

Analyzing the Cultural Beliefs based on Popular Venues

Esha Tulsian

August 16, 2020

1. Introduction

1.1 Background

With a total of 195 countries in the world currently, each country has its own culture that reflect its values. Just as food or clothes help define culture, a city's layout is planned beforehand to reflect cultural traditions and values. Each city's layout defines its purpose whether it is financial, entertainment, or port. These all depend on the demographics, geographic, and politics of the area. For instance, New York City developed towns like Chinatown and Little Italy that were developed based on the people who lived in the area. They spread their culture in that specific area and created a hub for their culture. To compare, a city from three different countries were chosen, such as New York City, United States; Toronto, Canada; and Seoul, South Korea. Each city, though all are financials hubs of their respective country, will be analyzed to see the prevalence of venue type allude to the demographics or beliefs of the country.

1.2 Business Problem

The three cities will be compared to each other and see how each represents their unique culture. In addition, the cities will be analyzed to see which venue type are most common.

1.3 Interest

This analysis will be helpful for any business that is opening in one of these cities. They will be able to see where the best place to open some businesses or what the general population of the city likes. Additionally, the general public will enjoy learning the comparison between the cities and their cultures.

2. Data Acquisition and Cleaning

2.1 Data sources

The postal codes, latitude, and longitude for New York City can be found [here](#). The postal codes of Toronto can be found [here](#). The geospatial data of Toronto can be found [here](#). The postal codes data for Seoul can be found [here](#). The geospatial data of Seoul were found using Geopy library. The rest of the data like venues and venue categories were found using Foursquare API.

2.2 Data cleaning and Features

The data sources for each location were combined to create one table so that each at least had postcode, latitude, and longitude data.

Focusing on the New York City data, a dataframe with features, such as borough, neighborhoods, latitude, and longitude. The dataframe had 5 unique boroughs and 306 unique neighborhoods.

Focusing on Toronto data, a similar dataframe was created with borough, neighborhood, latitude, and longitude features. This was a bit more difficult as data was taken from Wikipedia, so there were some boroughs and neighborhoods that were not assigned to a specific postcode. Those postcodes were removed. Additionally, there were multiple neighborhoods with the same postcode, so those were grouped up so that each postcode was seen once. After doing the cleaning, the dataframe had 10 unique boroughs and 103 unique postcodes. This dataframe so far only had the postcode, borough, and neighborhood. Next, the latitude and longitude of each postcode was added by merging a geospatial dataframe that was created and the postcode dataframe. The final dataframe had borough, neighborhood, postcode, latitude, and longitude features.

Finally, focusing on the Seoul data, the data was a bit different as Seoul has districts instead of boroughs. Wikipedia was scraped to get the postcode and district data. Initially, geocoder library was used to try to get latitude and longitude data for the districts; however, it took too long to run. So, instead the Geopy library was used and returned values of latitude and longitude for each district. These values were

added to the district dataframe. Creating a final dataframe with postcode, district, latitude, and longitude features. The final dataframe had 25 unique districts.

In addition to the features discussed above, data about venues, venue latitude, venue longitude, and venue category were looked at.

3. Methodology/Exploratory Data Analysis

3.1 New York City

After developing the final dataframe for the whole of New York City, all the 306 neighborhoods were plotted using the folium library (Figure 1).

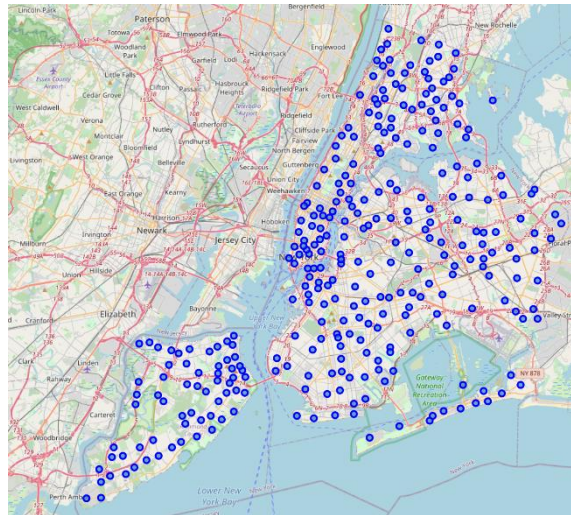


Figure 1 Map of New York City Neighborhoods

Using the Foursquare API, the venues for each neighborhood was found. The total number of venues that were given was 10032 venues with 421 unique venue categories. The top 10 venues for each neighborhood was found. Based on the top 10 venues and locations of neighborhoods, the neighborhoods were clustered into 5 clusters using k-means clustering.

