Kevin Bittner

Applied Data Science Capstone

January-16-21

**Introduction**

The problem presented is that an entrepreneur in Miami, Florida is interested in establishing a yoga studio but does not know the best neighborhood in which to establish the business. The individual is familiar with Miami neighborhoods but would like to use more advanced methodologies to inform the location decision. Clustering analysis is one likely methodology that could be beneficial and worth exploring. Any business owner or individual in establishing a business in Miami could be interested in the results.

**Data**

A data set is needed that describes the businesses in Miami neighborhoods as well as the geographic location of each neighborhood. This neighborhood data was found in the form of a wikipedia page (https://en.wikipedia.org/wiki/List\_of\_neighborhoods\_in\_Miami). This page includes Miami neighborhood names, coordinates, and also population data. This data was imported using the Python package mysoup, and converted into a Pandas data frame. Business data will be extracted from Foursquare and added to the data set in order to populate business type and location.

**Methodology**

A Python script was created to extract neighborhood data from a wikipedia page: <https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami>. One neighborhood did not include coordinates and was subsequently removed from the data set. The data set includes the names of Miami neighborhoods, locations, as well as population. Given that Miami is a heavily touristed area, population data alone is unlikely to be a good predictor of business traffic, although still valuable. Sorting the data frame by population shows that the neighborhood of Little Havana has the greatest population of 76,163. Next business data was extracted using Foursquare and a new data set was created. Then, this data was clustered into five groups. The clustering methodology was used to show which cluster of neighborhoods would be ideal for a yoga studio. Any neighborhood with a yoga studio as the most popular venue is likely not an ideal choice to establish a new yoga studio.

**Results**

The clustering process successfully sorted the neighborhoods into five clusters. The first cluster is the largest and includes 18 of the 24 Miami neighborhoods. The second cluster includes one neighborhoods, there are two neighborhoods in the third cluster as well as the fourth and there is one neighborhood in the final cluster.

**Discussion**

Given the size of the first cluster this was first reviewed as a potential cluster of neighborhoods to establish a yoga studio. However, Lummus Park in the first neighborhood shows the first most prevalent establishment is yoga studios. Review of the remaining fours clusters shows that yoga studios do not fall in the top ten venues for any of the four clusters making them suitable clusters to establish a yoga studio. All else equal, a cluster with more neighborhoods rather than fewer is likely to yield a more suitable location to establish a yoga studio, so the third and the fourth clusters are reviewed. Review of the businesses located in the third cluster shows that “Boat or Ferry” is the most common venue in each neighborhood of the cluster. This shows that the cluster is coastal and likely to be a large tourist location, which is not ideal as the entrepreneur wishes to focus on local residents. The fourth cluster appears to be perfect for establishing a yoga studio with no yoga studios present in the top ten venues. Ideal neighborhoods for establishing a yoga studio would be West Flagler or Little Havana. Little Havana also yields the highest population of any Miami neighborhood, making it the ideal location to establish a yoga studio.

**Conclusion**

In conclusion, the Little Havana neighborhood of Miami is ideal for a yoga studio. This is supported by the absence of yoga studios in the top establishments of both Little Havana and the other neighborhood in the cluster as a result of cluster analysis.