Exploring and Clustering Asian Restaurants in Austin

Author: Andersen Lin; Date: May 30th, 2020

1. Introduction

As the capital city of the state of Texas, Austin is the cultural and economic center of the Round Rock-Austin metropolitan area. Austin is home to almost 1 million residents estimated in 2019, with around 200 thousand new residents from 2010, it is one of the fastest-growing cities in the United States [1].

Being famous for its diversity, Austin is a city with people of various ethnic and cultural backgrounds. Among the fast-growing population in the city, the Asian community has been the fastest growing one in the country since 2000 [2].

As a result, Austin is one of the ideal places for investors to put their money in a restaurant concept in the 2020s, especially in Asian restaurants. However, when it comes to starting a business in an area, there are numerous factors to be taken into consideration.

The objective of this project is to provide intuitive visualization tools to assist in the decision-making process of opening an Asian restaurant in Austin by analyzing existing ones in neighborhoods from a data science perspective.

2. Data

To address the problem, below is a list of data we need for further analysis:

1. To present users with an interactive map, we need `folium` to render a map and allow us to put additional information
2. To localize Austin on the map, we need `geopy` to get the geographic coordinate of Austin
3. To assist in finding a location to open an Asian restaurant from a neighborhood basis. We need a GeoJSON file [3] that gives us the coordinates of boundaries of neighborhoods in Austin
4. To clearly display the number of existing Asian restaurants in each neighborhood. We need `folium` to draw choropleth maps
5. To get a list of existing Asian restaurants in Austin [4], we need `Foursquare API` to explore restaurants based on given locations

2.1 Get geographic coordinates of Austin for map initialization

By using geopy, the geographical coordinate of Austin is 30.2711286, -97.7436995.

2.2 Get GeoJSON with Austin neighborhoods central coordinates and boundaries

- Retrieved from a public dataset [3]
- Boundaries are used to plot a choropleth map that shows the number of Asian restaurants nearby for each neighborhood
- Central coordinates are used to add markers for each neighborhood with pop-up text on the map

Clean the GeoJSON data to extract neighborhood names and coordinates only.

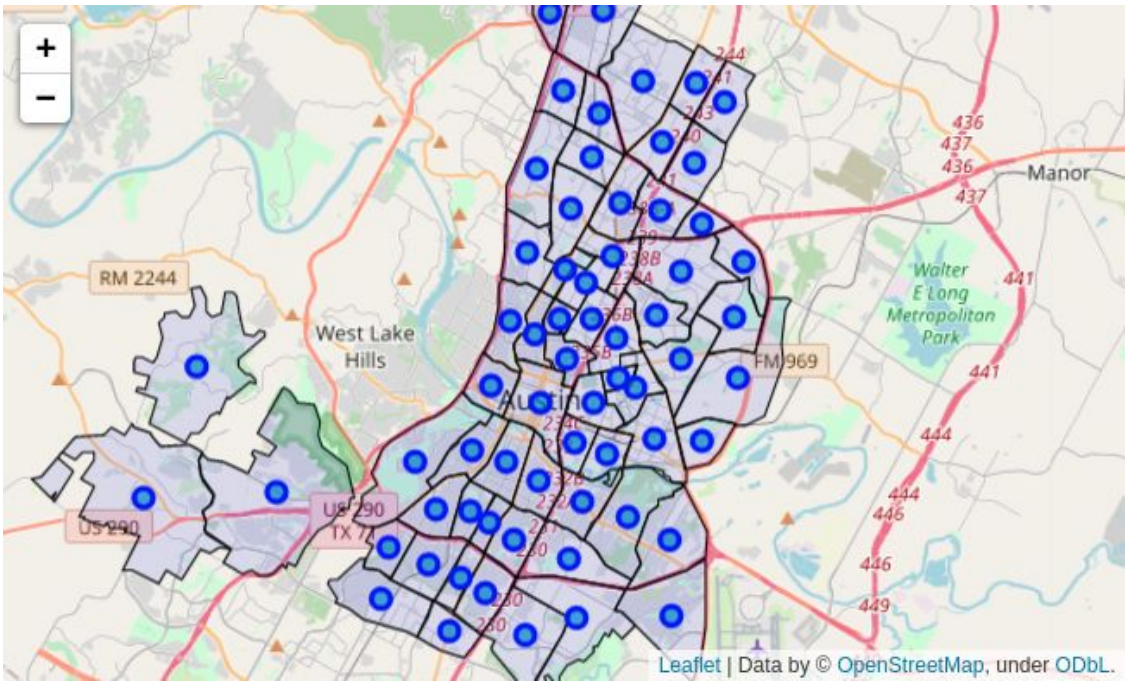
Part of the dataset extracted from a GeoJSON file [3] after filtering and cleaning:

	Neighborhood	Latitude	Longitude
0	Johnston Terrace	30.258532	-97.684605
1	Bouldin	30.251880	-97.755556
2	MLK 183	30.278141	-97.671725
3	Zilker	30.255278	-97.768350
4	Crestview	30.346569	-97.725053
5	Onion Creek	30.121934	-97.791569

According to the data from the GeoJSON file, there are 66 neighborhoods in Austin.

2.3 Render the map with cleaned data to ensure it is reliable

An example of a rendered map by folium with neighborhood boundaries and markers:



2.4 Return a table of Asian restaurants near each neighborhood

1. Call the FourSquare API using its explore feature
2. Query a list of Asian restaurant categories defined by FourSquare API
3. Set the radius of exploring to 1000 meters as the distance of the two closest neighborhoods are around 1000 meters
4. Set the number of returned venues to maximum (50 venues per call)

Store necessary values for analysis only which include venue names, venue location, and venue categories. Then, drop duplicate entries to avoid data overlapping when clustering in the modeling phase.

After cleaning the dataframe, there are 315 unique Asian restaurants near all the neighborhoods in Austin.

An example of part of the dataset returned by FourSquare API after merging and cleaning:

	Neighborhood	PostalCode Latitude	PostalCode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bouldin	30.251880	-97.755556	Uchi	30.257594	-97.759891	Japanese Restaurant
1	Bouldin	30.251880	-97.755556	Lucky Robot Japanese Kitchen	30.250943	-97.749165	Japanese Restaurant
2	Bouldin	30.251880	-97.755556	Soto - South Lamar	30.255706	-97.761870	Japanese Restaurant
3	Bouldin	30.251880	-97.755556	Otoko	30.247678	-97.749998	Sushi Restaurant
4	Bouldin	30.251880	-97.755556	Ramen Tatsu-ya	30.253895	-97.763134	Ramen Restaurant
5	Bouldin	30.251880	-97.755556	Dawa Sushi Kimchi	30.251554	-97.753995	Sushi Restaurant

Now we have collected all the data we need for the following phases.

3. Methodology

Now that we have collected, initially understood, and cleaned the data required. It is time to analyze and prepare data for visualization and modeling.

So far, we have initially formatted the following data:

- Dataframe df_neiaus contains names of Austin neighborhoods and their geographic coordinates
- Table asian_rest_austin.csv contains neighborhoods which have Asian restaurants nearby as well as returned values from FourSquare API

The purpose of using the data:

- Display the number of nearby Asian restaurants for each neighborhood on a choropleth map
- Mark neighborhoods on the map with clusters to reveal insights to categories of Asian restaurants

3.1 Data preparation

To display Asian restaurant density on a choropleth map. Create a dataframe arnum_asu with a sum of Asian restaurants near each neighborhood:

	Neighborhood	Number of Asian Restaurants
0	Bouldin	15
1	Brentwood	6
2	Central East Austin	9
3	Chestnut	1
4	Coronado Hills	1
5	Crestview	3

Then, based on returned data from FourSquare API, there are 18 unique Asian restaurant categories near neighborhoods in Austin.

To reveal characteristics of nearby Asian restaurants, compute the occurrence frequency of each category of Asian restaurants. This is done by first get dummies for each Asian restaurant category and then take an average on occurrence.

Below is part of the dataframe with the frequency of each Asian restaurant category:

	Neighborhood	Asian Restaurant	Bubble Tea Shop	Cantonese Restaurant	Chinese Restaurant	Dim Sum Restaurant	Filipino Restaurant	Himalayan Restaurant	Japanese Curry Restaurant	Japanese Restaurant	Korean Restaurant	Noodle House
0	Bouldin	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.200000	0.133333	0.066667
1	Brentwood	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.166667	0.166667	0.000000
2	Central East Austin	0.0	0.0	0.0	0.111111	0.0	0.0	0.0	0.0	0.111111	0.000000	0.111111
3	Chestnut	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.000000	1.000000
4	Coronado Hills	0.0	0.0	0.0	1.000000	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000

Also by looking at the row number of this dataframe, I noticed that 49 out of 66 neighborhoods in Austin have Asian restaurants nearby. We will keep this information in mind.

