

Clustering Neighborhoods by Mexican Food Type in Mexico City

Constantino Carreto Romero

Introduction

- We are in the Mexican street food (e.g. tacos, tortas, pizza, etc.)



- we want to start new branches in Mexico City.
- where to place our new branches?
- Solution: cluster neighborhoods according to food type presence

Data

- ▶ We firstly collect postal codes for Mexico City from the Mexican Postal Service's website. This data is used to obtain geo-coordinates for each neighborhood.
- ▶ We then use the Foursquare API to collect trending venues in each neighborhood.

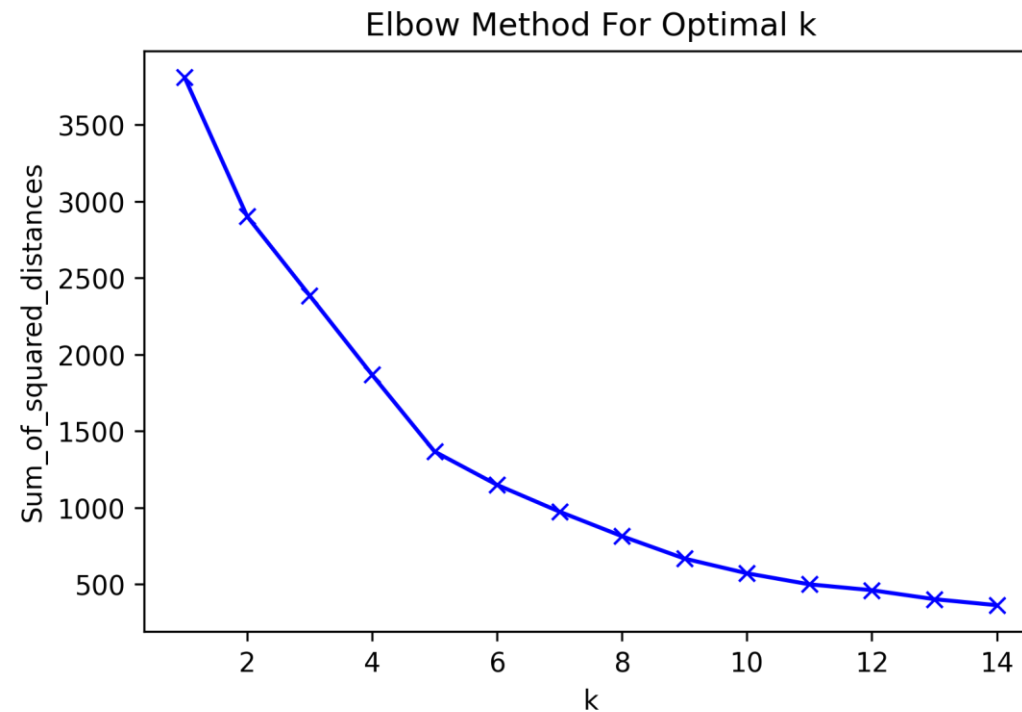
Index	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Latitude	Longitude	Venue Category
22103	San Pedro Mártir	Tlalpan	19.2665	-99.17	Centro de nutrición Casa Blanca	19.265388	-99.167554	Athletics & Sports
22104	San Pedro Mártir	Tlalpan	19.2665	-99.17	Las Quekas De "Don Fer"	19.265361	-99.167104	Snack Place
22105	San Pedro Mártir	Tlalpan	19.2665	-99.17	Café Año	19.265204	-99.166857	Snack Place
22106	San Pedro Mártir	Tlalpan	19.2665	-99.17	The Food Stop	19.268408	-99.166852	Food Truck
22107	San Pedro Mártir	Tlalpan	19.2665	-99.17	Aqua Sur	19.265319	-99.165622	Gym Pool

- Our final dataset is the input for our cluster analysis. This final dataset looks like this:

Index	Neighborhood	burgers	pizza	quesadillas	tacos	tortas
0	10 de Abril	0.25	0.0	0.0	0.500000	0.250000
1	12 de Diciembre, San Andrés Tomatlán, San Andr...	0.25	0.0	0.0	0.750000	0.000000
2	15 de Agosto	0.00	0.0	0.0	1.000000	0.000000
3	16 de Septiembre	0.20	0.0	0.0	0.800000	0.000000
4	1a Sección Cañada, 2a Sección Cañada	0.00	0.0	0.0	0.857143	0.142857

