

IBM Capstone Project Report

1. Problem Description

Many individuals that are not familiar with travel outside of their home country may not know where to start when beginning to travel internationally. A good rule of thumb is to find and research a city that has a similar sociocultural landscape to one you have lived in or visited. As an avid domestic and international traveler, I have been able to experience two somewhat similar cities in Austin, Texas, USA, and Seoul, South Korea. They share a similar area at around 250 sq. miles, and are both known for their plethora of food options, nightlife, music venues, and other exciting tourist sites. As such, for my capstone project I will be performing a k-Means clustering analysis to determine the similarities and differences between Austin and Seoul.

2. Data Description and Usage

In order to perform my analysis, I will be using the following data:

- Foursquare API : <https://foursquare.com/>
- Districts of Seoul - https://en.wikipedia.org/wiki/List_of_districts_of_Seoul
- Austin Neighborhoods - https://en.wikipedia.org/wiki/List_of_Austin_neighborhoods

The Foursquare API will provide information about the latitude and longitude of the districts/neighborhoods in both cities, as well as interesting spots to visit around each of their districts/neighborhoods. The information about the districts/neighborhoods themselves will help in naming each district/neighborhood and separating clusters of them by their similarity or dissimilarity to one another.

3. Methodology

There were 3 phases of my EDA and machine learning processes: Seoul Calculations & Mapping, Austin Calculations & Mapping, and Seoul vs. Austin Comparison. In the first phase, I defined the credentials and version for the Foursquare API. Then I explored venues in the first neighborhood in the dataset, Dobong-gu. Then I created a function to repeat the same process to all the neighborhoods in Seoul, and analyzed each neighborhood. This involved looking at the parameters such as the top 5 most common venues.

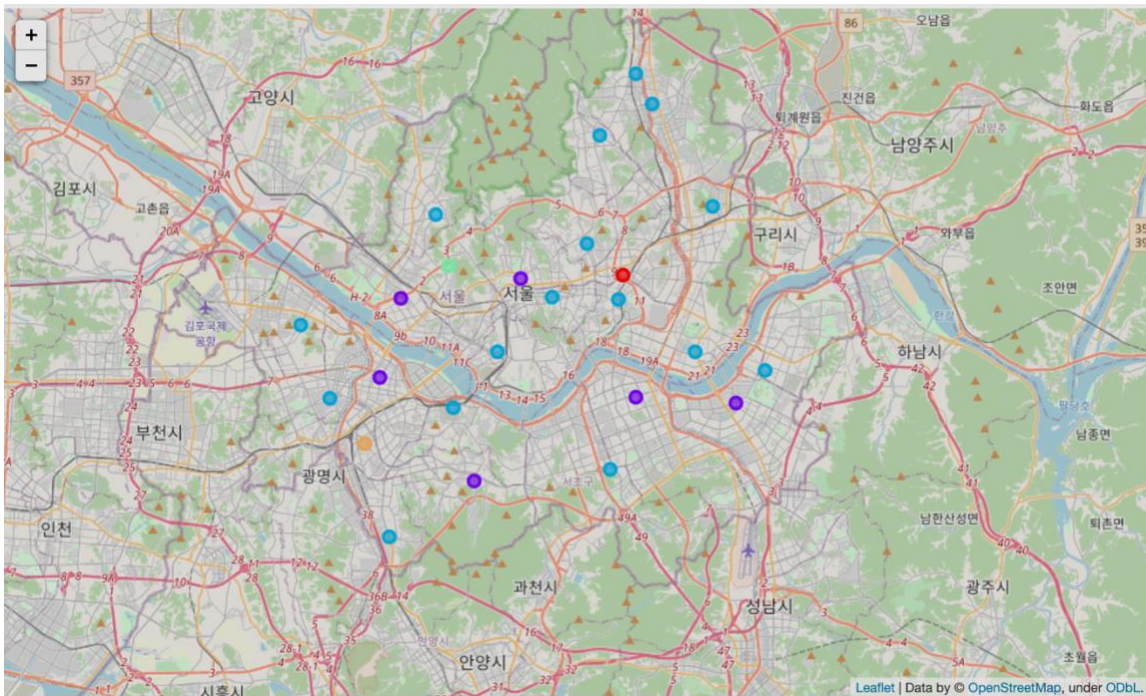
Next I used to k-means clustering to cluster the neighborhoods, then visualized these clusters using the Map() function in folium. I repeated this process with several different k values before deciding on an optimal one. For the next phase, I repeated this entire process for the city of Austin. K-means clustering was used here because it seemed to be the optimal method of effectively segmenting the unlabeled data that was acquired from the Foursquare API about the two cities into clusters so that conclusions could be drawn.