3.3 Toronto

After developing the final dataframe for Toronto, all the 103 unique postcodes were plotted using the folium library (Figure 3).

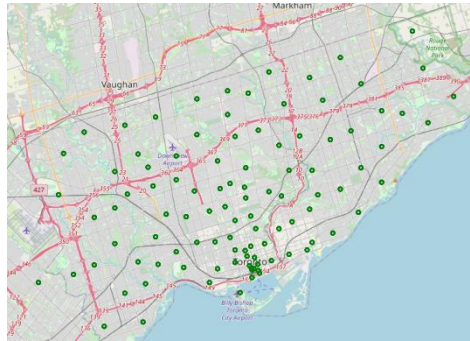


Figure 2 Map of Toronto Neighborhoods

Using the Foursquare API, the venues for each postcode was found. The total number of venues that were given was 2124 venues with 269 unique venue categories. The top 10 venues for each neighborhood was found. Based on the top 10 venues and location data, the postcodes were clustered into 5 clusters using k-means clustering.

3.4 Seoul

After developing the final dataframe for Seoul, all the 103 unique postcodes were plotted using the folium library (Figure 4).

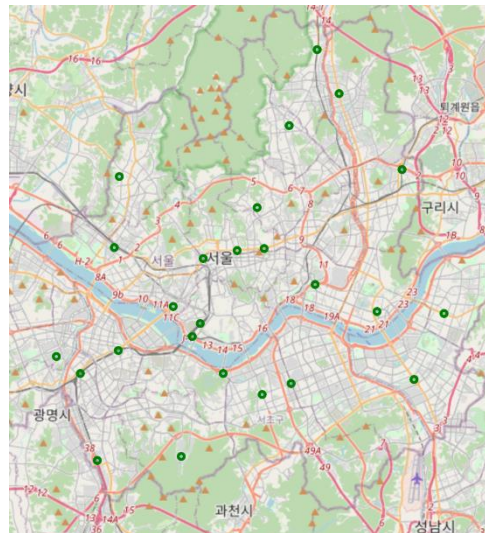


Figure 3 Map of Seoul Neighborhoods

Using the Foursquare API, the venues for each postcode was found. The total number of venues that were given was 791 venues with 124 unique venue categories. The top 10 venues for each neighborhood was found. Based on the top 10 venues and location data, the postcodes were clustered into 5 clusters using k-means clustering.

4. Results

4.2 New York City

First plot that was created was the top 20 venue categories overall in all of New York City (Figure 5).

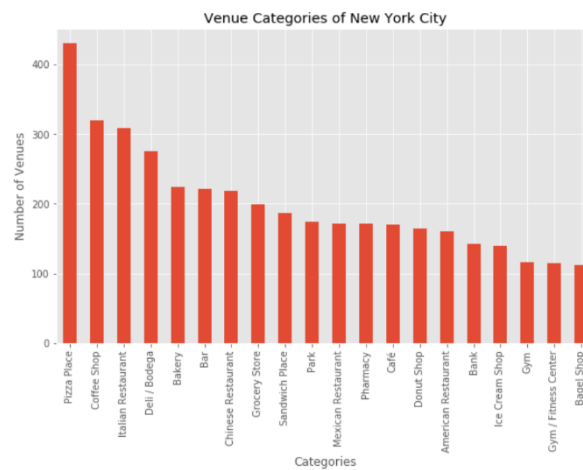


Figure 4 Top 20 Overall Venue Categories of all New York City Neighborhoods

As seen in Figure 5, the top 3 venues are pizza places, coffee shops, and Italian restaurants. Next, a plot of the Top 20 most popular venue category in the top 10 venues of each neighborhood (Figure 6).

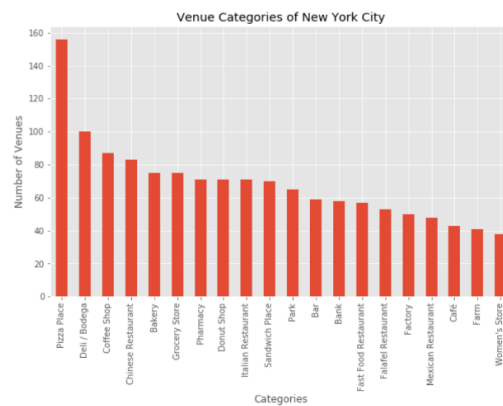


Figure 5 Top 20 Venue Categories after looking at the top 10 Venues of Each NYC Neighborhood

As seen in Figure 6, the top 3 categories are pizza places, deli/bodega, and coffee shops. Based on the top 10 venues and location, a cluster map was created (Figure 7).