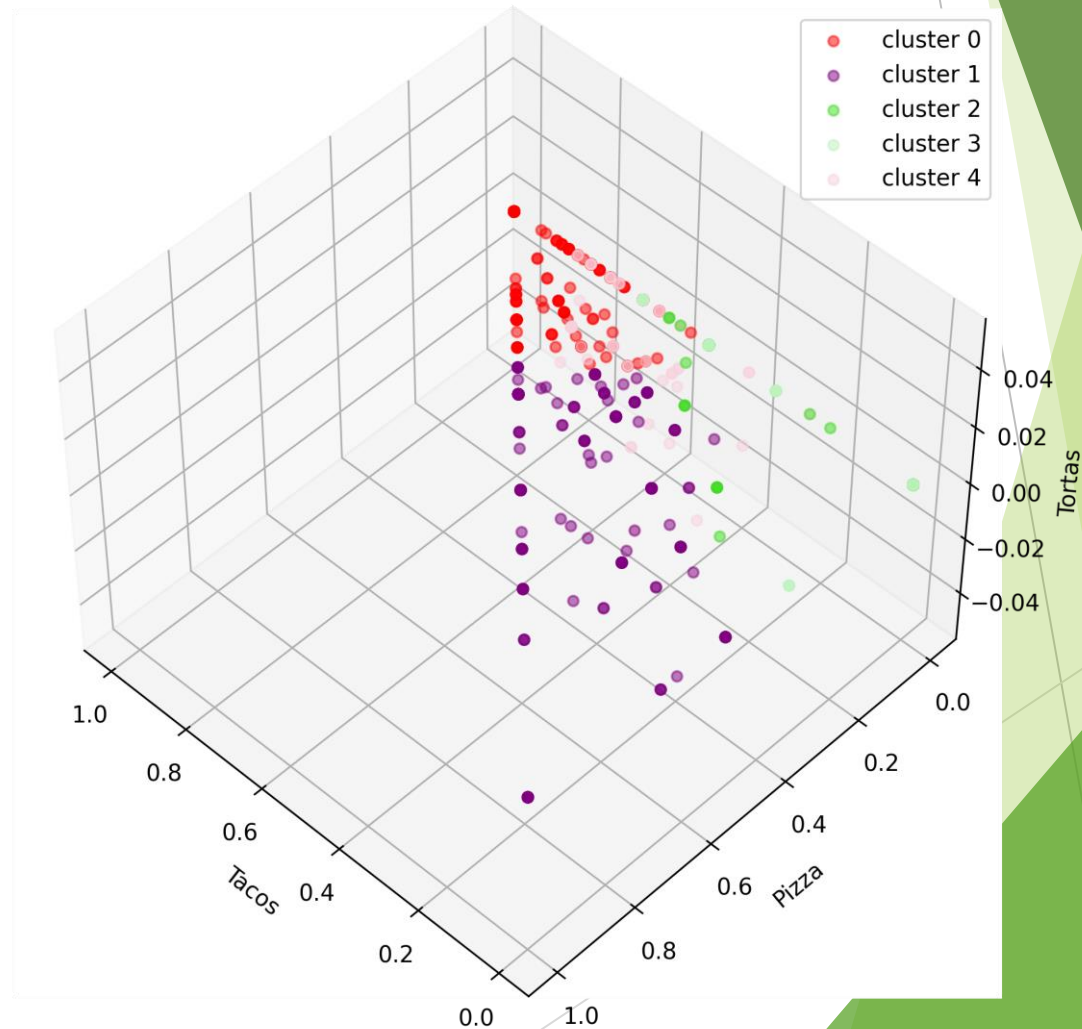
Clustering

- I use k-means clustering.
- To decide number of clusters I use the elbow method. It results $k=5$.

