FINDING SIMILAR NEIGHBORHOOD

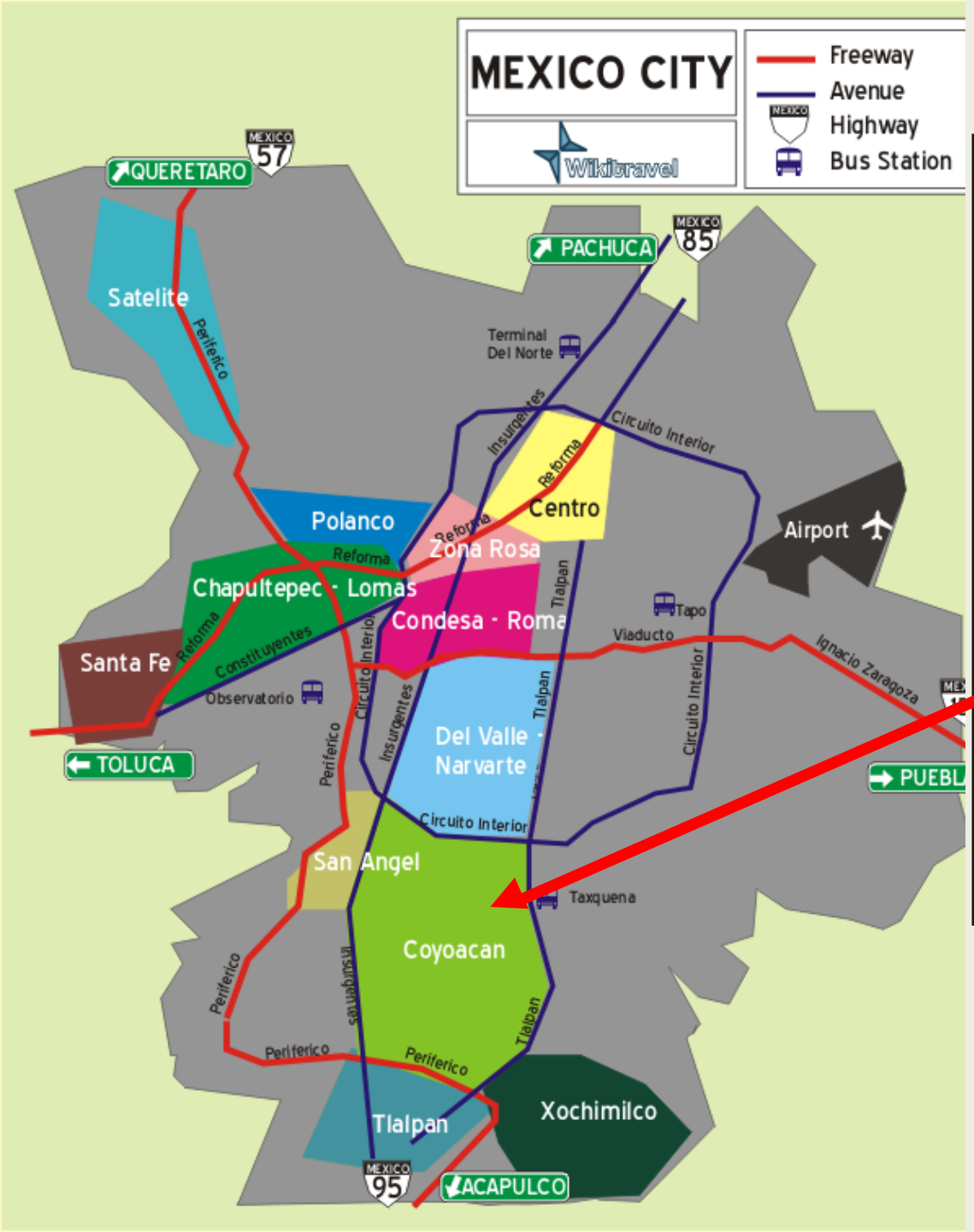
IBM Applied Data Science Specialization

Capstone Project

Author: RUBEN ADAD

Definition of the problem

- I am going to move from Mexico City to Madrid.
- I want to to find neighborhoods in Madrid that have the same types of attractions and places of interest (venues) as those in the neighborhood where I currently live.