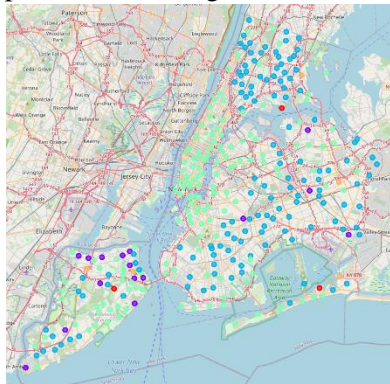


Figure 6 Cluster Map for New York City

The details of the cluster are found in Table 1. The clusters were determined on how similar the top 10 venues for each neighborhood were.

Table 1.

Cluster	0	1	2	3	4
Color	Red	Purple	Blue	Green	Orange
Neighborhood	3	16	133	148	5

4.2 Toronto

First plot that was created was the top 20 venue categories overall in all of New York City (Figure 11).

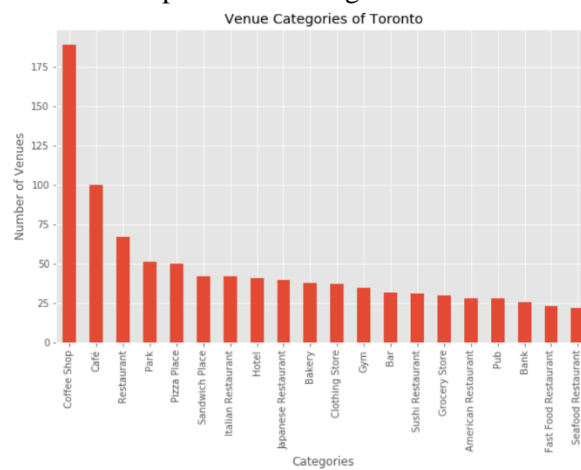


Figure 7 Top 20 Overall Venue Categories of all Toronto Neighborhoods

As seen in Figure 11, the top 3 venue categories in Toronto are coffee shops, cafes, and restaurants. Next, a plot of the Top 20 most popular venue category in the top 10 venues of each neighborhood (Figure 12).

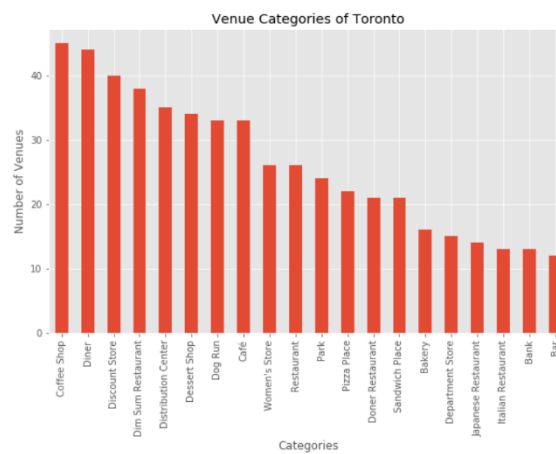


Figure 8 Top 20 Venue Categories after looking at the top 10 Venues of Each Toronto Neighborhood

As seen in Figure 12, the top 3 venue categories were coffee shops, diner, and discount store. Based on the top 10 venues and location, a cluster map was created (Figure 13).

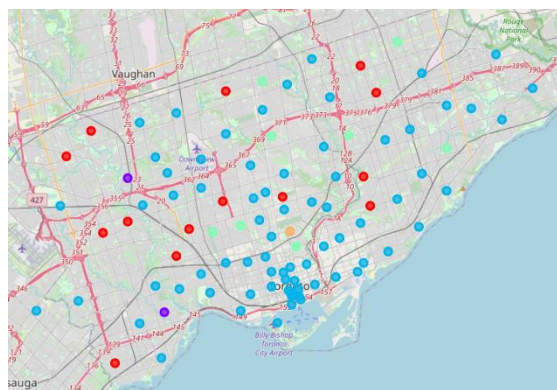


Figure 9 Cluster Map of Toronto

The details of the cluster are found in Table 3. The clusters were determined on how similar the top 10 venues for each neighborhood were.

Table 3

Cluster	0	1	2	3	4
Color	Red	Purple	Blue	Green	Orange
Neighborhood	14	2	74	8	1

4.3 Seoul

First plot that was created was the top 20 venue categories overall in all of Seoul (Figure 14).

