BATTLE OF NEIGHBOORHOODS – SUSHI RESTAURANT

Applied Data Science Capstone Project
- **Jahnavi Sandhya Lanka**

Notes:

Leveraging the knowledge gained from the IBM Data Science course to perform analysis and build machine learning models to further face and try to resolve real world's problems

Scenario

Sushi is popular around the world; the term sushi actually refers to foods that use a type of rice seasoned with vinegar and are garnished with raw fish or vegetable. There are several types of sushi, making it an appealing food to people with a wide range of tastes. One form of sushi, nigiri-zushi, are hand-formed mounds of rice with a dab of wasabi topped with various ingredients. Popular nigiri-zushi include maguro (tuna), toro (belly of tuna), hamachi (yellowtail), and ebi (shrimp). The main aim of the restaurnt is to take the food to the customers. The business plan would be successful if the restaurants was aimed at the right target customers. The turnaround time for the customer's orders should be quick to satisfy customers. The following business problem section is based on this scenario.

Business Problem

A client of mine who is based in Orlando, Florida, wants to start a Sushi Restaurant with the concept of selling japanese cuisine. It mainly concentrates on foods like shrimp, escolar, tuna, red snapper etc. These foods are pretty quick to cook and serve to the customers. The main goal of the client is to aim at the customers like students, business employees, residents, shopping malls, or mostly commonly visited places by the public. The challenge is to find the top three neighborhoods among the most popular neighborhoods in Orlando, Florida so that my client can kick-start their business and estimate the turn out profits in the next two to three years by concentrating on the top three neighborhoods.

This business problem mainly concentrates on finding the top three neighborhoods to start the Sushi Restaurant.

DataSource

- List of super neighborhoods in Orlando, FL: https://data.cityoforlando.net/Government-General/Neighborhoods/dpx3-qjrc
- Foursquare data to find the venues

How will the data be used to answer the business needs?

The data mentioned above will be used to explore and target locations across different venue categories present in the neighborhoods

- Use Foursquare and Geopy data to map top venues for the super neighborhoods of Houston and cluster them in groups

- City-Data to get the neighborhoods information

- Additional data will be added from open data sources if available in the future if the data is insufficient

By extracting the venues of the neighborhoods, we can determine which of the neighborhoods are closely located to each other which would give an idea that the customer count is high in that area. By using Foursquare data and the Orlando's neighborhood data, we can recommend the top three neighborhoods by performing machine learning techniques and can visualize them through a graph or a map.

Using Beautiful Soup package for python, I was able to extract the neighborhoods from the link and successfully appended the neighborhoods to a data frame. Next by implementing Geopy code, I was able to append longitude and latitude of each neighborhood. Certain neighborhoods will be excluded since few of the neighborhoods co-ordinates were not retrieved by Geopy.

Methodology

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.

With above data, I can use the concept of manually assigning weights to calculate a score and predict the most popular neighborhood to resolve the problem. Combine with Foursquare API which provides how many venues in different category of Residential Areas, Universities, Offices, and Shopping Malls can be created. Then the matrix can be used to calculate a score to form weighted matrix and the most favorable spots are ranked based on the neighborhood's score.

