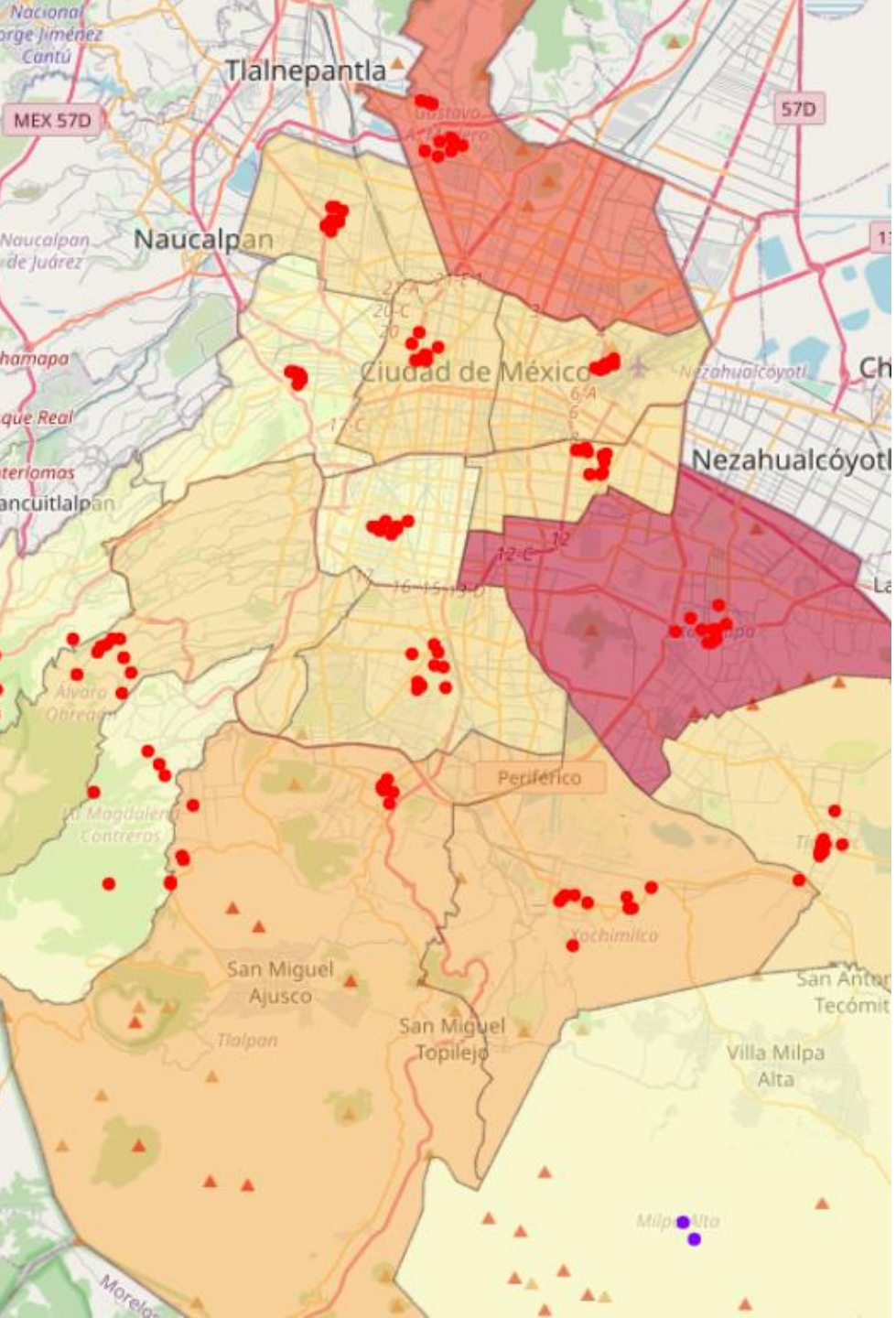



# Geographical cluster of the possible COVID-19 infection points in México City.

---

Andres J. Ramos

# Backgound

SARS-CoV-2 it’s a virus that has the initial outbreak in Wuhan China in December 2019 and has stopped most of the economic and social activities around the world.

In Mexico City, for example, there are boroughs that are presenting large outbreaks and want to return to economic and social activities.

**The problem**

- Is it possible to classify the busiest places to generate a map of risk of contagion?

