Capstone Project – Battle of the Neighborhoods

Exploring Business Opportunity in West Houston

- ## Introduction

Large cities can provide businesses the opportunity to grow, expand, and become more well known. Houston, Texas is the 4th largest city in the United States and rapidly growing, making it a great city for a new business to access a large group of new potential clients. The famous saying in real estate, "Location, location, location" is applicable to the business world as well. Location has the potential to be the deciding factor in whether a business succeeds or fails. For simplicity, different suburban areas in the West Houston will be the focal point of this project. Factors that will be considered include

- The number of other restaurants in each suburban area
- Other venues in the suburban area that might attract restaurant patrons (movie theaters, shopping centers, etc.)

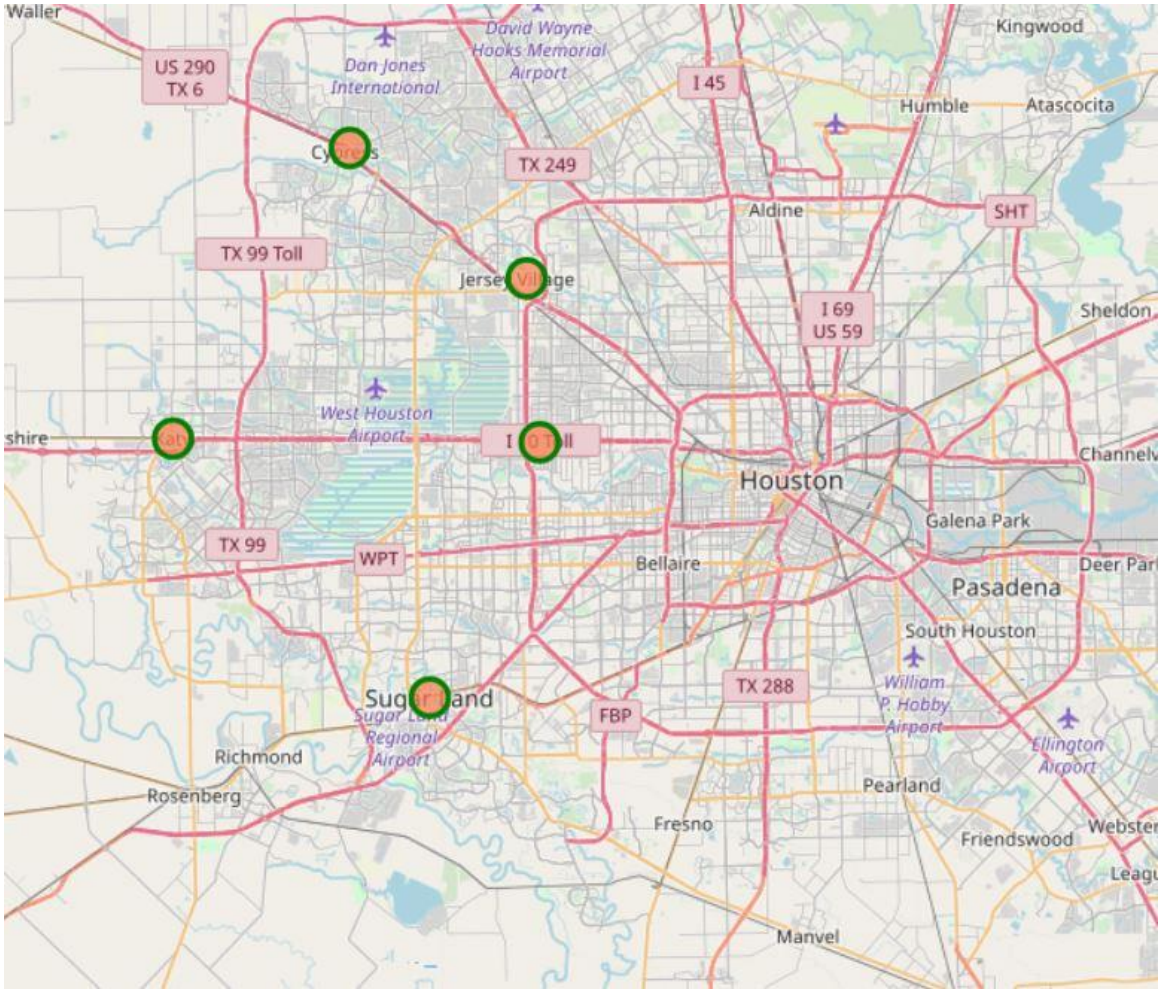
Potential people who could utilize the results of this project are entrepreneurs interested in opening a restaurant. However, the target audience would be the investors or lending institution. Showing them that research has been done to make a data driven decision on the location of the future business shows the investors that decisions aren't made on a whim and that logical thought has been put into the decision in such a way that the business would be most profitable.

