Battle of Neighborhoods – Exploring Restaurants in Columbus

1. Introduction

1.1 Background

According to the United States Census Bureau, Columbus becomes the14th largest city in the U.S. and is the second-largest city in the Midwest. Since the 2010 census, Columbus has grown 9.3 percent, from 787,033 residents in 2010 to 860,090 in 2016. Between 2011 and 2016, private employment in Ohio increased by 450,000 jobs, far outperforming the regional average of 261,000 jobs. City officials say they expect more than a million people to move to Central Ohio over the course of the next 20 to 25 years. The rapid growth presents big opportunities for investors from all areas. Along with the population expansion, Columbus is rapidly gaining a reputation as one of America's up-and-coming culinary destinations. Its restaurant scene is creating a buzz with its creativity, diversity and one-of-a-kind experiences. Food & Wine recognized Columbus as one of the best culinary destinations in the U.S.

An investor from Japan would like to take advantage of the rapid growth of Columbus and invest in the restaurant industry, specifically Japanese restaurants, in Columbus and its surrounding cities. He would like to make data-driven decisions on the new locations that are most suitable for the new restaurants. This project will examine and visualize current restaurants in Franklin County.

1.2 Questions to be addressed:

1. Which is the distribution of restaurants in each city in Franklin County?
2. Which cities have a potential market for Japanese restaurants?

2. Data Description

2.1 Data Description and Analytical plan

For this project we will use the following data:

1. Ohio data that contains lists of cities and coordinates
   - A list of cities: https://en.wikipedia.org/wiki/List_of_cities_in_Ohio
   - Coordinates: https://raw.githubusercontent.com/yywang118/Coursera-Capstone/master/OHIO_location.csv

2. Restaurants in each city from Franklin County in Ohio.
   - Data source: Fousquare API
   - Description: Use API to get all the venues in each city.

2.2 Approach

1. Collect the Ohio city data
   from https://en.wikipedia.org/wiki/List_of_cities_in_Ohio

2. Collect the latitude and longitude for each city from
   https://raw.githubusercontent.com/yywang118/Coursera-Capstone/master/OHIO_location.csv

3. Using FourSquare API we will find all venues for each city.
   a. Filter out all venues that are Restaurants.
   b. Find rating, tips and like count for each Indian Restaurants using FourSquare API.
   c. Using rating for each restaurant, we will sort that data.
   d. Visualize the Ranking of neighborhoods using folium library(python)

3. Data Preparation

3.1 Web-scrapping and Data Cleaning

First, a table containing a list of 247 cities in Ohio was scraped from Wiki (https://en.wikipedia.org/wiki/List_of_cities_in_Ohio). *Requests, Beautifulsoup4* and *pandas* libraries were used to scrap the date, create an initial data frame (df), which contains the names of the 247 cities, population, and their home counties for the Ohio State. The head of the data frame (df) is shown below:

```
from bs4 import BeautifulSoup
import matplotlib.pyplot as plt
import os
from sklearn.cluster import KMeans
import folium
from geopy.geocoders import Nominatim
```

```
List_url='https://en.wikipedia.org/wiki/List_of_cities_in_Ohio'
source = requests.get(List_url).text
soup = BeautifulSoup(source, 'xml')
table=soup.find('table')
column_names=['City','Population','County']
df = pd.DataFrame(columns=column_names)
for tr_cell in table.find_all('tr'):
    row_data=[]
    for td_cell in tr_cell.find_all('td'):
        row_data.append(td_cell.text.strip())
    if len(row_data)==3:
        df.loc[len(df)] = row_data
        df.head()
```

```
df.head()
```

	City	Population	County
0	Akron	198,006	Summit
1	Alliance	21,616	Mahoning
2	Amherst	12,021	Lorain
3	Ashland	20,423	Ashland
4	Ashtabula	18,079	Ashtabula

3.2 Coordinates of the Major cities in Ohio

Geo-location function was defined to automatically retrieve coordinates of the cities; however, timeout error prevents the function from working properly. Luckily, a csv. File containing coordinates information for all the US cities is available online. This file was uploaded on github repository and the coordinates were retrieved for each major city in Ohio.

```
geo_df.head()
```