Next, I ranked occurrence frequencies and encode them for clustering based on the top 10 common Asian restaurant categories.

Part of the dataframe up to this step:

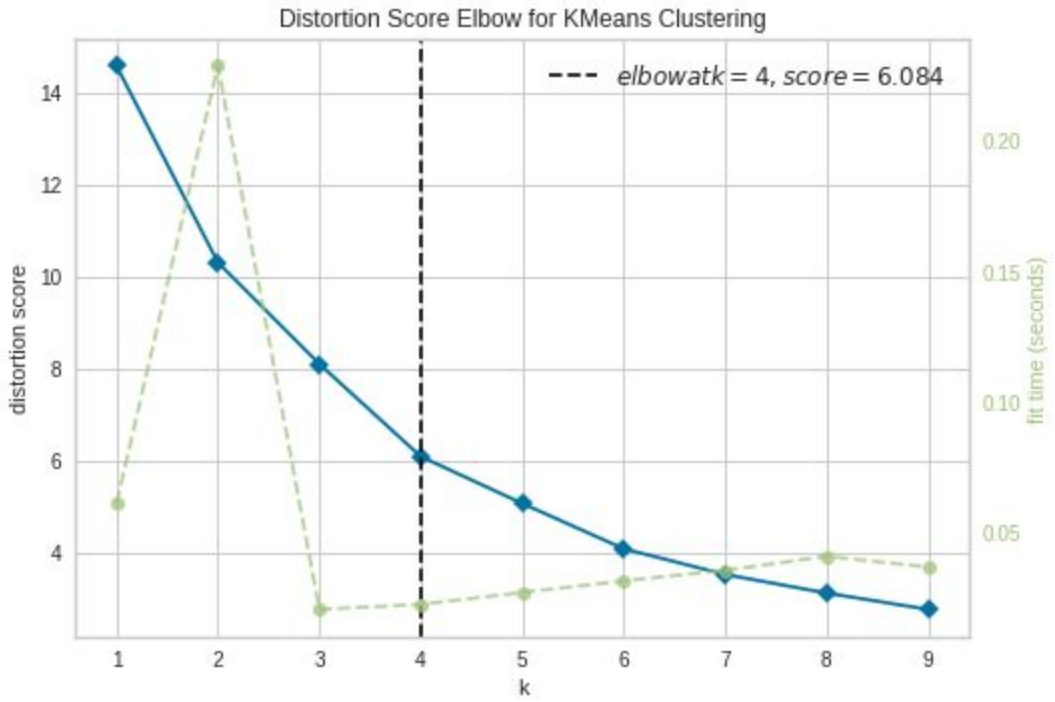
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Bouldin	Vietnamese Restaurant	Thai Restaurant	Japanese Restaurant
1	Brentwood	Vietnamese Restaurant	Thai Restaurant	Korean Restaurant
2	Central East Austin	Thai Restaurant	Japanese Restaurant	Chinese Restaurant
3	Chestnut	Noodle House	Vietnamese Restaurant	Japanese Curry Restaurant
4	Coronado Hills	Chinese Restaurant	Vietnamese Restaurant	Thai Restaurant

Given that the data we have is well-defined and are separated without overlapping, I used K-Mean clustering to robustly segment neighborhoods based on their 10 most common Asian restaurant categories.

3.2 Modeling and evaluation

Before moving on to clustering, I first visualized the elbow to determine an optimum number of clusters for modeling. I set the range of K from 1 to 20.

The result of the elbow visualization can be seen below:



After fitting the K-Means clustering model to the dataset, adding clustered labels as a new column back. We now have a new dataframe with their top 10 common Asian restaurant categories and their cluster labels based.

Part of the dataframe can be seen below, notice that I added back the neighborhoods without Asian restaurants nearby which did not participate in clustering (shown as NaN values for now). We will utilize this dataframe to visualize the clustering result and discuss observations later on.