- ## Data

For this project, Foursquare data was utilized to generate location data. In this project, I will focus on the following suburban areas of Houston

1. Memorial City
2. Katy
3. Cypress
4. Jersey Village
5. Cypress

For clarity, the latitude and longitude coordinates of each suburban area were acquired and transferred into a pandas dataframe. To visualize the locations of the suburban areas with respect to each other, I created a map of Houston and plotted each suburban area.



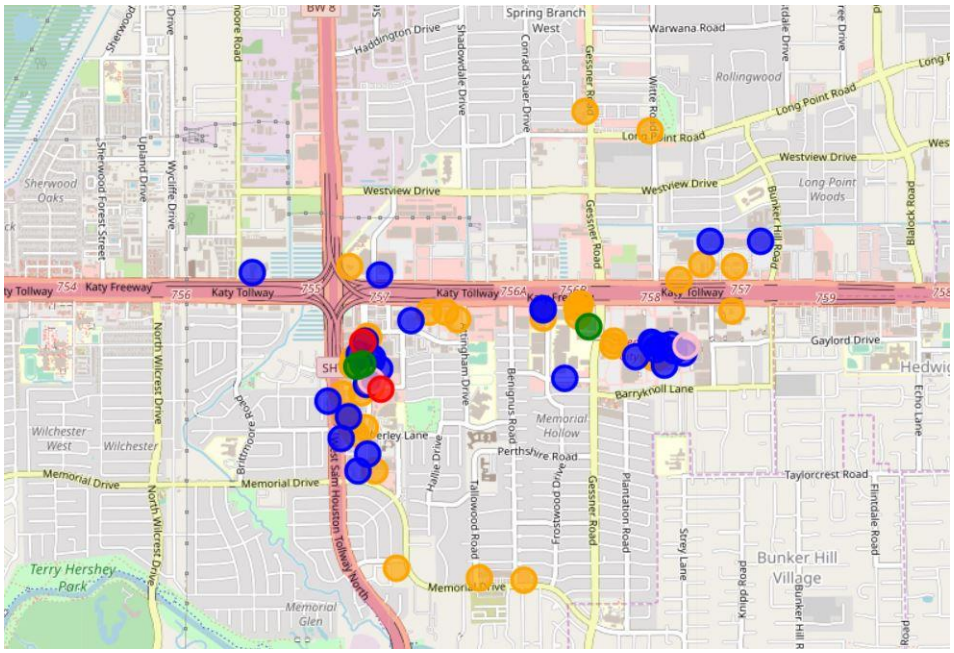
When running the Foursquare API to acquire the location data, it returns several aspects of data for each call. I filtered out the categories, latitudes, and longitudes of each venue in order to clean the data. The cleaned data was transferred to a dataframe, an example of which can be seen below.

:

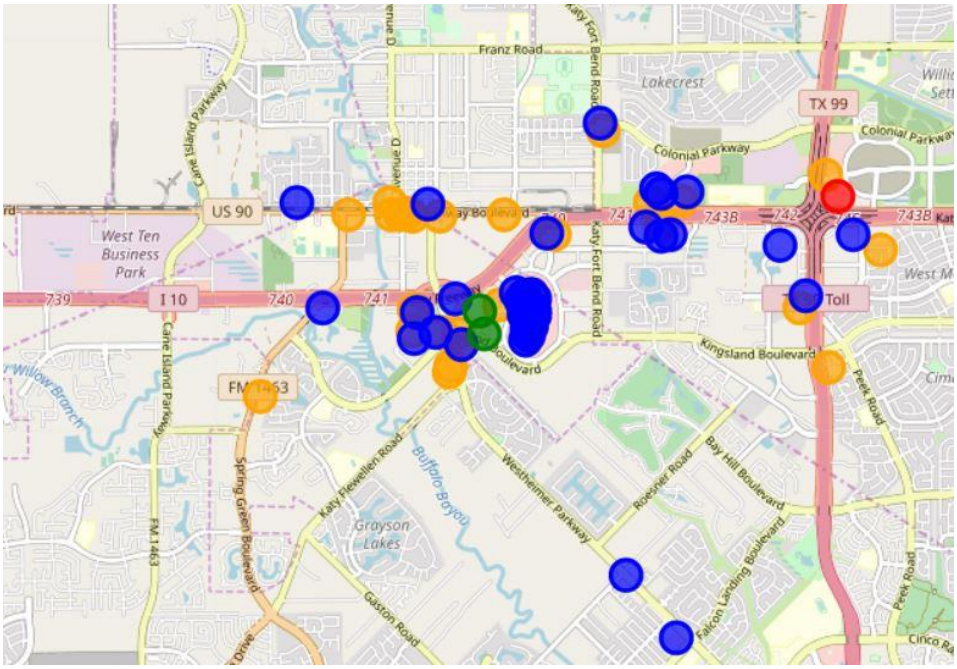
	name	categories	lat	lng
0	Fadi's Mediterranean Grill	Mediterranean Restaurant	29.782859	-95.554195
1	Taste of Texas	Steakhouse	29.782863	-95.556864
2	Pappadeaux Seafood Kitchen	Seafood Restaurant	29.783318	-95.556251
3	Pappasito's Cantina	Mexican Restaurant	29.783314	-95.555078
4	Life Time Athletic	Gym	29.779196	-95.558074

For each of the five suburban areas, a separate dataframe was created for each category of interest. These categories of interest included other restaurants, beauty salons, movie theaters, hotels, and stores and shops. For each suburban area, these venues were plotted on the following individual maps.