]:

	zip	lat	lng	city	state_id	state_name	zcta	parent_zcta	population	density	county_fips	county_name
0	NaN	39.969238	-82.936864	Bexley	OH	Ohio	True	NaN	2810	34.9	39089	Franklin
1	43001.0	40.088650	-82.613630	Alexandria	OH	Ohio	True	NaN	2810	34.9	39089	Licking
2	43002.0	40.063330	-83.173900	Amlin	OH	Ohio	True	NaN	3744	872.4	39049	Franklin
3	43003.0	40.410340	-82.975380	Ashley	OH	Ohio	True	NaN	3005	33.2	39041	Delaware
4	43004.0	40.015810	-82.800120	Blacklick	OH	Ohio	True	NaN	26139	678.2	39049	Franklin

The two data frames were merged and cleaned to only include variables of interest. Only cities from Franklin County were kept. There were a total of 10 cities were included.

```
geo_final_data=geo_merged[['City','Population','County_y','zip','lat','lng']]
```

```
franklin_data=geo_final_data[geo_final_data['County_y'].str.contains("Franklin",na=False)]
franklin_data
```

	City	Population	County_y	zip	lat	lng
34	Bexley	13,057	Franklin	NaN	39.969238	-82.936864
46	Canal Winchester	7,101	Franklin	43110.0	39.826340	-82.799450
141	Columbus	892,672	Franklin	43085.0	40.099860	-83.015700
207	Dublin	48,647	Franklin	43016.0	40.098590	-83.153760
228	Grove City	41,627	Franklin	43123.0	39.868550	-83.119880
229	Groveport	5,363	Franklin	43125.0	39.836640	-82.884460
235	Hilliard	33,672	Franklin	43026.0	40.022750	-83.185520
289	New Albany	7,724	Franklin	43054.0	40.080840	-82.803480
315	Reynoldsburg	35,893	Franklin	43068.0	39.956370	-82.785020
379	Westerville	36,120	Franklin	43081.0	40.110640	-82.891340

4. Exploring Major Cities using Foursquare API

4.1 Foursquare Location Data

I used foursquare API to retrieve popular venues around the major cities in Franklin county. I've chosen the limit of 100 popular spots for each major city within. Because we were only interested in the restaurant, we cleaned the data returned by Foursquare, and only included

