Finding similar cities, South US and Mexico

Walt Mayfield

May 1, 2019

**Introduction**

Every year more than 150,000 people immigrate to the US from Mexico. My project will cluster US and Mexican cities to show which are similar, to help guide immigrants in finding a good city.

**Data Source**

I will use city, population, latitude and longitude data from http://www.geonames.org/ to find the 50 most populace cities in the Southern US and Mexico. Then I will use Foursquare.com to find 100 venues within a 5km radius of the city to define a type of city.

**Methodology**

1. Download the dataset from geonames.org which contains city, population, latitude and longitude data
2. Select the 50 most populace cities in US states which share a border with Mexico and the 50 most populace states in Mexico.
3. Use Foursquare to pull 100 venues within 5 km of each city’s coordinates.
4. Extract the type of venue for each row
5. One hot encode the type of venue for each venue
6. Take a mean of the one hot encoded values to find frequency of each venue type in each city. This creates a venue type vector for each city.
7. Use the venue type vector to group cities using k-means.
8. Explore the groups to learn which cities are similar.

**Results**

I segmented the 100 cities into 30 clusters. Four of those clusters representing 15 cities contained both us and Mexican cities. The remaining 26 clusters only contained cities from one country. This shows that most cities of the countries are more similar to cities of the same country.

Of the four clusters which contained cities from both countries the largest contained the following data:



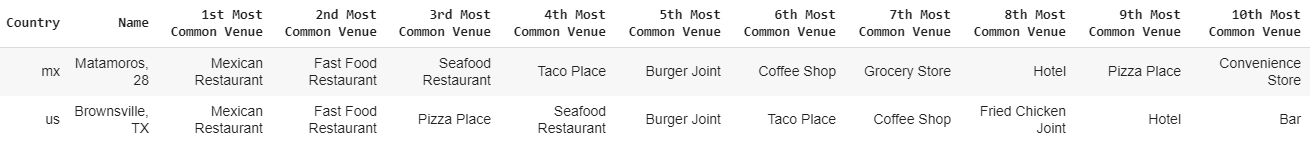
This cluster shows that Corpus Christi, TX is similar to the Mexican cities Zapopan, Mexicali, Hermosillo, Morelia, Veracruz, Lopez Mateos, and Villahermosa. The large frequency of Mexican and Italian Restaurants make these cities similar.

Here is the second cluster



El Paso would be a good target city for a person who likes the Mexican Cities of Juarez or Torreon.

The US town of Brownsville is similar to the Mexican city Matamorros due to the prevalence of Mexican and fast food.



And unsurprisingly the US city of Laredo is similar to the Mexican city Nuevo Laredo.



**Discussion**

There were fewer clusters containing both US and Mexican cities than I expected. The population sizes were bigger for most of the US cities than Mexican ones. I recommend doing this analysis with cities with similar population sizes. That may create more clusters which mix US and Mexican cities.

**Conclusion**

This report shows which US and Mexican cities are similar.