

# BATTLE OF NEIGHBORHOODS – FOOD TRUCK EDITION

Applied Data Science Capstone Project

- Varun Duvalla

## 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

Food trucks have been growing since it was known to the public. The main aim of the food truck is to take the food to the customers. The business plan would be successful if the truck was aimed at the right target customers. Unlike traditional restaurant and fast food chains, the food is cooked at that moment and served fresh to the customer. 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 Houston, Texas, wants to start a food truck with the concept of selling Middle Eastern cuisine. It mainly concentrates on foods like shawarma, wraps, salads with hummus etc. These foods are pretty quick to cook and easy to serve the customers. The main goal of the client is to aim at the customers like students, business employees, constructions sites, public parks, or mostly commonly visited places by the public. The challenge is to find the top three neighborhoods among the most popular neighborhoods in Houston, Texas so that my client can kick-start their business and estimate the turn out profits in the next one to two years by concentrating on the top three neighborhoods.

This business problem mainly concentrates on finding the top three neighborhoods to start the food truck business.

## Data

Sources:

- List of super neighborhoods in Houston, TX:  
<http://www.city-data.com/nbmaps/neigh-Houston-Texas.html> (Link changed due to inaccurate data in Wikipedia)
- 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 Houston'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 Schools, Universities, Offices, and Shopping Malls a 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.