restaurants. Below is the head of the data-frame including restaurant information fort the 10 major cities from Franklin county in Ohio.

```
# Create a Data-Frame out of it to Concentrate Only on Restaurants

franklin_venues_only_restaurant = franklin_venues[franklin_venues['Venue Category'].str.contains('Restaurant
franklin_venues_only_restaurant.index = np.arange(1, len(franklin_venues_only_restaurant)+1)
print ("Shape of the Data-Frame with Venue Category only Restaurant: ", franklin_venues_only_restaurant.shape
franklin_venues_only_restaurant
```

```
Shape of the Data-Frame with Venue Category only Restaurant:  (239, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Bexley	39.969238	-82.936864	Giuseppe's Ritrovo	39.957316	-82.938282	Italian Restaurant
2	Bexley	39.969238	-82.936864	Brassica	39.957333	-82.940051	Mediterranean Restaurant
3	Bexley	39.969238	-82.936864	Tavern Olde Towne	39.962913	-82.976560	Restaurant
4	Bexley	39.969238	-82.936864	Schmidt's Restaurant und Sausage Haus	39.946222	-82.990993	German Restaurant
5	Bexley	39.969238	-82.936864	Brassica	39.976273	-83.003219	Restaurant
6	Bexley	39.969238	-82.936864	Lindey's	39.950775	-82.993778	American Restaurant
7	Bexley	39.969238	-82.936864	Arepazo Tapas Bar Grill	39.951473	-82.998798	Latin American Restaurant
8	Bexley	39.969238	-82.936864	The Guild House	39.975110	-83.002959	American Restaurant

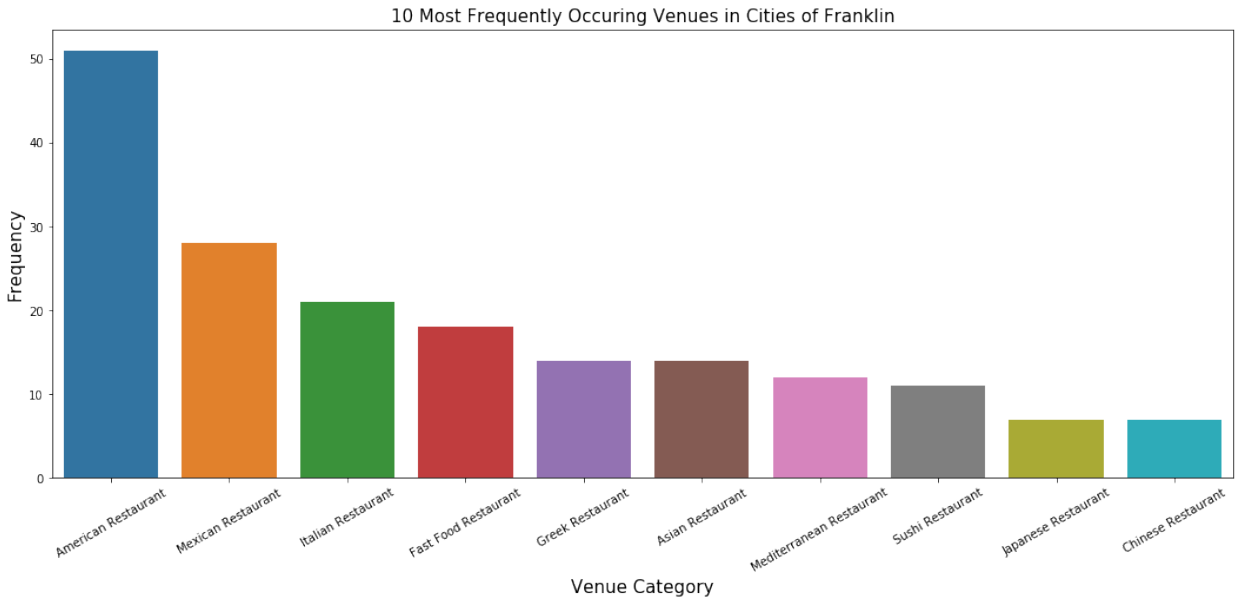
4.2 Most Popular Restaurant Style

Below are the number of restaurants grouped by type. We can see the American restaurant is the most popular, and the Mexican restaurant is the second. These findings are not surprising as American and Mexican people constitute the major population in Ohio.