The weighted matrix can be applied on several target locations with each venue category information to generate a ranking result by using indexes. The top three on the ranking list can be recommended to the client.
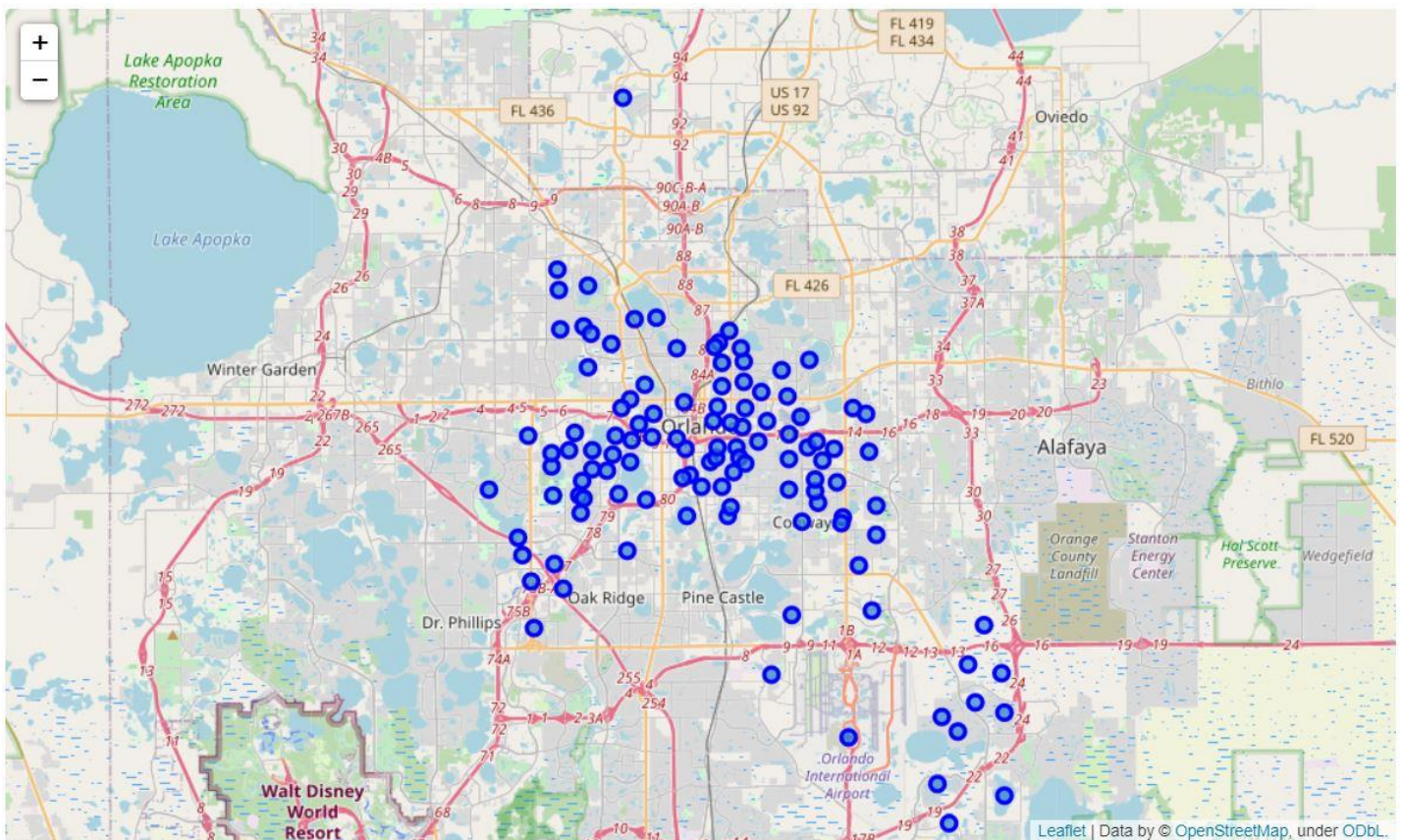
With the help of folium package from the above figure we can observe that the Orlando's neighborhoods are spread across the state. This image is to get an overview of what type of geographical map we are working with and what would be the initial thoughts.

Since Sushi Restaurant is a booming business, the next step was to determine the count of Sushi Restaurants across each neighborhood.


