessons learned from this exercise?

1) It was a very interesting exercise, especially because it states that 80% of the time in data analysis, it takes to organize the data.
2) It would be interesting to know the business perspective to manage null values
3) The work would be easier through a data pipeline, and have some other visualization tool to automate the process. I think this is where the work of the data engineer would come in.
4) It is good to do the cleaning process for the given sample, however it is better to work with the total data because there are aspects that the initial sample did not have.

Can you identify other ways to approach this?

As I said, make a data pipeline would be ideal to manage data and make it available to users to carry out this type of analysis.

Having a meeting with the business to know the real interest of the analysis could speed up the work and generate greater value to the investigation

Thanks!

I really enjoy this experience!

It took me more than 3 hours, but I reinforced a lot of knowledge, I learned new ones.

I would really like to have some feedback from you.