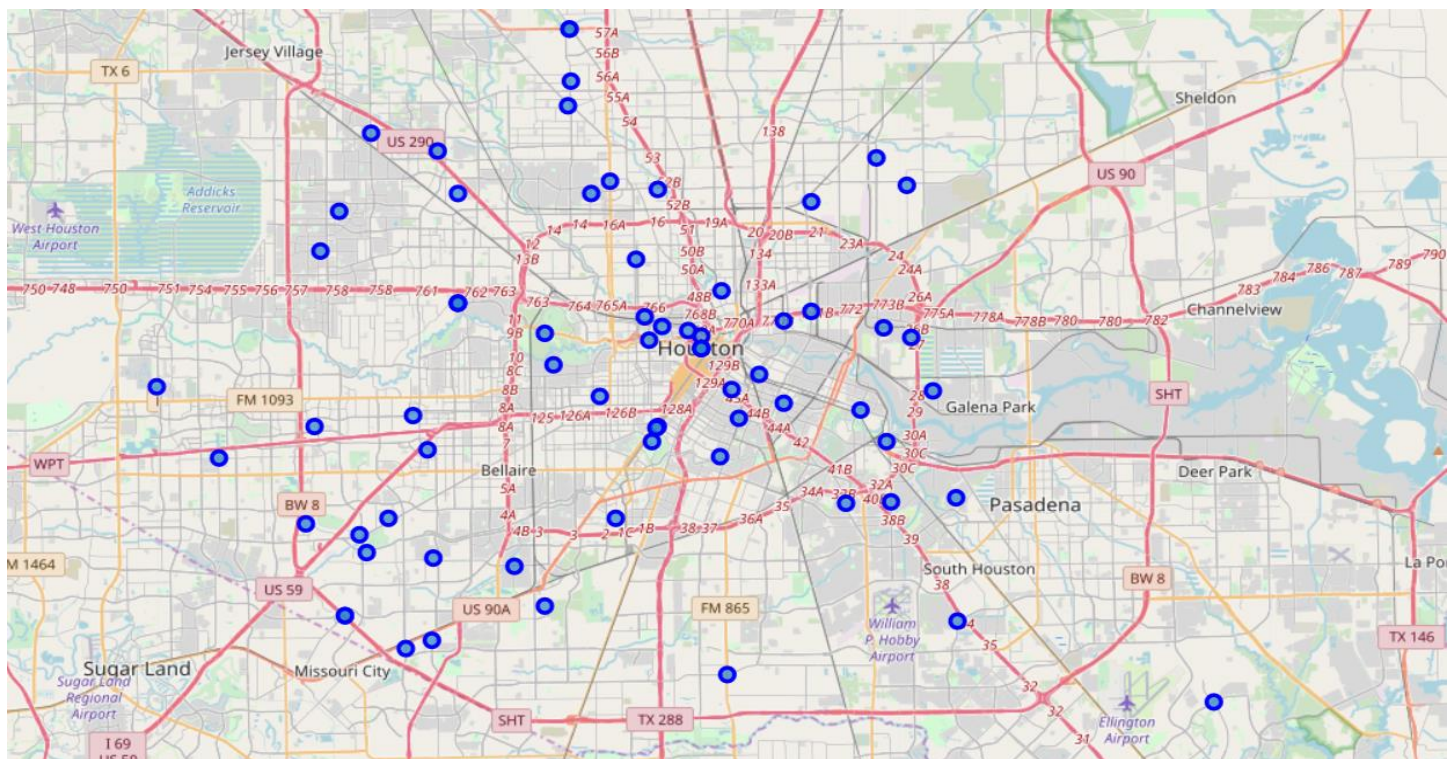


Fig 1. Map of Houston's Super Neighborhoods

With the help of folium package from the above figure we can observe that the Houston's super 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 food trucks is a booming business, the next step was to determine the count of food trucks across each neighborhood.

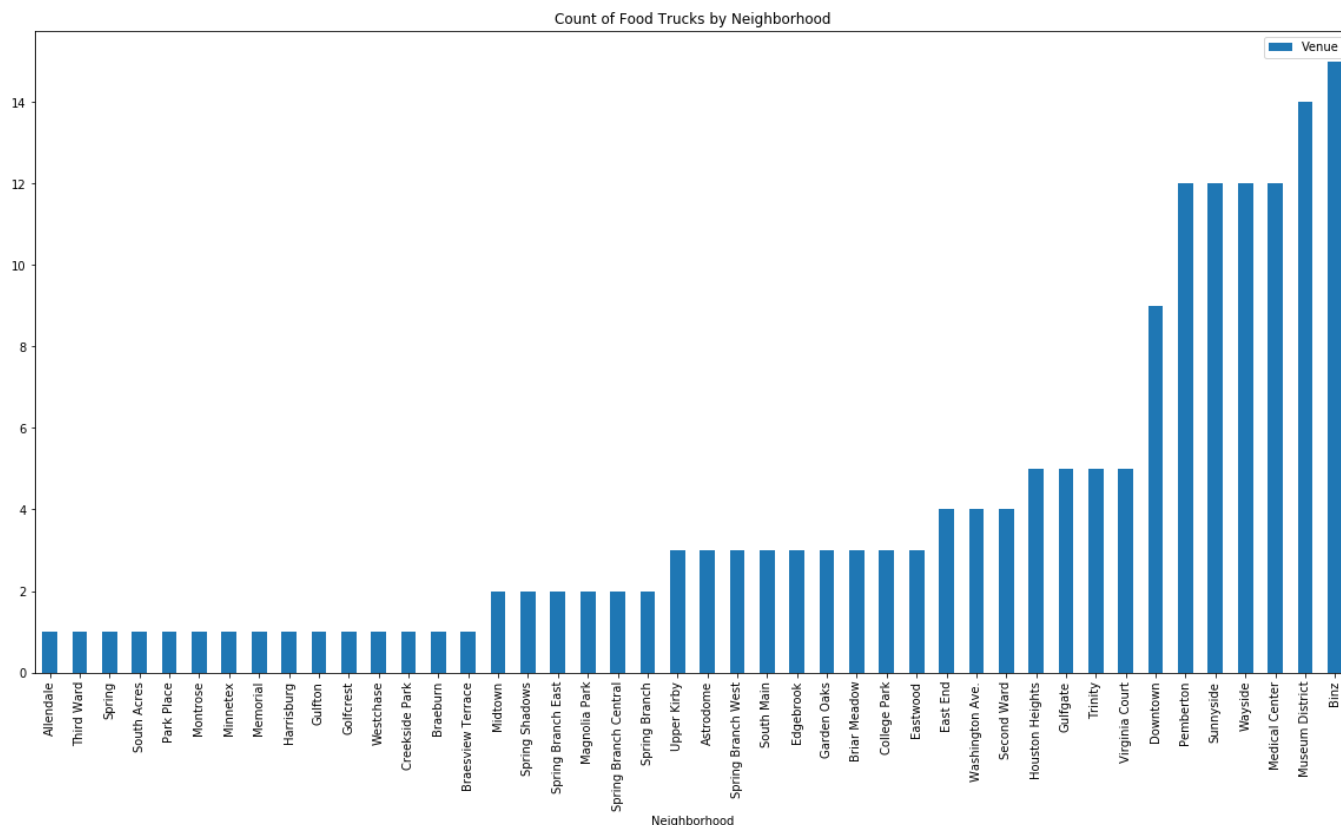


Fig 2. Count of Food Truck by Neighborhood

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 food truck are high in the following neighborhoods,

- Binz
- Museum District
- Medical Center

Client's concern was that the competition would be high and there would several concurrences of other food trucks. Next step was to determine the target places/audience which are schools, 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 food truck business. Schools are plotted in green, Universities are plotted in red, offices are plotted in blue, and shopping malls are plotted in yellow.



