

Exploring and Clustering Asian Restaurants in Austin

...

Course Project of Applied Data Science Capstone by IBM on Coursera

Developed by: Andersen Lin

Date: May 30th, 2020

Introduction/Background - Austin and Asian Community

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].

Business Understanding - Where to Open an Asian Restaurant

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.

Data Requirements - Geospatial and Venue Data

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

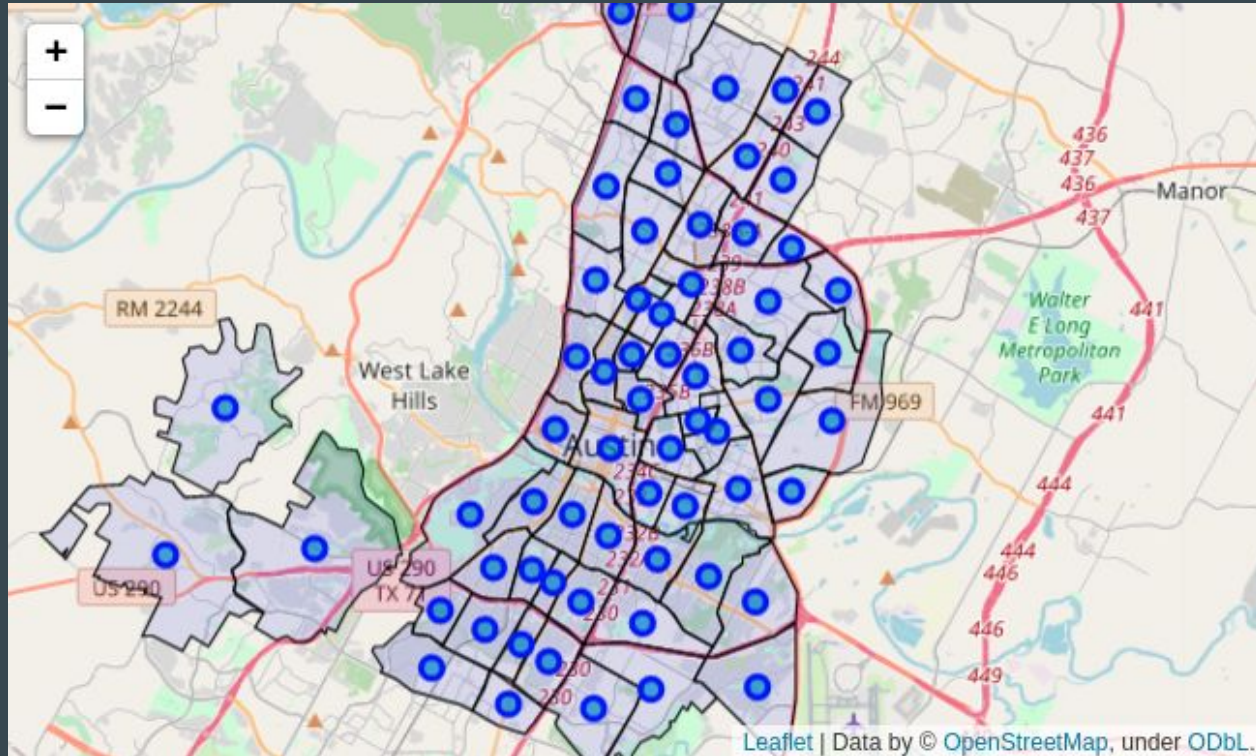
Data Collection - Neighborhoods and Coordinates

1. By using geopy, the geographical coordinate of Austin is 30.2711286, -97.7436995.
2. Get GeoJSON with Austin neighborhoods central coordinates and boundaries
 - a. Retrieved from a public dataset [3]
 - b. Boundaries are used to plot a choropleth map that shows the number of Asian restaurants nearby for each neighborhood
 - c. Central coordinates are used to add markers for each neighborhood with pop-up text on the map

	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

Data Collection - Render a Demo Map Using Folium

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



Data Collection - Explore Venues with FourSquare API

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)

Venue	Venue Latitude	Venue Longitude	Venue Category
Uchi	30.257594	-97.759891	Japanese Restaurant
Lucky Robot Japanese Kitchen	30.250943	-97.749165	Japanese Restaurant
Soto - South Lamar	30.255706	-97.761870	Japanese Restaurant
Otoko	30.247678	-97.749998	Sushi Restaurant
Ramen Tatsu-ya	30.253895	-97.763134	Ramen Restaurant
Dawa Sushi Kimchi	30.251554	-97.753995	Sushi Restaurant

Data Understanding

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

Data Preparation - For Choropleth Map

To display Asian restaurant density on a choropleth map. Create a dataframe `ar_num_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

Data Preparation - For Clustering Neighborhoods

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.