```
print (franklin_venues_only_restaurant['Venue Category'].value_counts())
```

```
American Restaurant              52
Mexican Restaurant               30
Italian Restaurant               19
Fast Food Restaurant             18
Greek Restaurant                 14
Asian Restaurant                 13
Mediterranean Restaurant         12
Sushi Restaurant                 11
Japanese Restaurant               7
Chinese Restaurant                7
Thai Restaurant                   6
Restaurant                        5
Seafood Restaurant                5
German Restaurant                 4
French Restaurant                 4
Vietnamese Restaurant             4
Middle Eastern Restaurant         3
Latin American Restaurant         3
Indian Restaurant                 3
Vegetarian / Vegan Restaurant     2
Korean Restaurant                 2
Arepa Restaurant                  2
```

Below is a plot showing the 10 most popular restaurants in Franklin county.



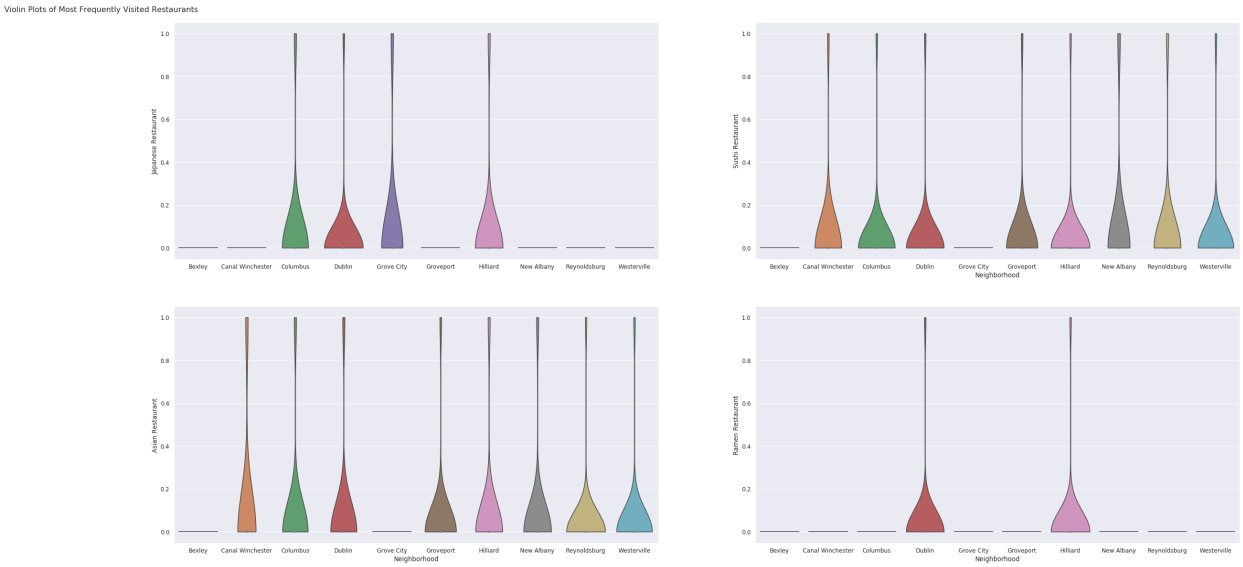
4.3 Ten Most Popular Restaurant Styles by City

I also examined data and explored the 10 most popular restaurants for each city. Special attention was given to the potentially restaurants providing Japanese style cuisines, including Asian restaurant, Ramen restaurant, Sushi Restaurant, and Japanese restaurants.

]:

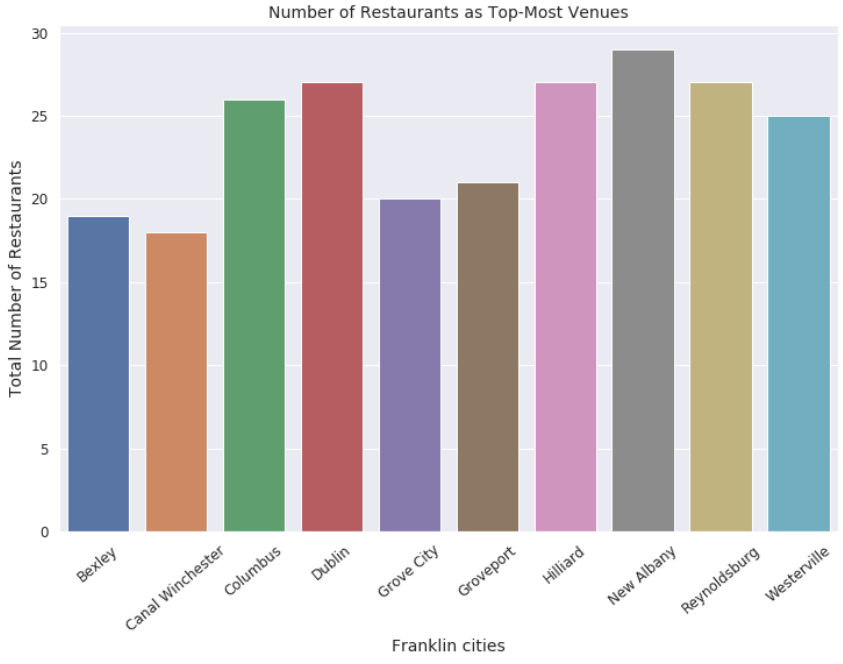
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bexley	American Restaurant	Italian Restaurant	New American Restaurant	German Restaurant	Restaurant	Latin American Restaurant	Mediterranean Restaurant	Mexican Restaurant	Vietnamese Restaurant	Ramen Restaurant
1	Canal Winchester	American Restaurant	Fast Food Restaurant	Italian Restaurant	Asian Restaurant	Korean Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Greek Restaurant	Sushi Restaurant	Thai Restaurant
2	Columbus	American Restaurant	Greek Restaurant	Mexican Restaurant	Asian Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Thai Restaurant	Seafood Restaurant	Mediterranean Restaurant
3	Dublin	American Restaurant	Mexican Restaurant	Fast Food Restaurant	Italian Restaurant	Asian Restaurant	Greek Restaurant	Chinese Restaurant	French Restaurant	Indian Restaurant	Mediterranean Restaurant
4	Grove City	American Restaurant	Mexican Restaurant	Fast Food Restaurant	German Restaurant	Latin American Restaurant	Japanese Restaurant	Seafood Restaurant	Greek Restaurant	Chinese Restaurant	Mediterranean Restaurant