Count of Sushi Restaurants by Neighborhood
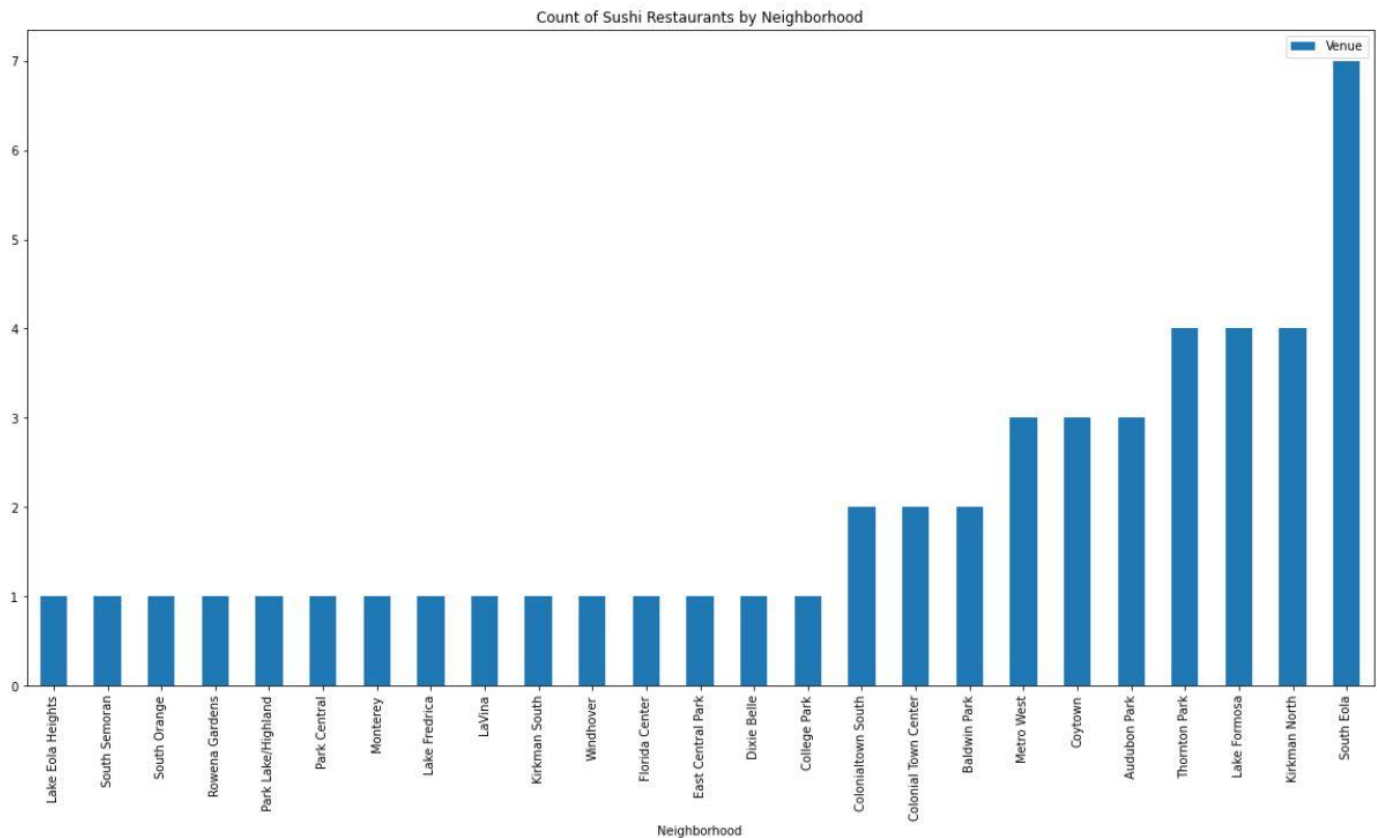
With the help of Matplotlib library and Foursquare data, I was able to plot the graph which visually it helped me, and the client determine that the count of Sushi Restaurants is high in the following neighborhoods,

- South Eola
- Krikman North
- Lake Formosa

Client's concern was that the competition would be high and there would several concurrences of other sushi restaurants. Next step was to determine the target places/audience which are residential areas, universities, shopping malls and offices on the map by using foursquare data.