CURRENT NEIGHBORHOOD: Coyoacan Mexico City

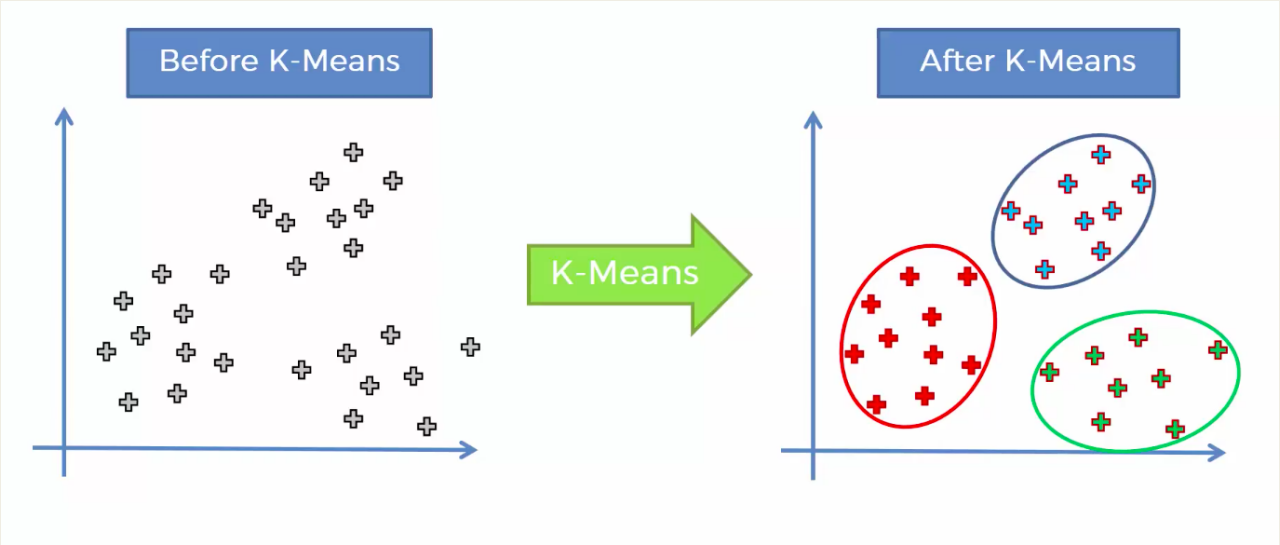
MADRID
Neighborhoods

Objective

- Find neighborhoods similar to another one that serves as a point of comparison.
- In this case, the similarity is based on the venues in the neighborhoods, according to the information provided by the Foursquare API.