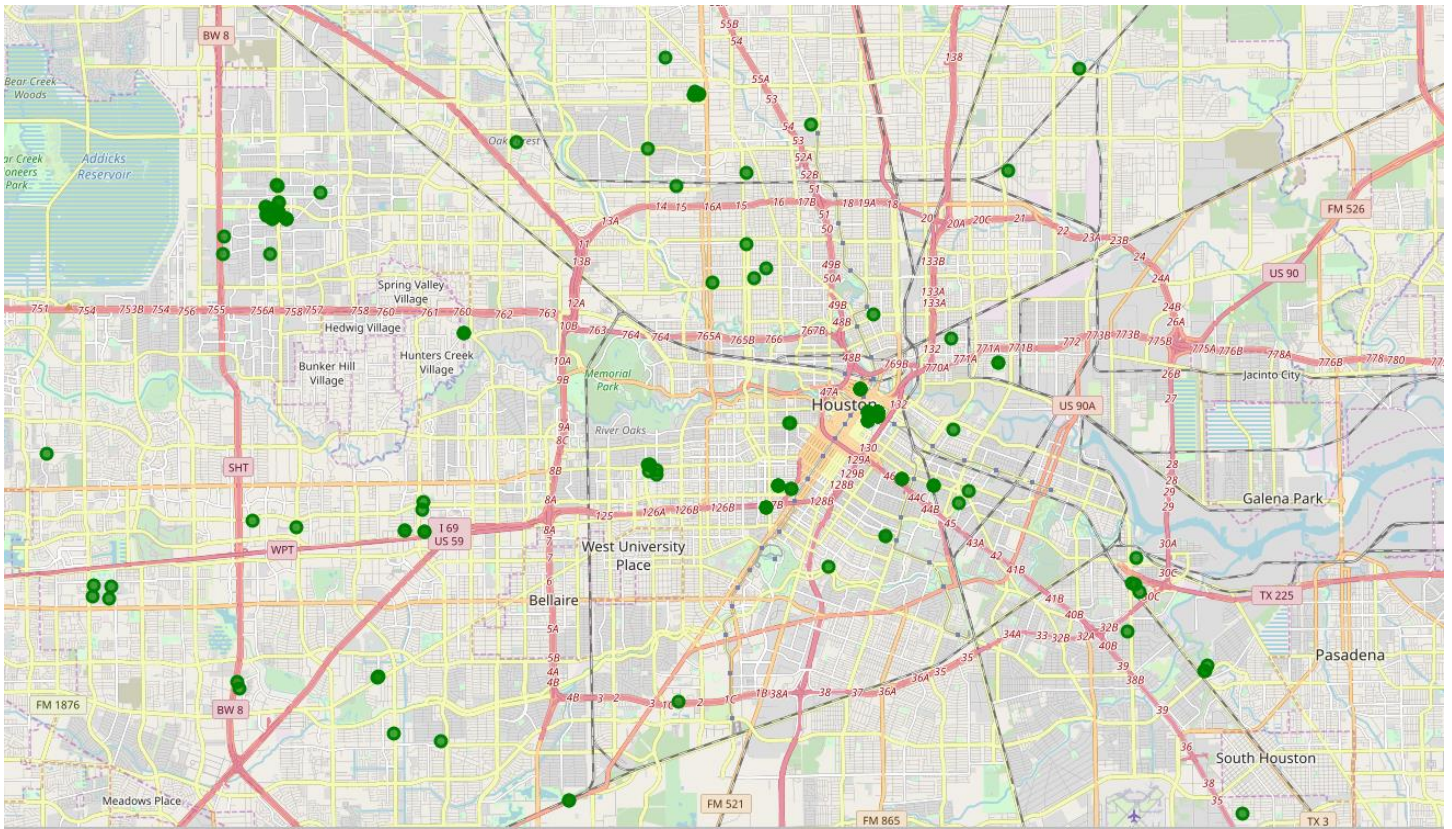


Fig 3. Map of Schools in the Houston Neighborhoods

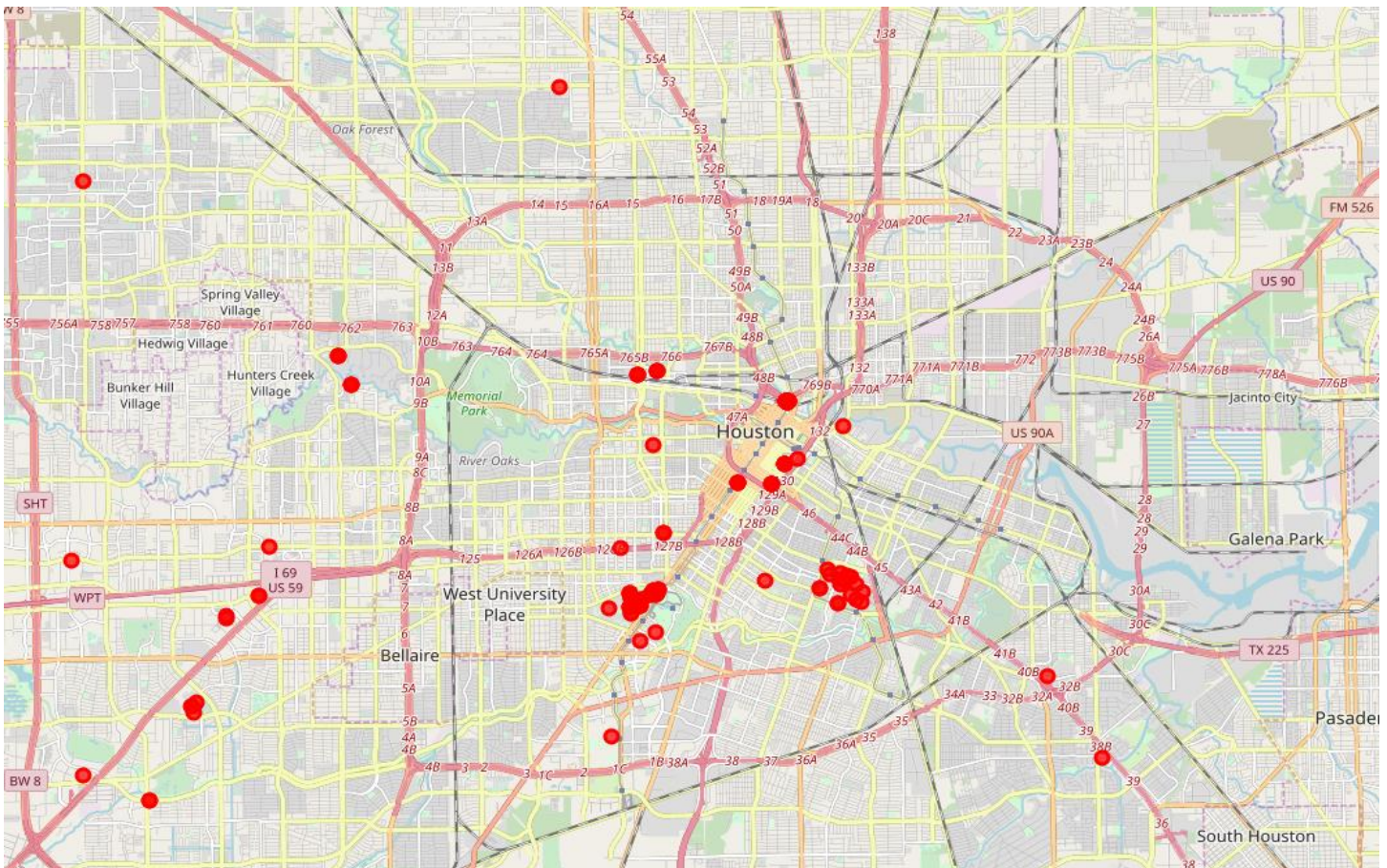


Fig 4. Map of Universities in the Houston Neighborhoods



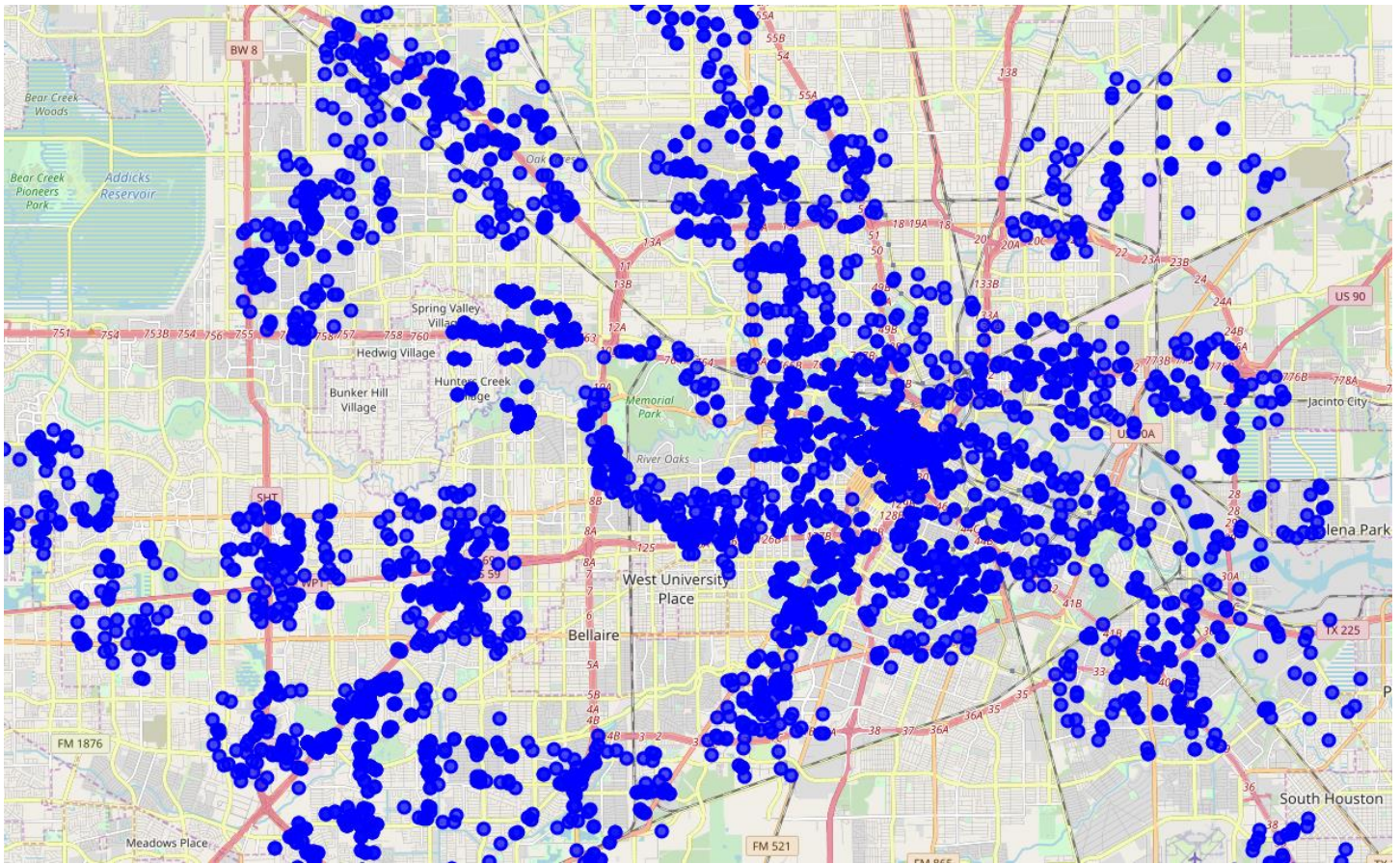


Fig 5. Map of Offices in the Houston Neighborhoods