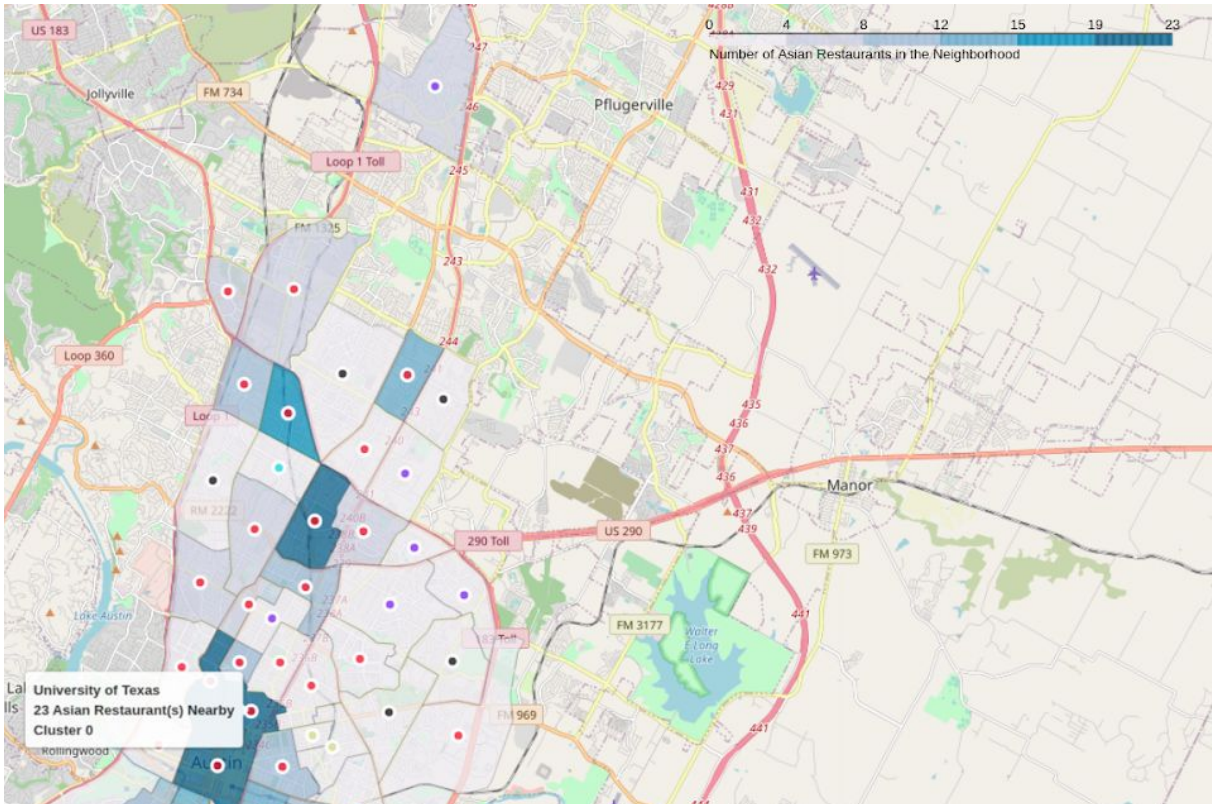
	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Johnston Terrace	30.258532	-97.684605	NaN	NaN	NaN	NaN
1	Bouldin	30.251880	-97.755556	0.0	Vietnamese Restaurant	Thai Restaurant	Japanese Restaurant
2	MLK 183	30.278141	-97.671725	0.0	Vietnamese Restaurant	Thai Restaurant	Bubble Tea Shop
3	Zilker	30.255278	-97.768350	0.0	Thai Restaurant	Japanese Restaurant	Ramen Restaurant
4	Crestview	30.346569	-97.725053	2.0	Japanese Restaurant	Korean Restaurant	Japanese Curry Restaurant

4. Results

4.1 Render a choropleth map with clustered markers that describes:

- The distribution of Asian restaurants among neighborhoods with different coloring
- The number of nearby Asian restaurants in each neighborhood
- The cluster of each neighborhood which indicates similar Asian restaurant categories

Part of the rendered choropleth map with markers indicate clusters and pop-up texts can be seen below:



4.2 Make a list of clustered neighborhoods

By simply observing the 3 most common venues for neighborhoods in the same cluster, we can see they have similar Asian restaurant categories in common. In the Discussion section, I will visualize based on the results to reveal this observation.

An example can be seen below, it is obvious that neighborhoods within cluster 1 all have Chinese restaurant(s) as their most common Asian restaurant nearby. However, I will visualize based on these data frames to prove my assumption.