	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

Data Preparation - For Bar Plots Showing Distribution

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

	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

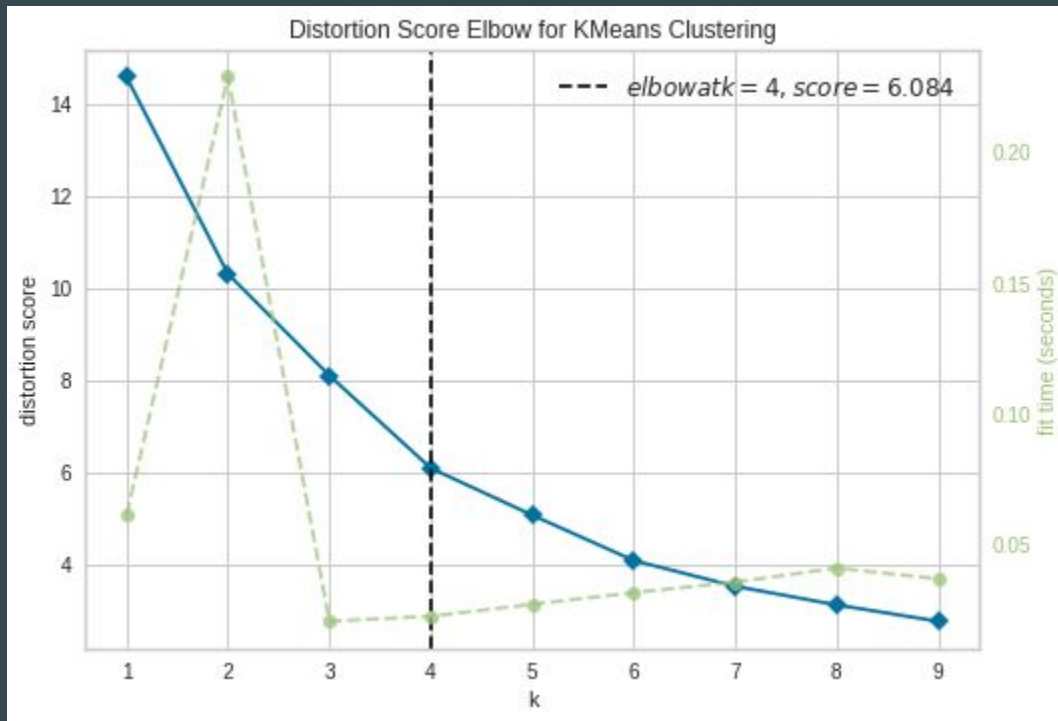
Modeling - K-Means Clustering

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.

	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

Evaluation - Run the Elbow Method to Yield an Optimal K

I visualized the elbow to determine an optimum number of clusters for modeling. I set the range of K from 1 to 20, with returned K being 4 as the optimal number of clusters.