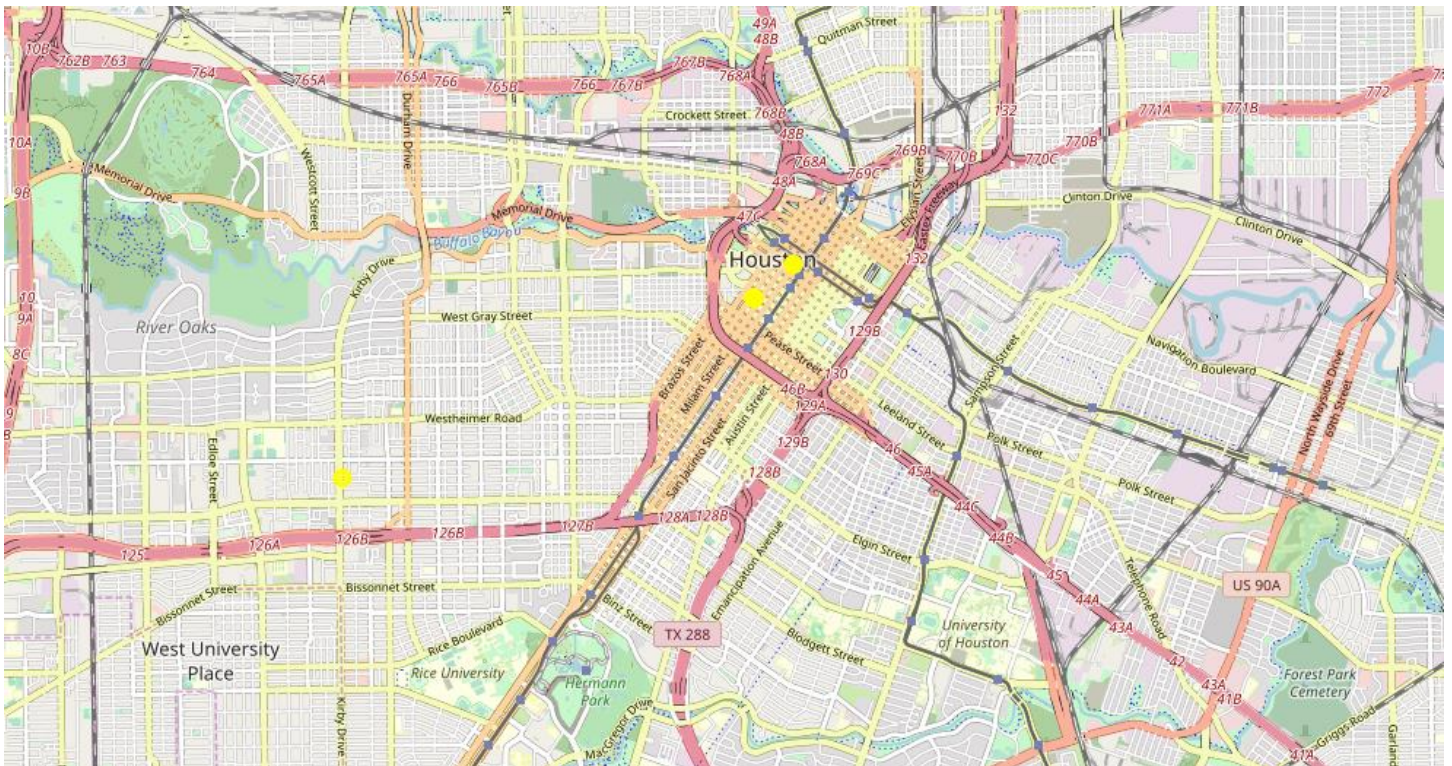


Fig 6. Map of Shopping Malls in the Houston neighborhood



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

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

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

	Neighborhood	Latitude	Longitude	Schools	University	Offices	Shopping Mall	Food Truck	Weighted Score
0	Memorial	29.7186	-95.3909	1.0	15.0	100.0	0.0	1.0	222.5
1	Third Ward	29.7284	-95.3499	3.0	15.0	96.0	0.0	1.0	216.5
2	Medical Center	29.7589	-95.3677	8.0	7.0	100.0	2.0	12.0	209.5
3	Wayside	29.7589	-95.3677	8.0	7.0	100.0	2.0	12.0	209.5
4	Museum District	29.7247	-95.3891	4.0	13.0	100.0	0.0	14.0	209.5
5	Pemberton	29.7589	-95.3677	8.0	7.0	100.0	2.0	12.0	209.5
6	Sunnyside	29.7589	-95.3677	8.0	7.0	100.0	2.0	12.0	209.5
7	Sharpstown	29.6849	-95.5156	2.0	4.0	100.0	0.0	0.0	208.0
8	Binz	29.7249	-95.3881	4.0	12.0	100.0	0.0	15.0	207.0
9	Spring Shadows	29.8185	-95.539	59.0	1.0	74.0	0.0	2.0	206.5