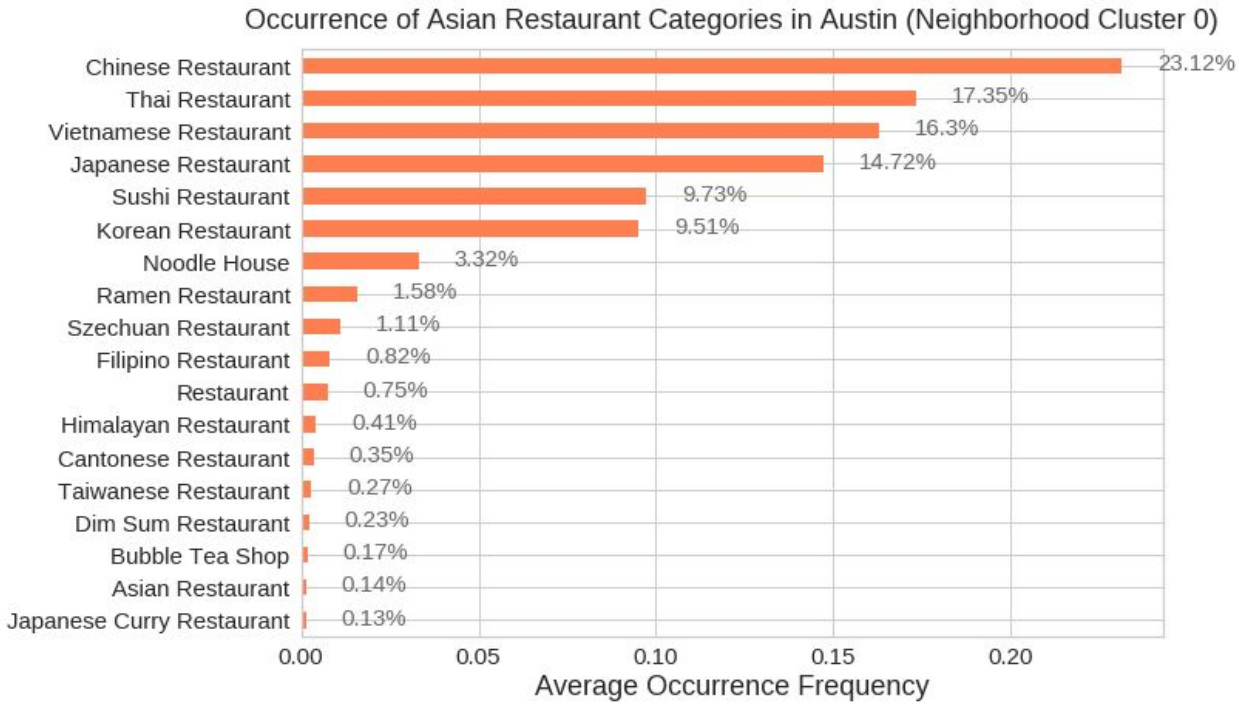
	Neighborhood	Cluster Labels	Number of Asian Restaurants	1st Most Common Venue
9	Wells Branch	1	4	Chinese Restaurant
37	Heritage Hills	1	1	Chinese Restaurant
43	Pleasant Valley	1	1	Chinese Restaurant
48	Hyde Park	1	1	Chinese Restaurant
52	University Hills	1	1	Chinese Restaurant
53	Dawson	1	3	Chinese Restaurant