5	Groveport	American Restaurant	Fast Food Restaurant	Thai Restaurant	Italian Restaurant	Middle Eastern Restaurant	Asian Restaurant	Ethiopian Restaurant	Greek Restaurant	Mediterranean Restaurant	Vietnamese Restaurant
6	Hilliard	American Restaurant	Mexican Restaurant	Italian Restaurant	Asian Restaurant	Greek Restaurant	Japanese Restaurant	Ramen Restaurant	Sushi Restaurant	Mediterranean Restaurant	Cajun / Creole Restaurant
7	New Albany	American Restaurant	Mexican Restaurant	Sushi Restaurant	Mediterranean Restaurant	Fast Food Restaurant	Italian Restaurant	Asian Restaurant	Restaurant	Fondue Restaurant	Chinese Restaurant
8	Reynoldsburg	American Restaurant	Mexican Restaurant	Fast Food Restaurant	Sushi Restaurant	Greek Restaurant	Mediterranean Restaurant	Arepa Restaurant	Asian Restaurant	Chinese Restaurant	Ethiopian Restaurant
9	Westerville	Mexican Restaurant	American Restaurant	Chinese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Asian Restaurant	Brazilian Restaurant	Fast Food Restaurant	Fondue Restaurant	Greek Restaurant

Below is a plot showing Japanese restaurants separated by City. From the plot, we can see that the cities which have the smallest number of Japanese style restaurants are Bexley; Canal Winchester, Grove city.



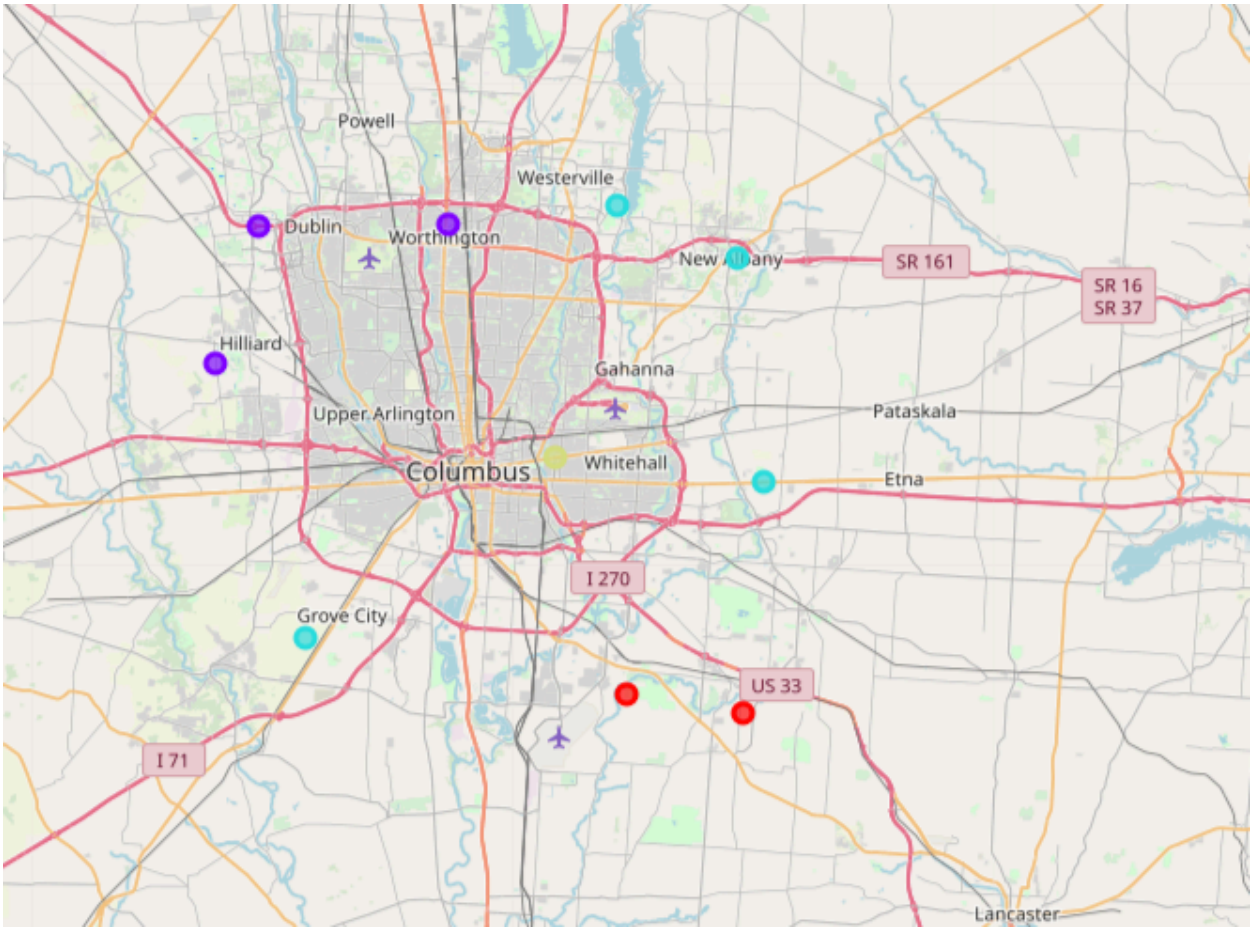
4.4 Number of Restaurants by City

I also examined the total number of restaurants by city, as this information would also influence the decision making. Based on the plot, we can see that Bexley, Canal Winchester, Grove City, and Groveport have smaller number of restaurants.



4.3 Clustering the major cities of Franklin

Finally, we clustered these 10 cities of the number of restaurants based on the venue categories using K-means clustering. The 4 clusters in a leaflet map using folium library were shown below.



5. Results and Discussion

This project used data from web resources like Wikipedia, python libraries like Geopy, pandas, folium, and Foursquare API, to explore restaurants in the 10 major cities in Franklin count in Ohio. The stakeholder would like to identify a place for opening a Japanese restaurant in one of the cities in the county. We focus the discussion on restaurants providing Japanese style cuisines, including Asian Restaurants, Sushi Restaurants, Japanese Restaurants, and Ramen Restaurnats. Below are the observations from the data:

- In general, American style restaurants are the most popular restaurants among the 10 major cities; Mexican style restaurants are ranked second; Italian style restaurants ranked third. For restaurants which may provide Japanese cuisines: Asian restaurant ranked 6th, Sushi restaurant ranked 8th, and Japanese restaurant ranked 9th.

- Asian restaurant ranked 2nd in Columbus, and 3rd in Groveport as the most common venue. Sushi restaurant ranked 3rd in New Albany.

- Columbus, Dublin, Hillard, and New Albany are ranked top in terms of the total number of restaurants, whereas Bexley and Canal Winchester ranked lowest.

- Grove city and Bexley have the smallest number of Japanese style restaurants.

- Since the clustering was based only on the most common venues of each district, Canal Winchester and Groveport fall under the same cluster and, Columbus, Dublin, and Hillard fall under the same cluster; Bexley stands on its own, and the rest fall under the same cluster.

Based on the above analysis, Grove city and Bexley will provide the least competition for an upcoming Japanese restaurant as the frequency of Japanese style restaurants as common venue are very low compared to other cities. However, other factors need to be taken into consideration are population and land price. Based on the data from wiki, Columbus has the largest population which far surpasses other cities in Franklin County. Given that Japanese style restaurant is not among the most common venue in Columbus, it is possible that Columbus may actually be a good place to open a new Japanese restaurant. Another factor that needs to be taken into consideration is the land price. I suggest the stakeholder exploring the land prices in these cities and factor the land price into their decision making.

6. Conclusion

In this project, I've made use of some frequently used python libraries to scrap web-data, clean the data, and use Foursquare API to explore the major cities in Franklin county to explore potential locations to open up a new Japanese restaurant. The data and results from the analysis were discussed. The stakeholder is also recommended to take into consideration the land price and population in their decision making.