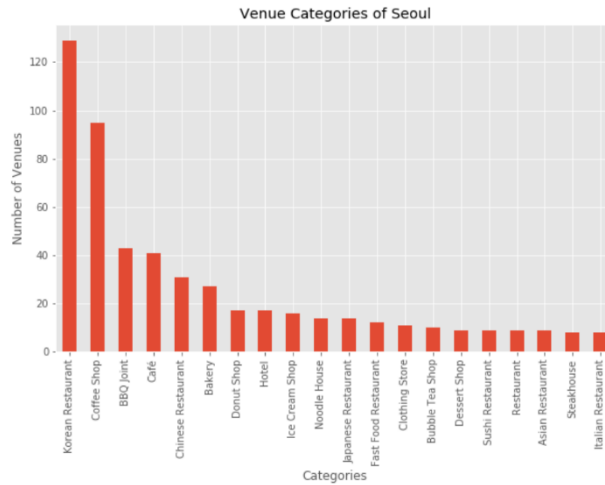


Figure 10 Top 20 Overall Venue Categories of all Seoul Neighborhoods

As seen in Figure 14, the top 3 venues are Korean restaurants, coffee shops, and BBQ joint. Next, a plot of the Top 20 most popular venue category in the top 10 venues of each neighborhood (Figure 15).

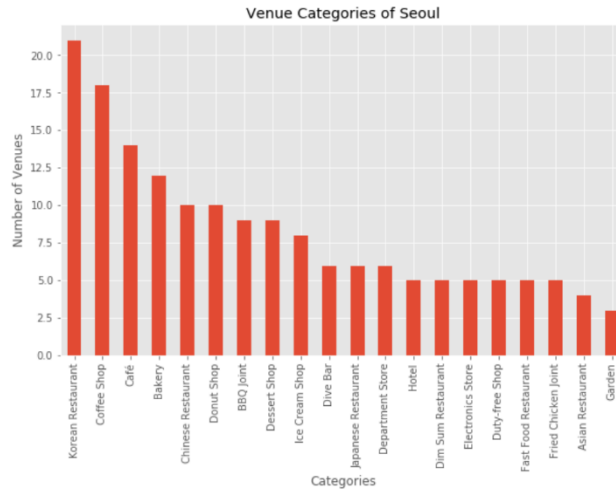


Figure 11 Top 20 Venue Categories after looking at the top 10 Venues of Each Seoul Neighborhood

As seen in Figure 15, the top 3 venues are Korean restaurants, coffee shops, and cafes. Based on the top 10 venues and location, a cluster map was created (Figure 16).

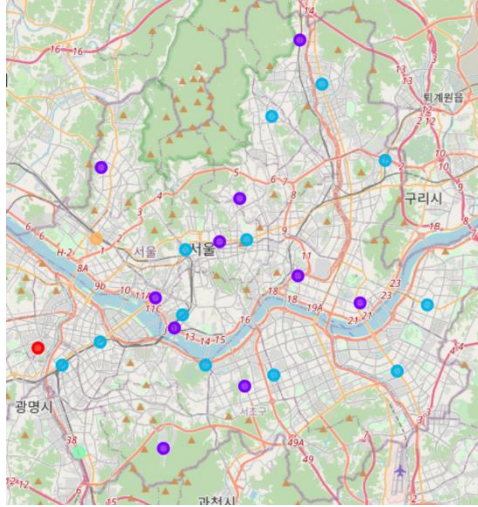


Figure 12 Cluster Map of Seoul

The details of the cluster are found in Table 4. The clusters were determined on how similar the top 10 venues for each neighborhood were.

Table 4

Cluster	0	1	2	3	4
Color	Red	Purple	Blue	Green	Orange
Neighborhood	1	10	12	1	1

5. Discussion

Coffee shops or cafes are frequented by all the cultures. This suggests that coffee shops are important in all 3 cultures. They provide for a small comfy place for customers to enjoy their coffee or pastries while doing their work. Though they all frequent them, it does not mean the décor of the coffee shop or items in the coffee shop are the same. Each culture will decorate in their own way. For instance, in NYC, they have very similar coffee shops like Starbucks. However, in Seoul, they have themed cafes based on popular cartoons or locations. Additionally, almost all the cities had food and drink places as their top 3. However, Toronto was the only city that had a non-food place, such as a discount store. This reveals that, in addition to food, shopping plays a big role in their culture. For NYC, it made sense that pizza, Italian restaurants, and deli/bodegas are frequented because NYC is known for its delicious pizza, pasta, and subs. Focusing

on Seoul, it made sense that there were more Korean restaurants as the general population is Korean. Korean food is known for having delicious BBQ, which is reflected by BBQ joints being frequented often as well.

6. Conclusion

In this project, the whole data science process was followed throughout. The steps were identifying the business problem, specifying and cleaning data required, performing exploratory data analysis, performing machine learning by clustering, and analyzing the results. To answer the business question, each city is different and the top venue categories reflect the culture. For instance, coffee shops are popular in all cities but decorations and menu items should be related to the culture. The findings in this report are important for businesses and for the general public to learn about cultural differences.