For the final phase of my analysis, I compared Seoul and Austin. I found the Neighborhood with the highest number of venues for Austin and Seoul, then plotted this information on a horizontal bar chart to illustrate the similarities and differences between the two cities. I used these charts to draw conclusions about the two cities, and if they indeed are similar enough for the purpose of international travel, as discussed in the introduction.

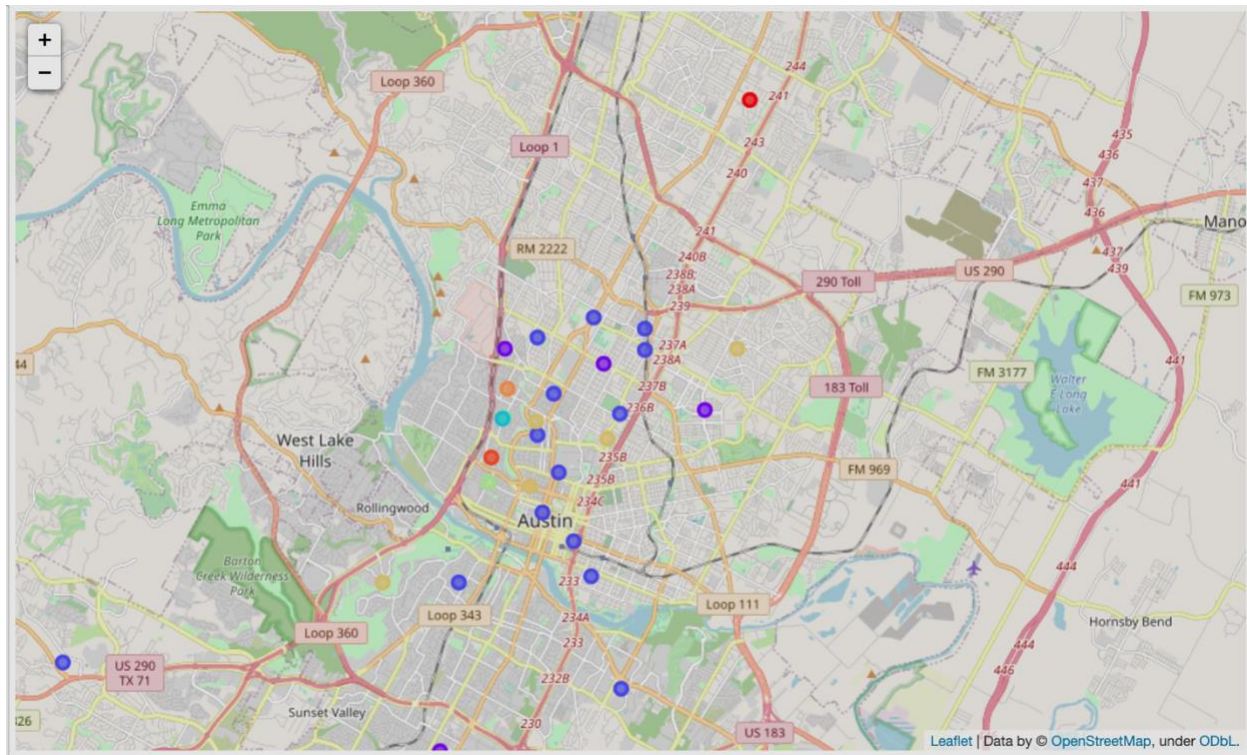
4. Results

The clustering maps for the two cities are as follows:

Seoul (k=5)