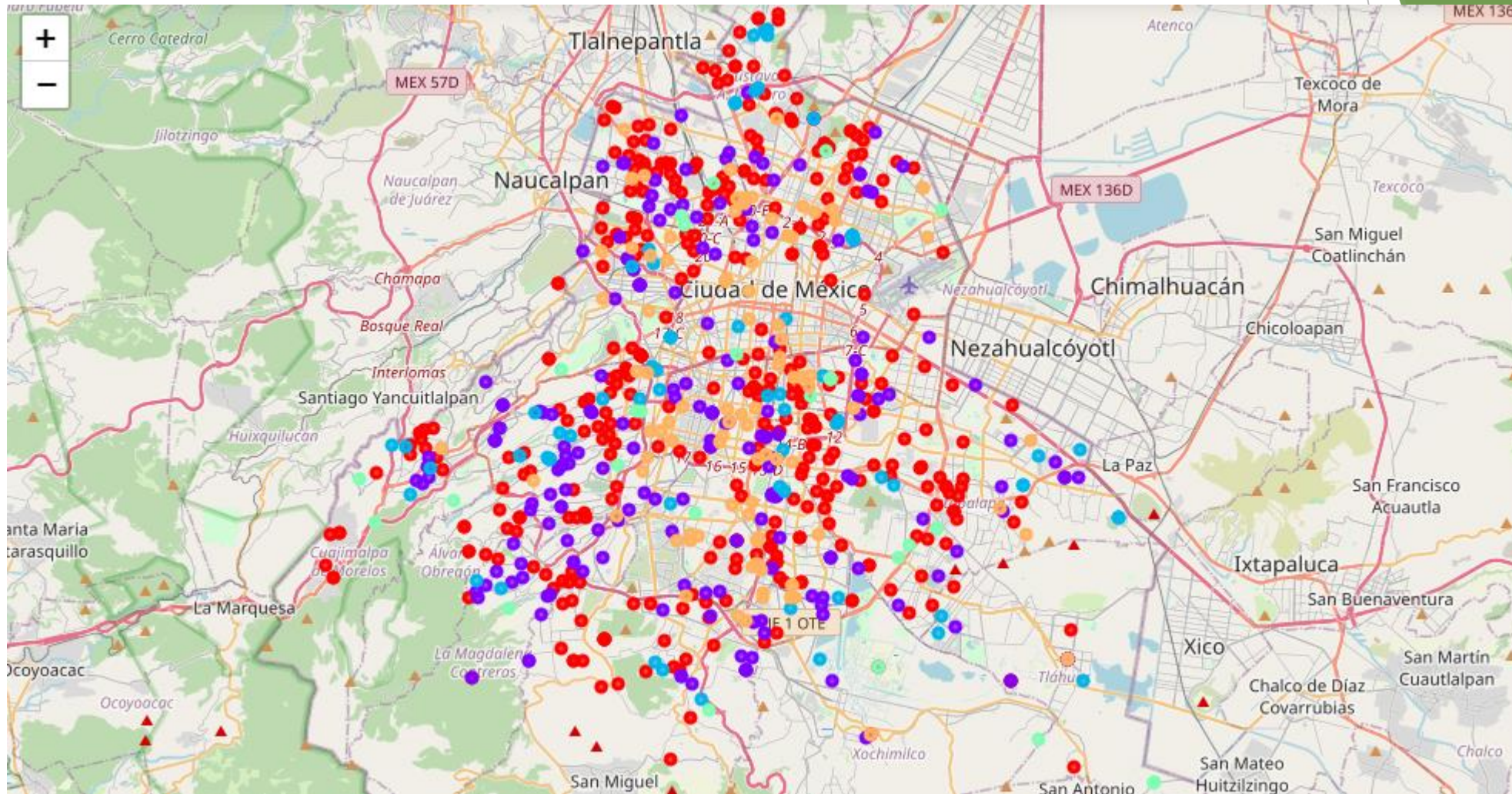


- I obtain the following cluster occurrences:
- If we plot three characteristics (food categories), the cluster distribution is as follows:

Cluster	Occurrences
0	390
1	215
4	102
2	75
3	28



- And the geographical distribution is as follows:



Results

- The percent of predominance of each food type:

Clus_km	burgers	pizza	quesadillas	tacos	tortas
0	0.063307	0.039313	0.018002	0.875768	0.003609
1	0.134216	0.409322	0.010247	0.441836	0.004378
2	0.700968	0.024325	0.004167	0.26454	0.006
3	0.077381	0.083333	0.52381	0.315476	0
4	0.087537	0.109599	0.016388	0.543003	0.243473

- ▶ Cluster 0:
 - ▶ This is the most common cluster. We note it is characterized by an overwhelming presence of tacos venues.
- ▶ Cluster 1:
 - ▶ It is characterized by a predominance of pizza and tacos businesses.
- ▶ Cluster 2:
 - ▶ It is characterized by an important predominance of hamburger places with a weaker presence of tacos.
- ▶ Cluster 3:
 - ▶ It is characterized by an important predominance of quesadillas businesses with a weaker presence of tacos.
- ▶ Cluster 4:
 - ▶ It is characterized by an important predominance of tacos business with a weaker presence of tortas and pizza.

Conclusions

- ▶ This information might be helpful to decide where to start a new business, depending on the type of food we want to offer.
- ▶ Let's say we want to start a new quesadillas business, then we now know that in neighborhoods in cluster 2 there is high demand for this type of food given the predominance of these businesses.
- ▶ At the same time, this also implies there is high competition here.
- ▶ A final decision on whether a neighborhood is suitable to start a new branch must come accompanied by a market analysis about the competition conditions in that neighborhood.