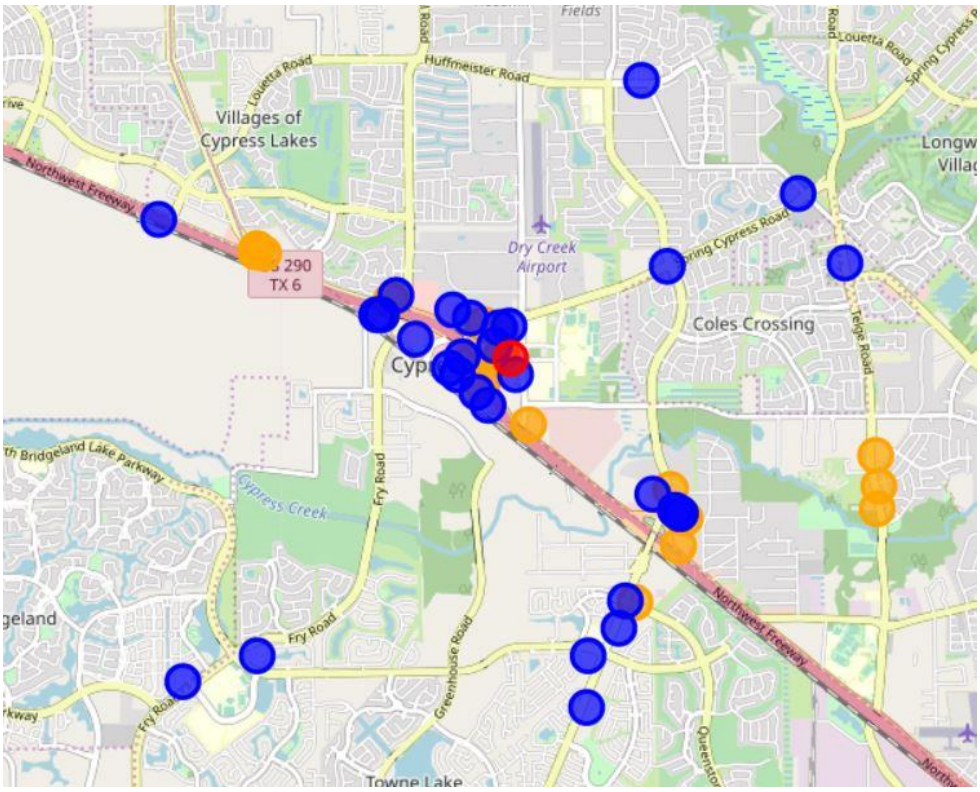
Memorial City



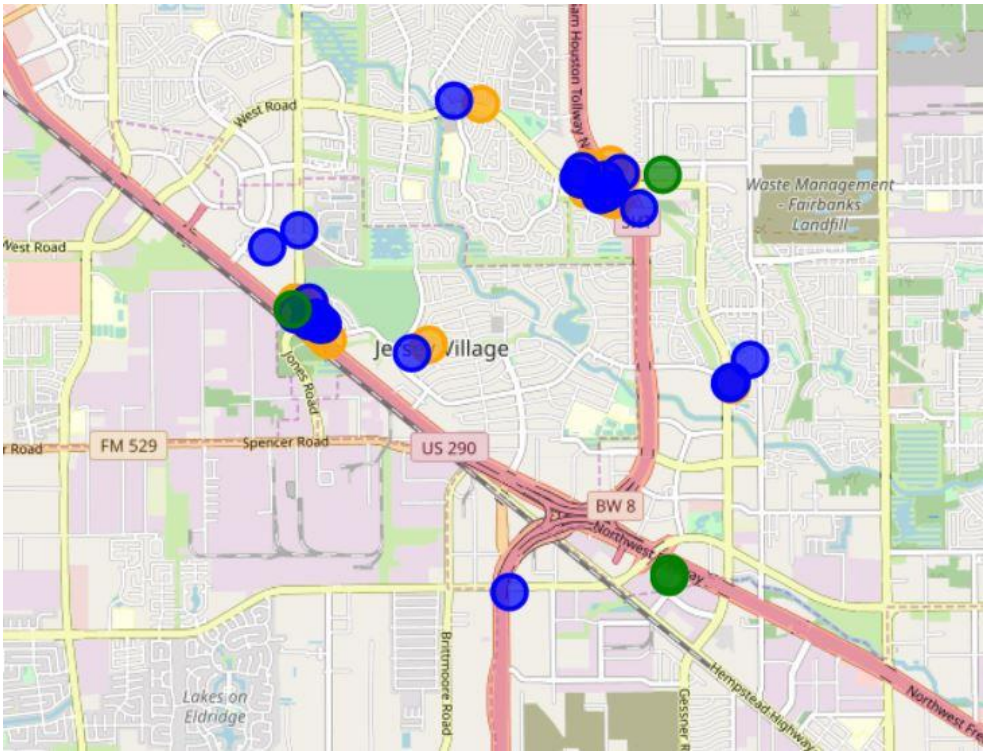
Katy



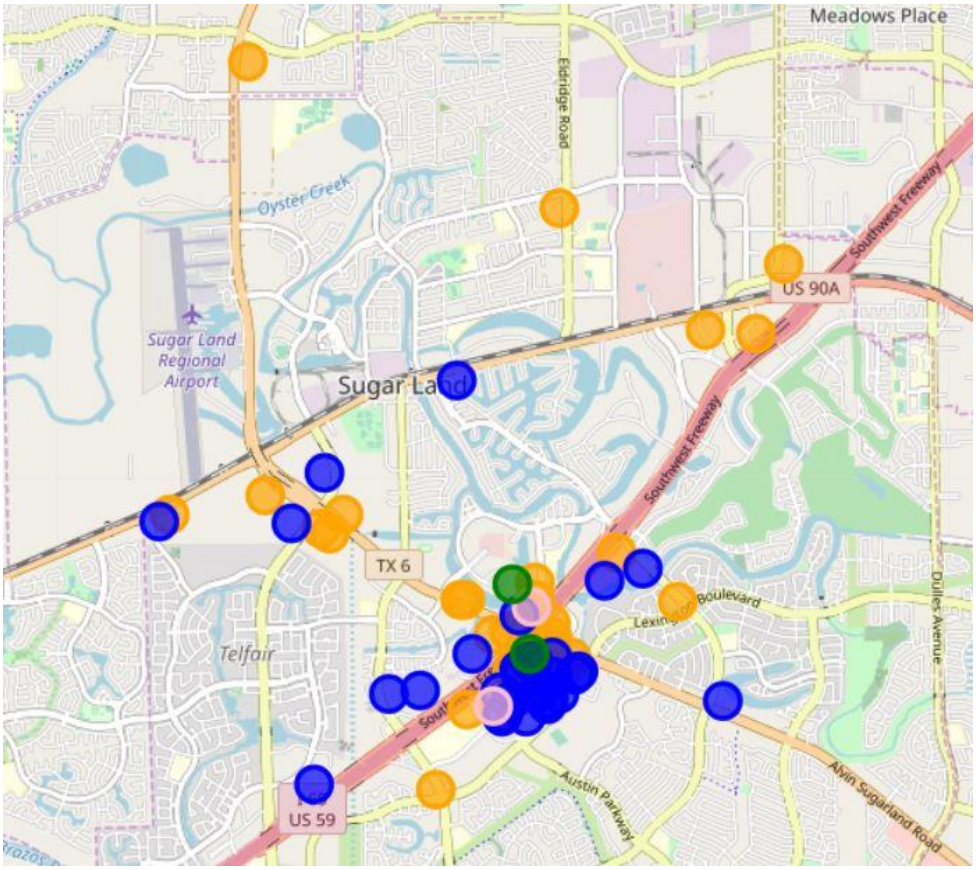
Cypress



Jersey Village



Sugar Land



These quantities of these locations were added to the following dataframe, with "other venues" meaning all shops/stores, movie theaters, hotels, and beauty services.

	Area	Number of Restaurants in Suburban Area	Number of Other Venues in Suburban Areas
0	Memorial City	38	36
1	Katy	27	43
2	Cypress	18	31
3	Jersey Village	21	28
4	Sugar Land	36	33

The quantities and visual representation of all venues in the suburban areas will give considerable insight into which area is best suited for opening a new restaurant and aid the target audience in making a well informed and data driven decision.

- ## Methodology

While there are many more suburban areas in the city of Houston, the five discussed in this report were chosen so that they were far enough apart, and the same venues wouldn't pop up for multiple suburban areas. Also, as seen in the corresponding notebook with the code used in this project, the radius of each area was selected based on the relative size of each suburban area. This was done to ensure that all the venue and location data was included and not left out due to not being in the set radius, but still within the area.

Because I wanted to view how many other restaurants were in the area, venues that had the word "restaurant" in their category label were transferred to a restaurants dataframe. Other categories of interest in this project were venues that people might visit either just before or right after eating. For simplicity, the venues analyzed were stores/shops, beauty salons, movie theaters, and hotels.

For each of the previously shown maps, the colors orange, blue, pink, red, and green are representative of restaurants, stores/shops, beauty services, movie theaters, and hotels, respectively. To determine the best location to open a restaurant, it would be important to not only have a list of these venues, but also maps of them. It would be unwise to open a restaurant in close proximity to several other restaurants, even if it was near to a movie theater or hotel, etc.

- ## Results and Discussion