# Data acquisition and cleaning

- The coronavirus data was obtained from the official government website https://datos.cdmx.gob.mx/explore/dataset/casos-asociados-a-covid-19/table/?disjunctive.resultado.
- Data containing patient health characteristics were removed
- Only the active cases were kept.
- Cleaned data contains 2 features

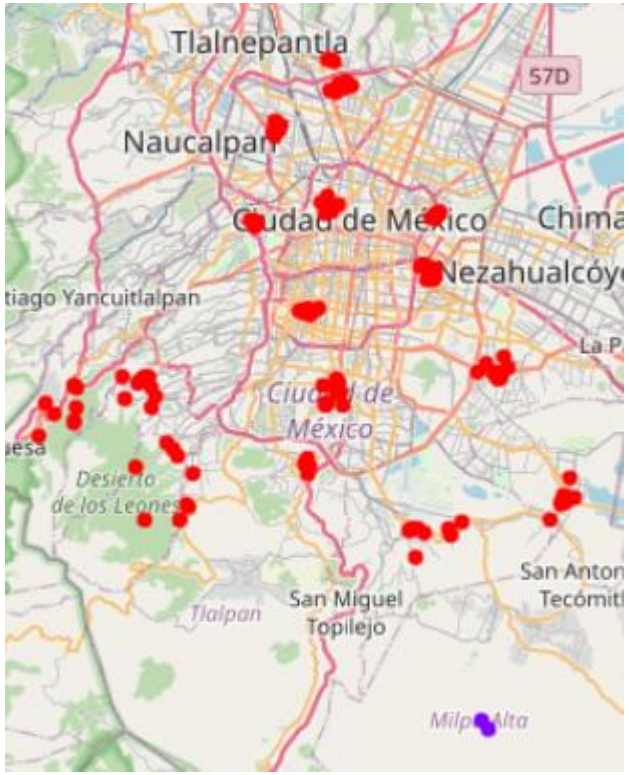
- For geographic data, the Foursquare API was used.

---

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Azcapotzalco | 19.485815 | -99.184206 | La Concherìa CDMX | 19.483789 | -99.185843 | Bakery |
| 1 | Azcapotzalco | 19.485815 | -99.184206 | Café ONCE28 | 19.484427 | -99.185720 | Breakfast Spot |
| 2 | Azcapotzalco | 19.485815 | -99.184206 | La Perla Tapatía | 19.483741 | -99.185856 | Mexican Restaurant |
| 3 | Azcapotzalco | 19.485815 | -99.184206 | Neko Café | 19.484152 | -99.183326 | Japanese Restaurant |
| 4 | Azcapotzalco | 19.485815 | -99.184206 | Centro Verde Azcapotzalco | 19.487757 | -99.182125 | Garden |

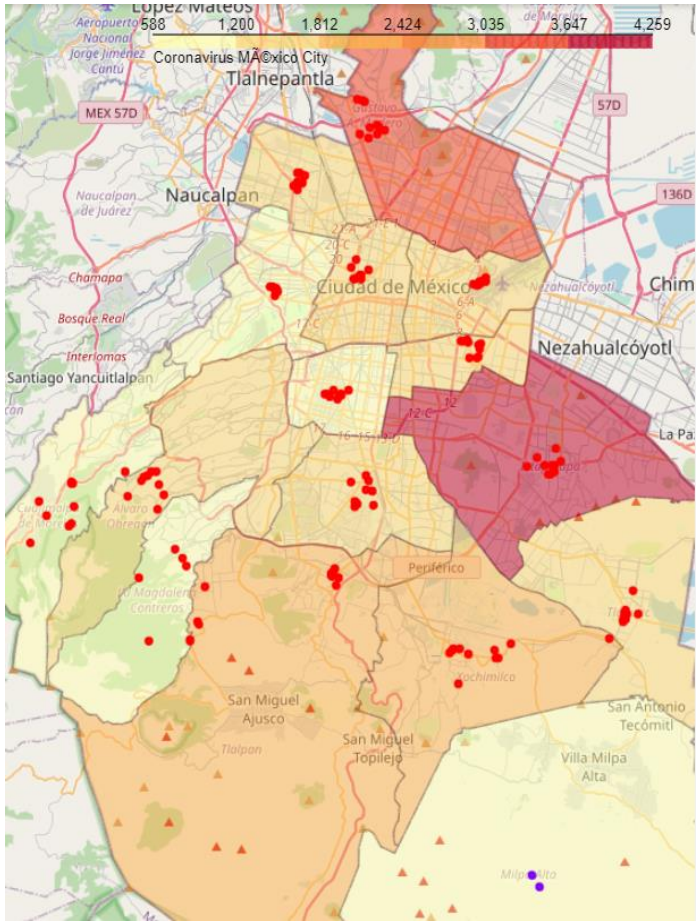
# Classification model

- The model used is K-means
- The model classified the locations into 2 clusters, this may mean that the locations are very similar to each other or that there is a underfitting.



# Plot the data into a map

- For this step I used a choropleth map and combine whit the coronavirus data to obtain the risk of infection



# Conclusion and future directions

- Knowing the possible points of risk can help generate better health measures
- Accuracy can be improved by sorting places by:
  - Most visited areas
  - Places  around the infected people