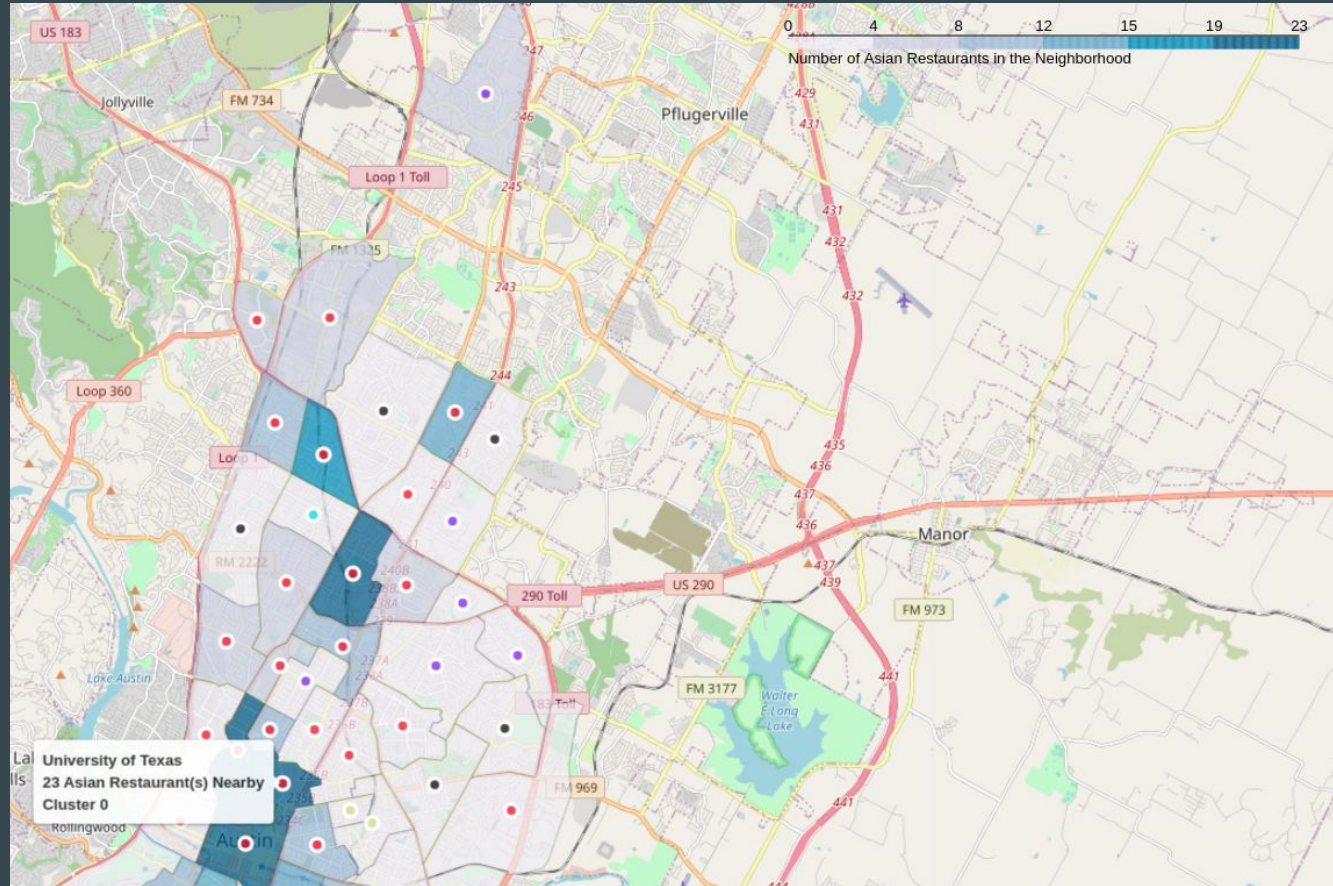


Deployment - Map-based Intuitive Visualization Tool

Render a choropleth map with clustered markers that describes:

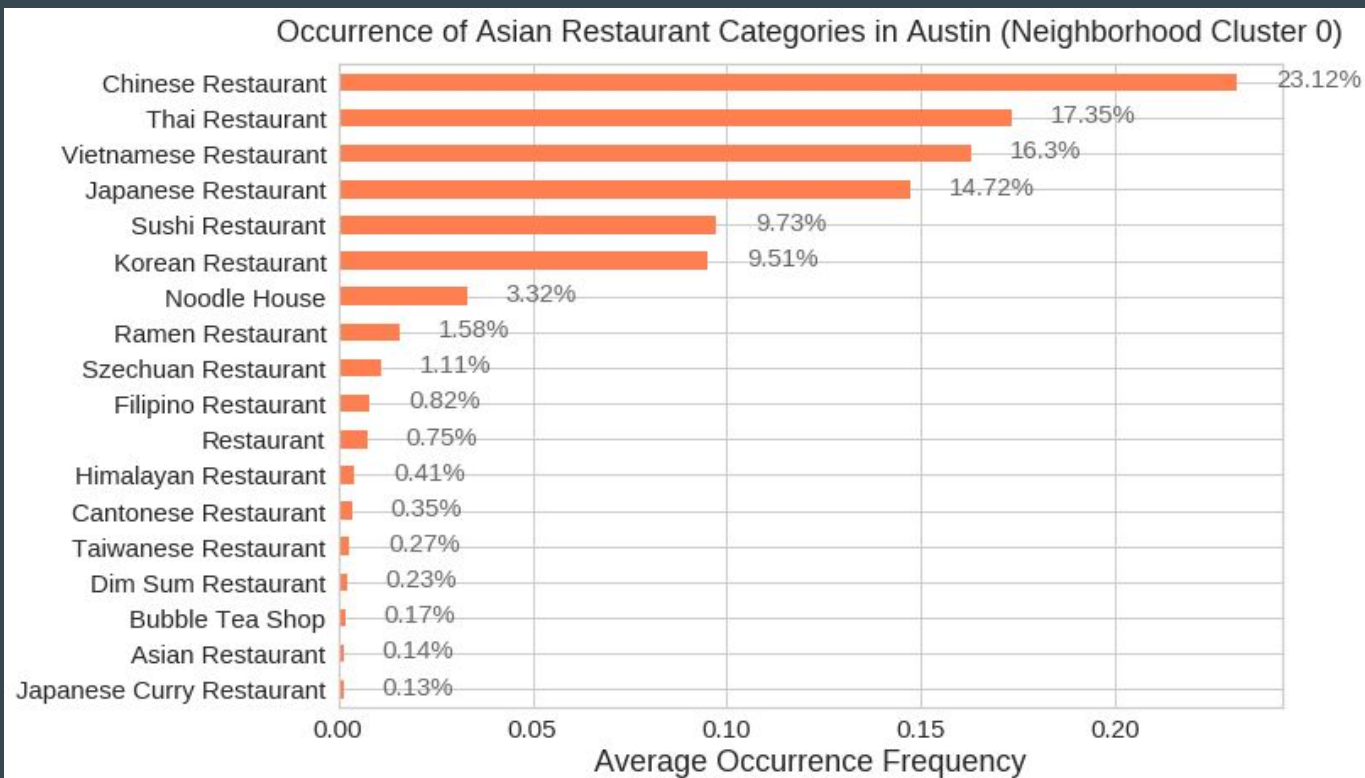
- The distribution of Asian restaurants among neighborhoods with different coloring
- The number of nearby Asian restaurants in each neighborhood
- The cluster of each neighborhood which indicates similar Asian restaurant categories

Deployment - Map-based Intuitive Visualization Tool



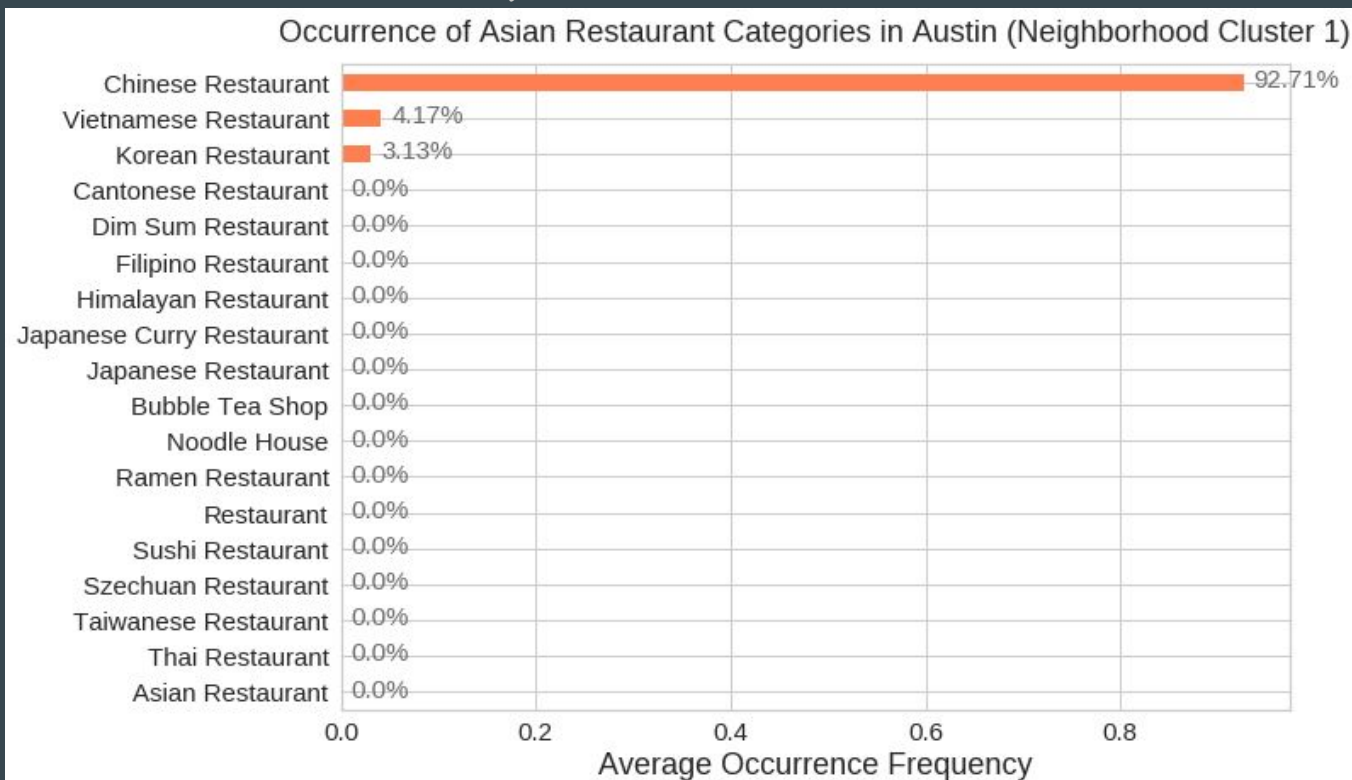
Deployment - Insights into Clustered Neighborhoods