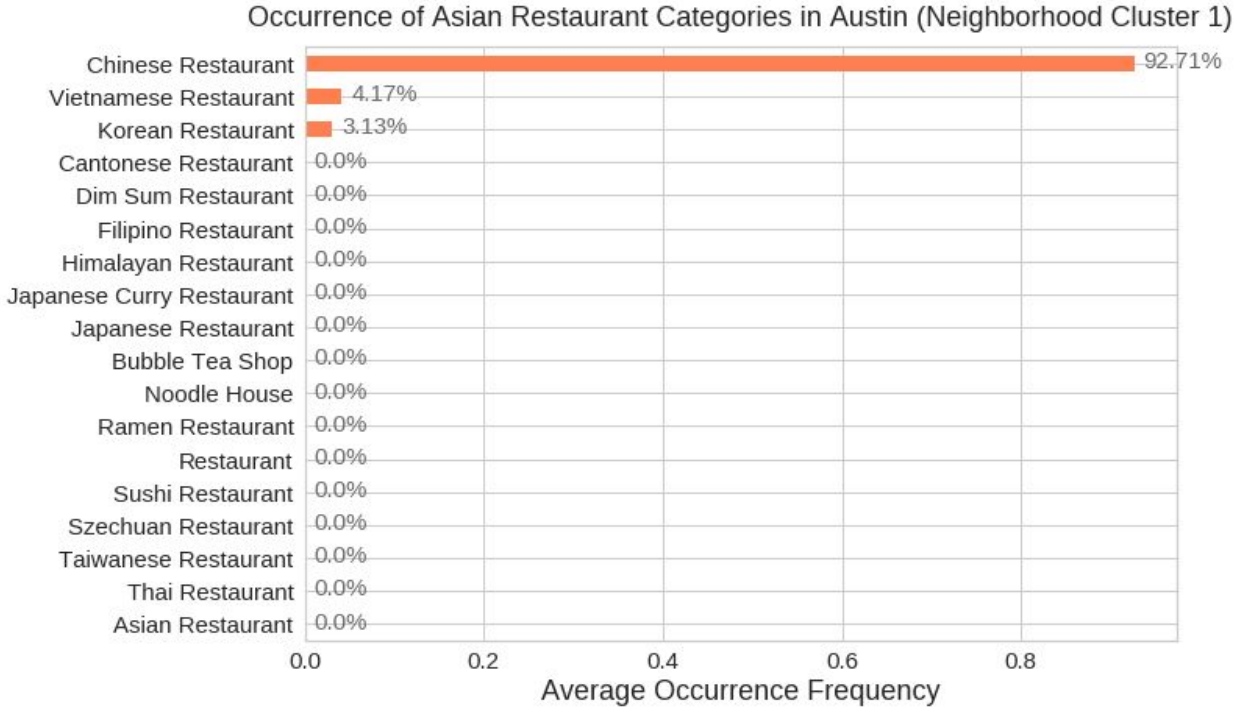
To make the visualization more detailed, I utilized the dataframe with the average frequency of each Asian restaurant category of each neighborhood for visualization. Part of the dataframe can be seen below:

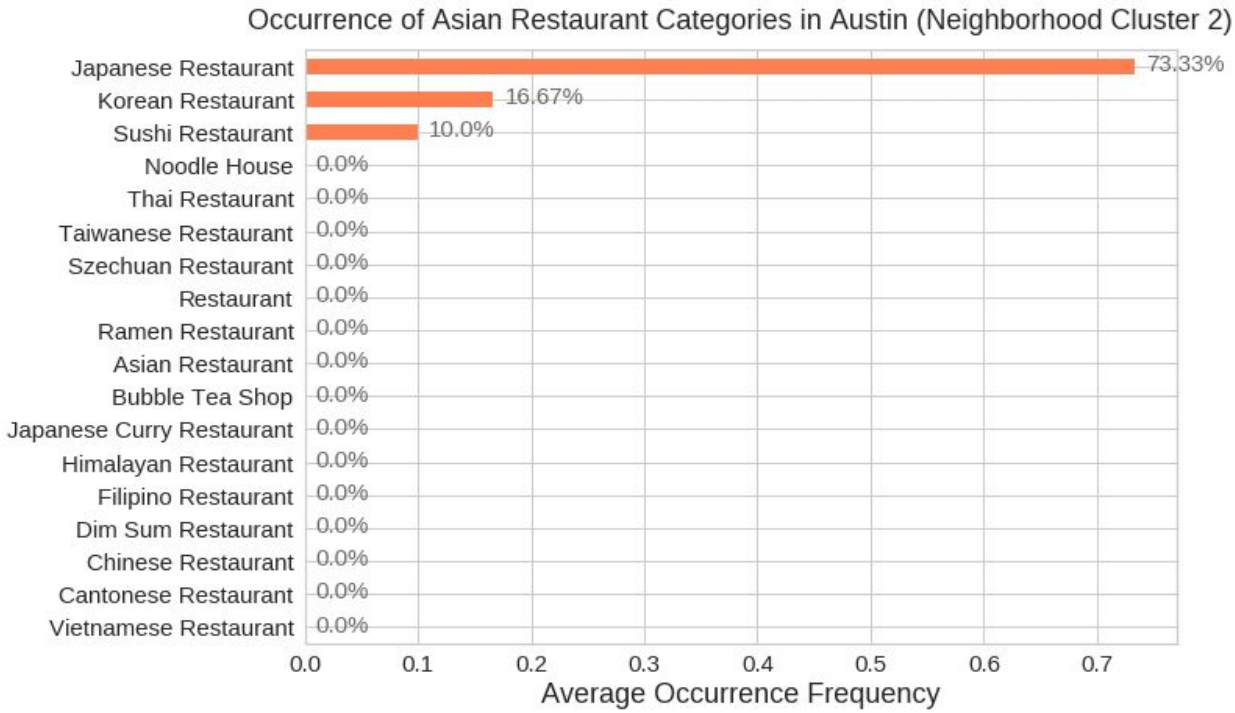
	Frequency
Chinese Restaurant	0.231171
Thai Restaurant	0.173453
Vietnamese Restaurant	0.162999
Japanese Restaurant	0.147161
Sushi Restaurant	0.0973161
Korean Restaurant	0.0951052
Noodle House	0.0332274
Ramen Restaurant	0.0158079
Szechuan Restaurant	0.0110827
Filipino Restaurant	0.00816993
Restaurant	0.00746965
Himalayan Restaurant	0.00407436
Cantonese Restaurant	0.00354122
Taiwanese Restaurant	0.00274936
Dim Sum Restaurant	0.00226244
Bubble Tea Shop	0.0017301
Asian Restaurant	0.00140056
Japanese Curry Restaurant	0.00127877