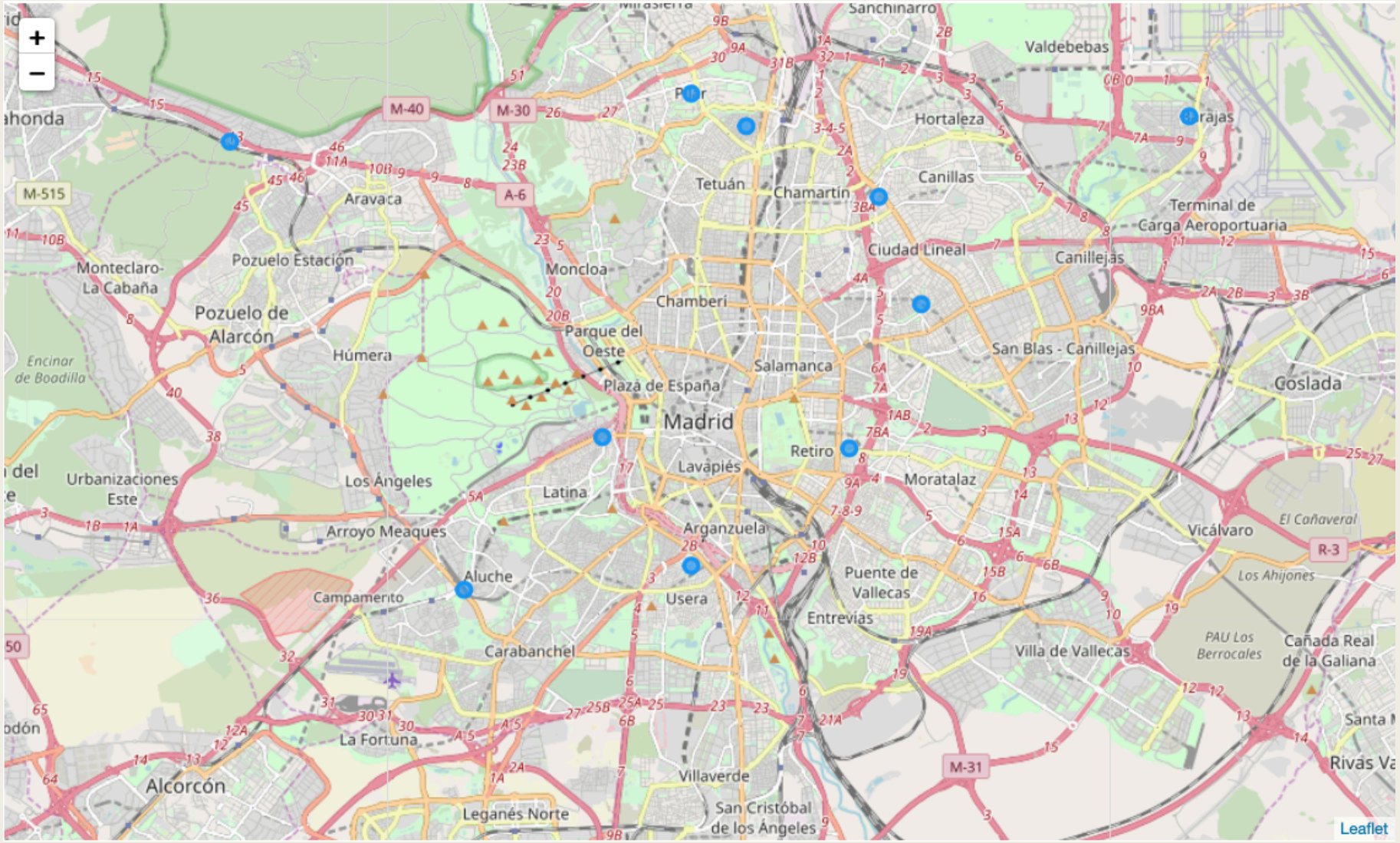
Steps followed

- Use Foursquare API to collect venues from current neighborhood.
  - *Result: 84 rows for 1 neighborhood*
- Use Foursquare API to collect venues from Madrid neighborhood.
  - *Result: 3,767 rows for 118 neighborhoods*
- Join both results and transform “venue category” using “onehot- encoding”.
  - *Result: 256 unique categories*
- Group ”venue categories” for each neighborhood using the mean.
- Use KMEANS to create clusters of similar neighborhoods.
- Find the closest neighborhoods to the current neighborhood using the distance to centroid from KMEANS.



KMEANS ALGORITHM

Similar neighborhoods in Madrid



Conclusions

- It is possible to help people who are in a situation similar to the one described in this case, using public data available through the Foursquare API.
- The final decision cannot be based solely on the results of this analysis.
  - *Rather it should be considered as a tool to narrow the options that must be investigated in greater detail.*
- One way to enrich the results of the analysis would be by adding demographic and socioeconomic attributes to each of the neighborhoods.