By retrieving the data from foursquare and mapping those venue of each category on the map we get a visual with the most venues closely located to each other which would help determine the hot spots for the sushi restaurant. Universities are plotted in green, Residential Areas are plotted in red, offices are plotted in blue, and shopping malls are plotted in yellow.
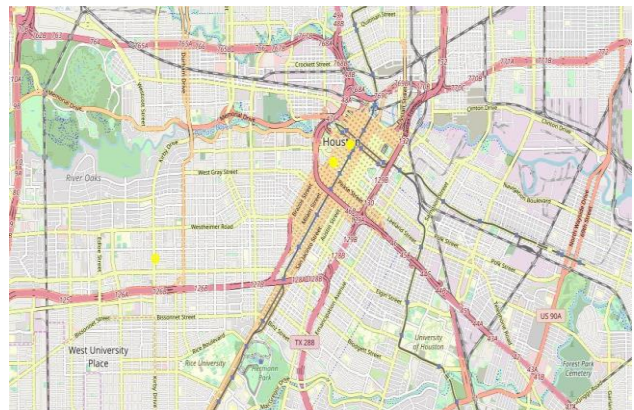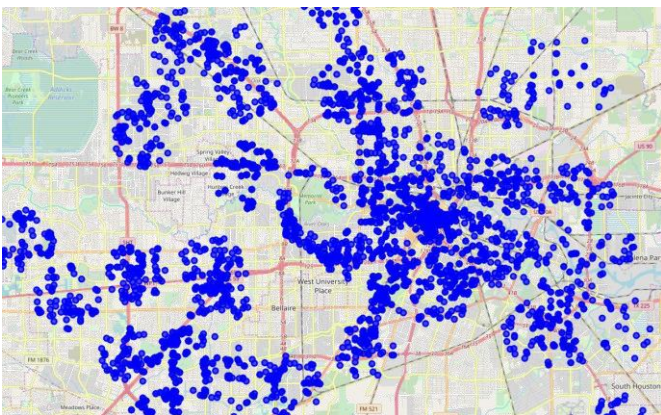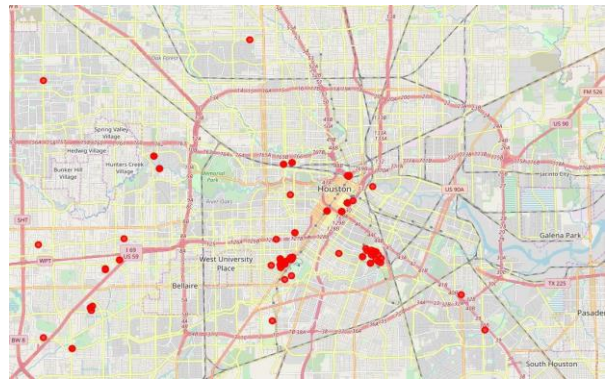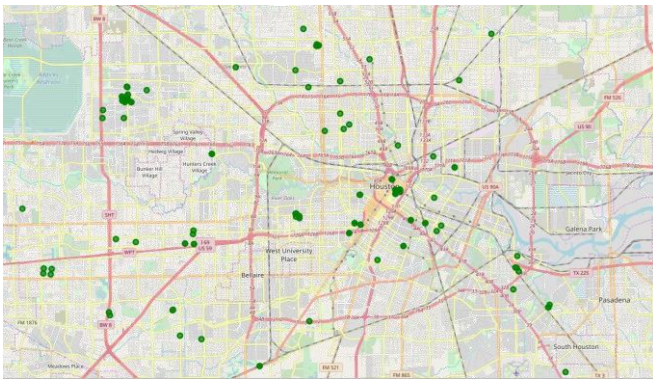
From the above figures we get an idea that the count of venues which are closer to the downtown are the high comparatively to the outskirts of the downtown which is an expected result.

Weights are assigned to each venue category, Food Truck being in negative since we need to avoid the concurrences of other food trucks, positive values for the other categories since they are the target audience. Offices have high weights than the other categories since they would be the prime target for this business.

After performing calculation using this formula

$$x = (y * i) + (y_1 * i_1) + ………… + (n * m)$$

we calculate the weighted score for each neighborhood by creating a matrix

	Neighborhood	Longitude	Latitude	Residential Area	University	Offices	shopping mall	Sushi	Weighted Score
0	Lake Eola Heights	-81.3706	28.5499	39.0	2.0	100.0	0.0	1.0	241.0
1	Holden/Parramore	-81.3899	28.537	35.0	1.0	100.0	0.0	0.0	236.5
2	South Eola	-81.373	28.544	39.0	2.0	100.0	0.0	7.0	235.0
3	Lake Dot	-81.3867	28.5521	31.0	1.0	100.0	0.0	0.0	232.5
4	Lake Cherokee	-81.3712	28.5333	30.0	1.0	100.0	0.0	0.0	231.5
5	Thornton Park	-81.3646	28.5433	31.0	2.0	100.0	0.0	4.0	230.0
6	Lake Copeland	-81.3747	28.5272	28.0	1.0	100.0	0.0	0.0	229.5
7	Callahan	-81.3861	28.5329	28.0	1.0	100.0	0.0	0.0	229.5
8	Delaney Park	-81.3716	28.5291	29.0	0.0	100.0	0.0	0.0	229.0
9	Park Lake/Highland	-81.3691	28.5586	25.0	1.0	100.0	0.0	1.0	225.5

Results

After the analysis and calculations, we can observe the top three neighborhoods in Houston to start are food truck are,