Bar plots showing how Asian restaurant categories distributed in each neighborhood cluster are below: (Cluster 0: Mainstream and Diverse Asian Restaurants)



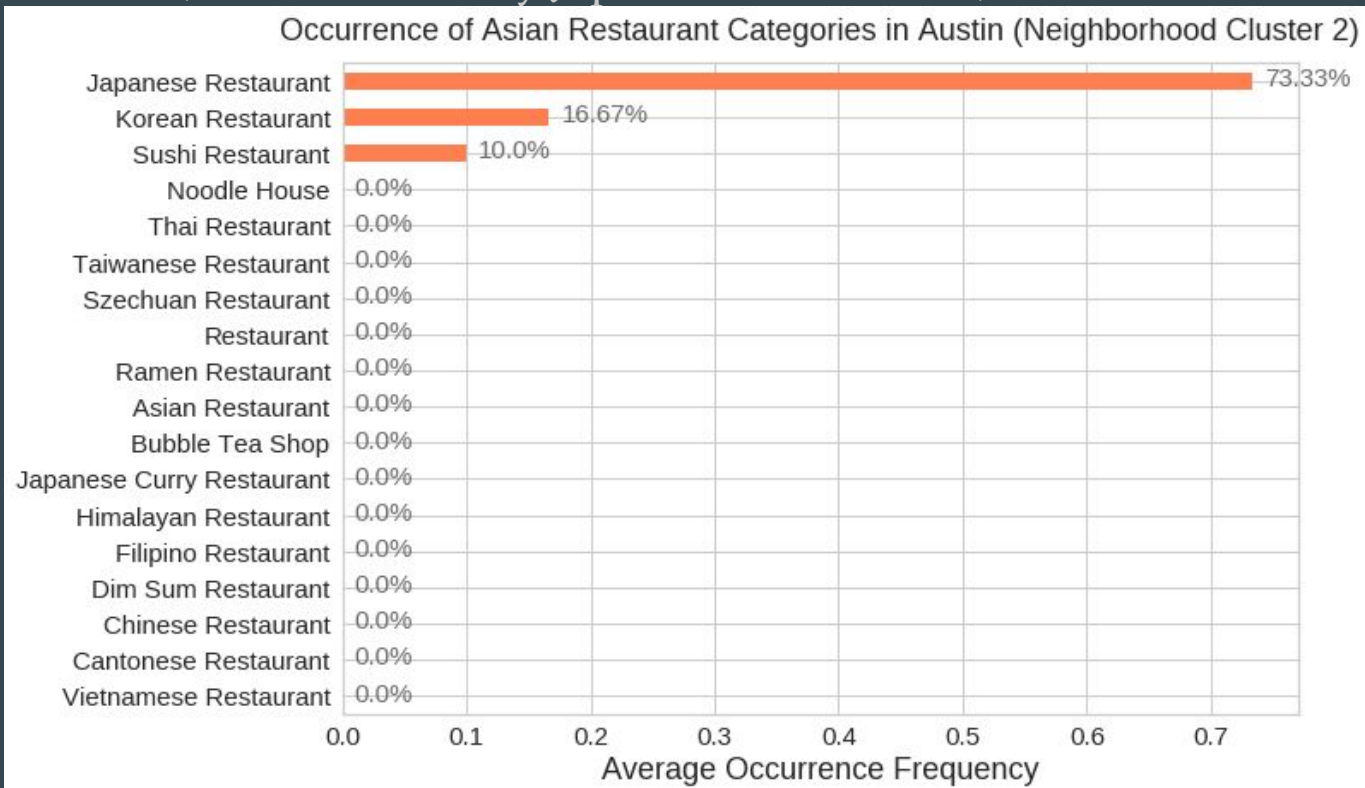
Deployment - Insights into Clustered Neighborhoods