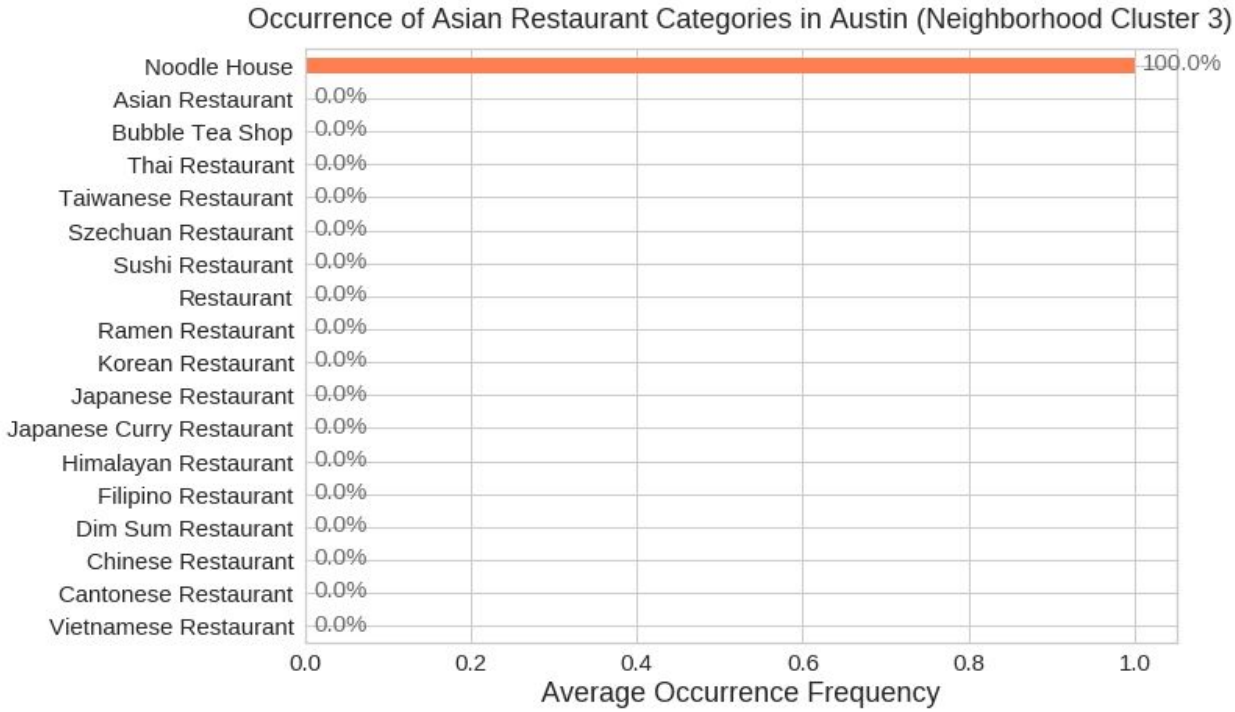
5. Discussion

Bar plots showing how Asian restaurant categories distributed in each neighborhood cluster are below:


Occurrence of Asian Restaurant Categories in Austin (Neighborhood Cluster 0)
Chinese Restaurant
Thai Restaurant
Vietnamese Restaurant
Japanese Restaurant
Sushi Restaurant
Korean Restaurant
Noodle House
Ramen Restaurant
Szechuan Restaurant
Filipino Restaurant
Restaurant
Himalayan Restaurant
Cantonese Restaurant
Taiwanese Restaurant
Dim Sum Restaurant
Bubble Tea Shop
Asian Restaurant
Japanese Curry Restaurant
23.12%
17.35%
16.3%
14.72%
9.73%
9.51%
3.32%
1.58%
1.11%
0.82%
0.75%
0.41%
0.35%
0.27%
0.23%
0.17%
0.14%
0.13%
0.00
0.05
0.10
0.15
0.20
Average Occurrence Frequency


Occurrence of Asian Restaurant Categories in Austin (Neighborhood Cluster 1)
Chinese Restaurant
Vietnamese Restaurant
Korean Restaurant
Cantonese Restaurant
Dim Sum Restaurant
Filipino Restaurant
Himalayan Restaurant
Japanese Curry Restaurant
Japanese Restaurant
Bubble Tea Shop
Noodle House
Ramen Restaurant
Restaurant
Sushi Restaurant
Szechuan Restaurant
Taiwanese Restaurant
Thai Restaurant
Asian Restaurant
92.71%
4.17%
3.13%
0.0%
0.0
0.2
0.4
0.6
0.8
Average Occurrence Frequency





With the above bar plots, we can now name out clusters.

- Cluster 0: Mainstream and Diverse Asian Restaurants
- Cluster 1: Chinese Restaurants
- Cluster 2: Japanese Restaurants
- Cluster 3: Noodle House

Based on the map and visualizations above, we have the following observations:

- Almost half of neighborhoods (34 out of 66) in Austin have relatively diverse Asian restaurant categories nearby (within 1000 meters from neighborhood's central location)
- Around a quarter of neighborhoods (15 out of 66) in Austin have relatively identical Asian restaurant categories (within 1000 meters)
- Around a quarter of neighborhoods (17 out of 66) in Austin have no Asian restaurants nearby (within 1000 meters)
- Asian restaurants in Austin are mainly distributed along Interstate 35 which is the highway that cuts through the heart of downtown Austin
- Mainstream Asian restaurant categories in Austin are: Vietnamese, Thai, Chinese, Japanese, and Korean
- Regardless of clusters, the frequency of a neighborhood with Chinese restaurants nearby is dominant

6. Conclusion

Investors can decide where to open Asian restaurants based on the distribution, density, and the type of existing Asian restaurants in Austin neighborhoods. Recommendations can be made with the help of observations from the results. As mentioned at the beginning of this report, there are many factors to consider when opening a business. Observations from this report can only assist in giving basic information on a general perspective; further research can be done using data science from a different perspective.

Several areas for future improvement:

- Display the Top 3 Asian restaurants as pop-up text on the marker of each neighborhood (by requesting venues ratings based on venue IDs on FourSquare API)
- Explore other types of diverse restaurants, i.e., Indian, middle eastern, Turkish, and etc.
- Explore and analyze neighborhoods in Round Rock-Austin metropolitan area
- Consider other factors that may determine a new restaurant's location and category, i.e., population density, crime rate. folium map is a versatile tool that can be used to overlay multiple layers of GeoJSON data for visualization

7. References

[1]. Austin, Texas, Wikipedia. Retrieved from: https://en.wikipedia.org/wiki/Austin,_Texas
[2]. The Asian Community in Austin: a Demographic Snapshot, The City of Austin. Retrieved from: https://austintexas.gov/
[3]. Austin Travis TX US Neighborhoods, OpenDataSoft. Retrieved from: https://public.opendatasoft.com/explore/dataset/zillow-neighborhoods/export/?refine.state=TX&refine.county=Travis&refine.city=Austin
[4]. Queried Asian Restaurant Recommendations within 1000 Meters of Each Neighborhood in Austin, FourSquare API. Retrieved from: https://developer.foursquare.com/docs/api-reference/venues/explore/