1. Lake Eola Heights

2. Holden/ Parramore

3. South Eola

Mapping the top three neighborhoods that are predicted during our analysis
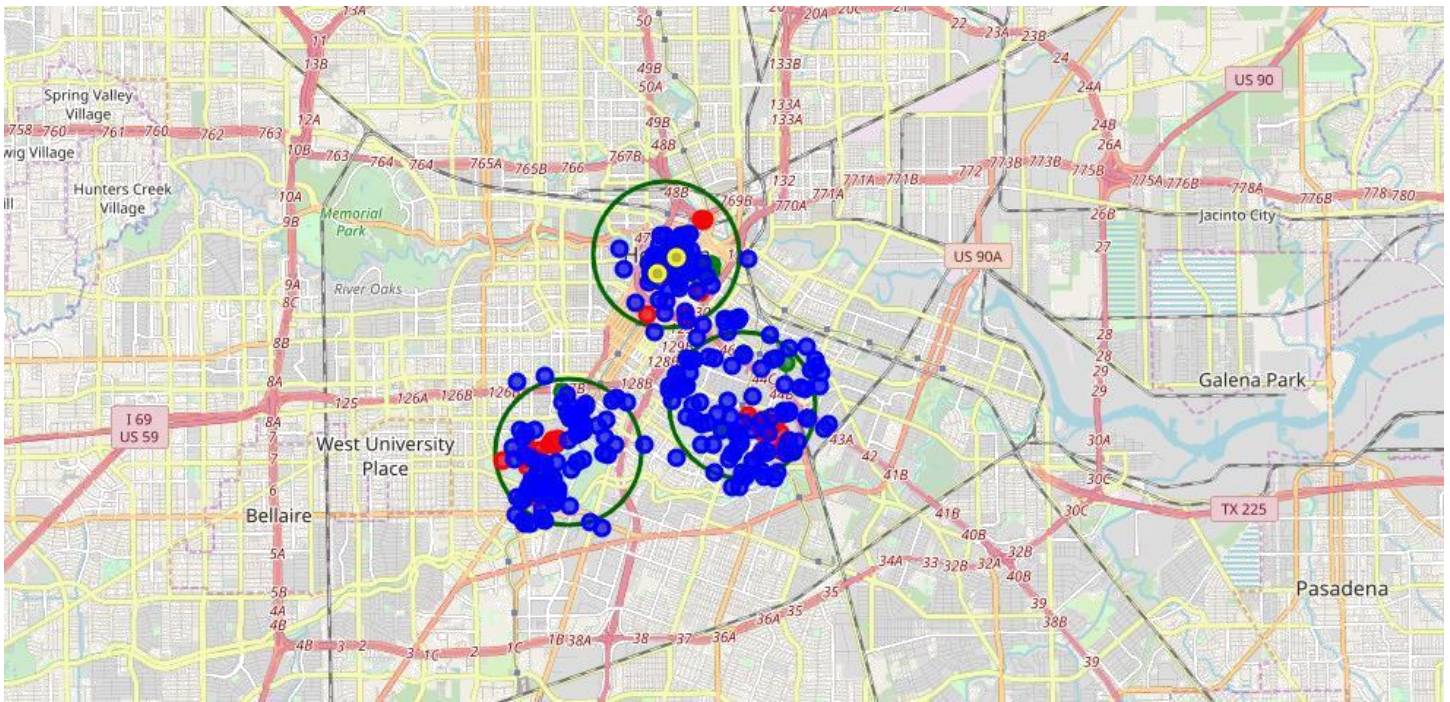
Discussion

From our initial analysis we could observe that the top 3 neighborhoods based on the count of Sushi Restaurants per neighborhood were South Eola, Krikman North and Lake Formosa

- Client was worried about their competition and wanted me find the least concurrent places where the client can avoid other competitors

- For further analysis we downloaded residential, office and university data since they would be great target audience and assigned them weights according to their traffic, and also assigned weight to the trucks data too to avoid concurrence with the other competitors

- Weighted scores are calculated and the result is sorted according to their weighted score which yielded us the top 3 neighborhoods

- We can also suggest the next 3 neighborhoods for the clients according to our analysis which would be
    - Lake Dot
    - Lake Cherokee
    - Thornton Park

- Limitations,
  - Geopy could not yield the results for all the super neighborhoods in Houston which were excluded from the data and few of the other neighborhoods have the same co-ordinates which could mean that the neighborhoods are close to each other or the Geopy is producing faulty co-ordinates
  - Accuracy of venues purely depends on Foursquare

Conclusion

Based on our analysis the client should start their food truck business in the following three neighborhoods, Lake Eola Heights, Holden/Parramore and South Eola

Upon further analysis, the client can also do business in the next three neighborhoods, Lake Dot, Lake Cherokee and Thornton Park.