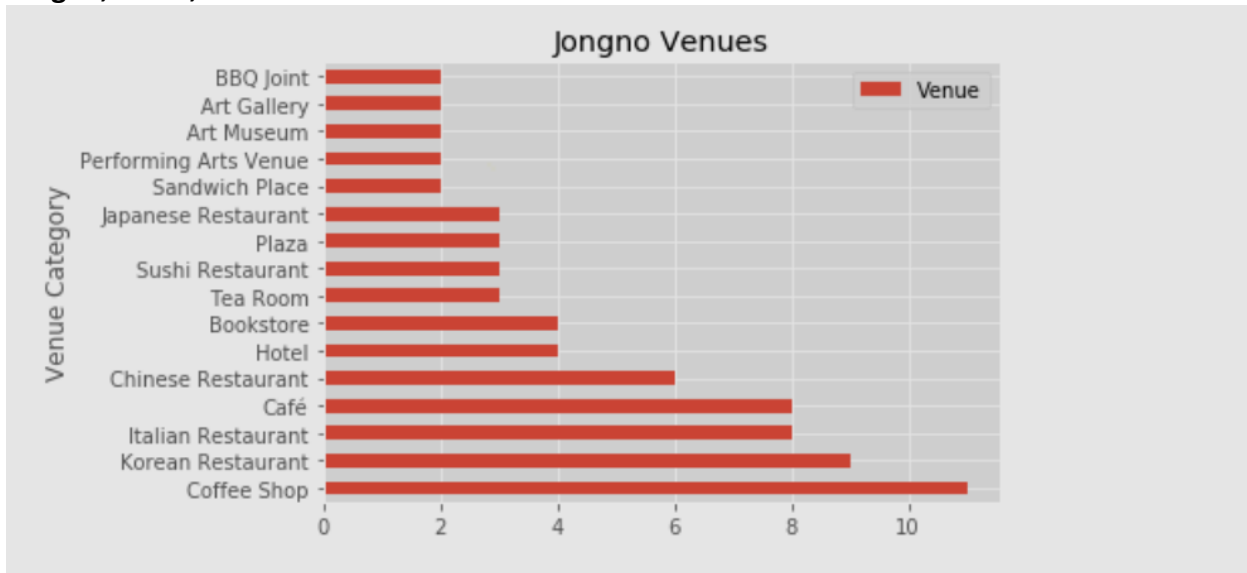


Austin (k=11)

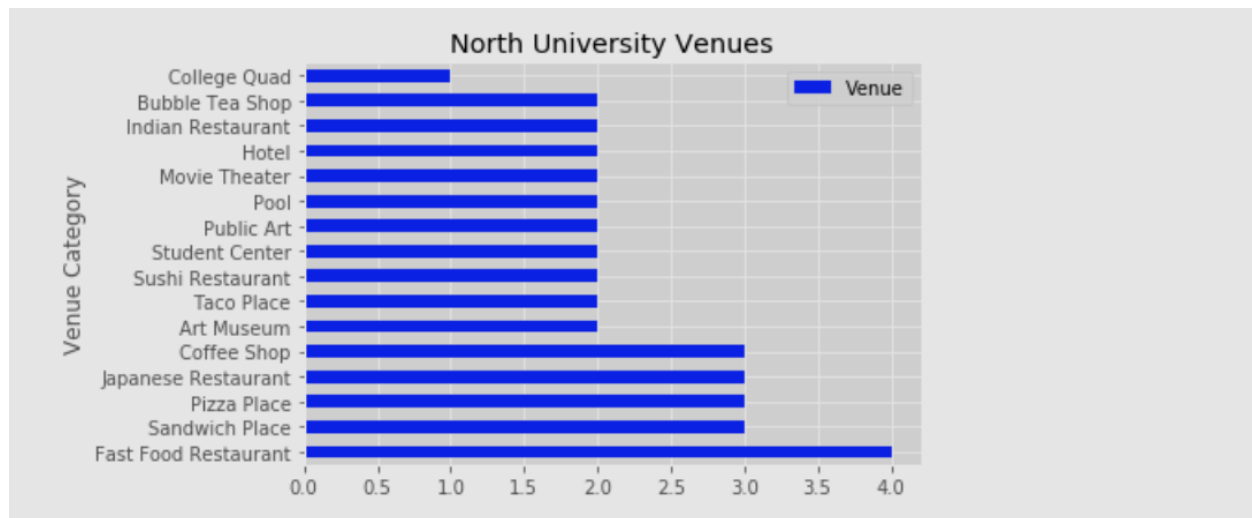


The barplots of the top 15 venue categories in the neighborhoods with the most venues in Seoul and Austin are as follows:

Jongno, Seoul, South Korea – 100 Venues



North University, Austin, Texas, USA – 73 Venues



5. Discussion

Considering my initial preposition that Austin and Seoul are similar, the results of this analysis are not entirely surprising. If we consider Jongno and North University as snapshots of their respective cities, we can see that they have similarities in having a lot of different food options in their variety of restaurants and cafes, while also have other attractions such as movie theaters and art venues. Further research on North University in particular reveals that it is situated at the University of Texas at Austin campus, which explains why there several attractions that are university-related in comparison to average metropolitan Jongno. What is surprising is the fact that despite Seoul as a whole clustering into a mere 5 clusters in comparison to Austin's 11, Jongno has more venues within its expanse than North University.

If this analysis is reiterated in the future, perhaps it would be better to incorporate different forms of machine learning and/or statistical analyses to fortify this information gathered here. Also, perhaps it would be beneficial to scrape the data using webscraping as opposed to manual compilation, which is what I did here. It would also be interesting to create a recommender system based on the types of venues an individual likes to visit, to be given an output of attractions to visit in either Seoul or Austin

6. Conclusion

Overall, this analysis suggests that Seoul is just as, if not even more interesting than Austin considering its venues are independent of a university campus, which can often pigeonhole attractions into being interesting to only a younger subset of people. As such, it seems that anyone, especially stateside, considering international travel should give both Seoul and Austin a visit for their rich variety of food and entertainment.