Fig 7. Weighted matrix data frame of each Neighborhood

## Results

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

1. Memorial
2. Third Ward
3. Medical Center

Mapping the top three neighborhoods that are predicted during our analysis

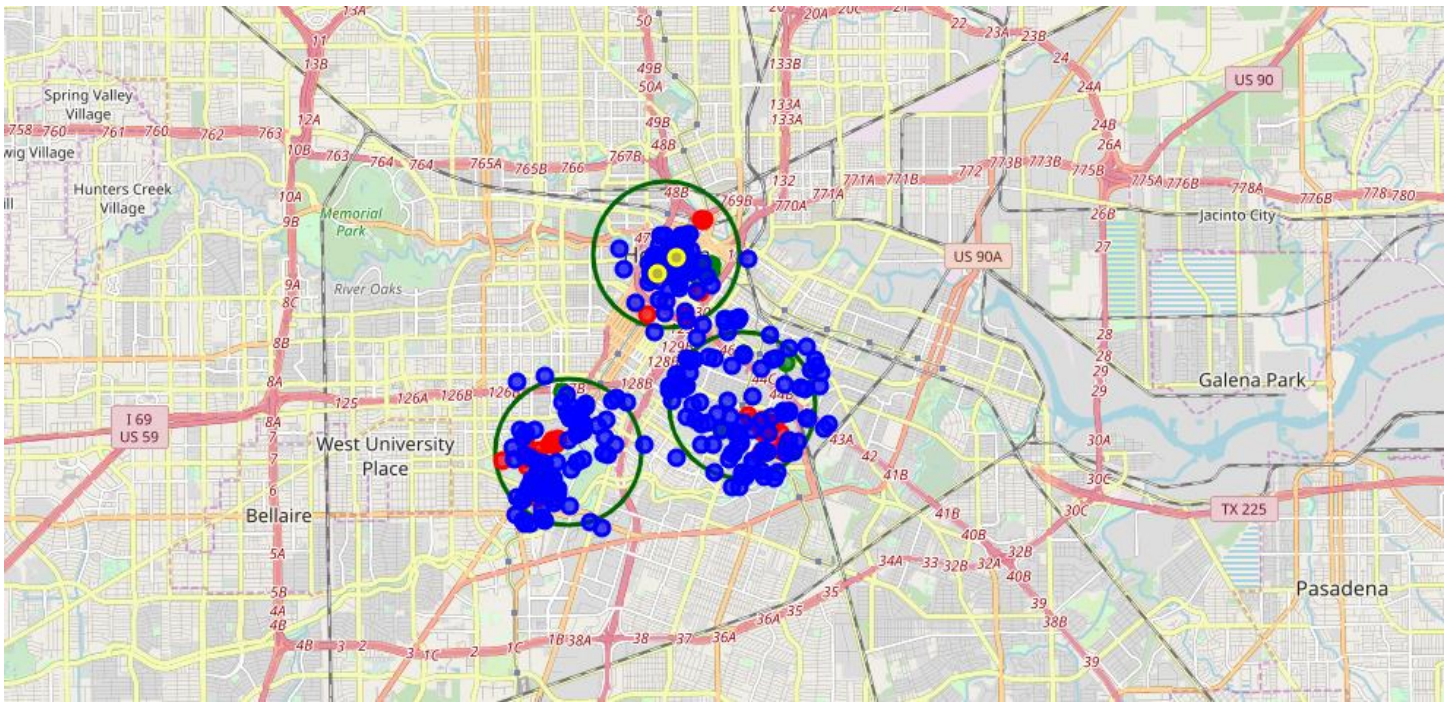


Fig 8. Top 3 neighborhoods encircled with the green color circle

## Discussion

- From our initial analysis we could observe that the top 3 neighborhoods based on the count of food truck per neighborhood were Binz, Museum District and Medical Center
- 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 school, 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,
  - Wayside
  - Museum District
  - Pembertown
- 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,

- Memorial
- Third Ward
- Medical Center

Upon further analysis, the client can also do business in the next three neighborhoods,

- Wayside
- Museum District
- Pembertown