Bar plots showing how Asian restaurant categories distributed in each neighborhood cluster are below: (Cluster 1: Mainly Chinese Restaurants)



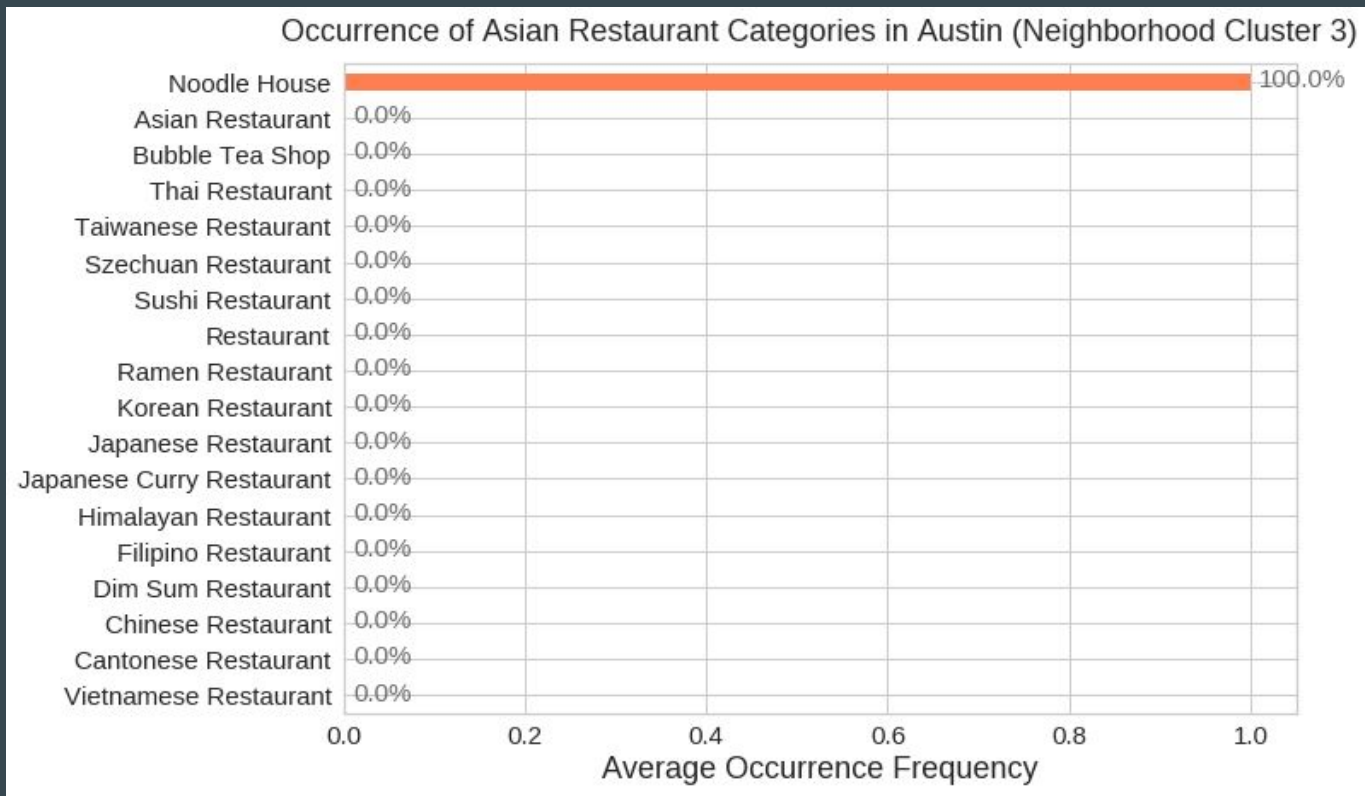
Deployment - Insights into Clustered Neighborhoods

Bar plots showing how Asian restaurant categories distributed in each neighborhood cluster are below: (Cluster 2: Mainly Japanese Restaurants)



Deployment - Insights into Clustered Neighborhoods

Bar plots showing how Asian restaurant categories distributed in each neighborhood cluster are below: (Cluster 3: Noodle House)



Deployment - Observations from Results

- 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

Feedback - 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.

Feedback - Improvements

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

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/>