Quantifying the numbers of restaurants and other venues in these suburban areas helps determine which area is ideal for a new business. The number of restaurants in the area will determine how much competition is in the area. Opening a restaurant near stores or hotels would be wise because potential customers won't have to go far in order to purchase a meal.

The area with the fewest number of restaurants is Cypress, while the area with the most stores/shops, hotels, theaters (other venues) is Katy. Upon analysis of the maps shown in the data section, although Katy has the most locations in our "other venues" category, it would not be wise to open a restaurant near these. This is due to the fact that the location of all the shops/stores, hotels, theaters, and beauty services are very saturated with restaurants. Opening a restaurant in Katy solely based on the fact that it has several other venues would not be the best decision because these locations are already crowded and clustered with several other restaurants.

The Jersey Village map shows that this area could be a potential place to open a successful business, mainly because there aren't many types of venues that we're analyzing in the area to begin with. However, the lack of data does not always deliver the best results, it would be analogous to a shot in the dark. Out of the five areas, Sugar Land has the second highest number of restaurants in this area. While Sugar Land does have several venues in our "other

venues" category, these are also surrounded by many restaurants in the area, making it an area of high competition for a new eatery. Memorial City and Sugar Land are similar in the fact that they both have a large amount of restaurants (Memorial City has the largest number of restaurants in the area) and the restaurants surround many of the hotels, theaters, stores/shops, etc. This would mean Memorial City is also not the best fit to ensure a successful and thriving business.

Analysis of the Cypress map, not only shows the fewest number of restaurants in all of the suburban areas of interest, but many of the stores/shops, hotels, and theaters are distanced from other restaurants in Cypress. The location data acquiring using Foursquare API clearly indicates that Cypress would be the best out of the five suburban areas to open up a restaurant.

Upon completion of this project, the analysis has been to discovered to have room for improvement in various areas. When acquiring the location data, the number of restaurants was chosen based on if the category column explicitly contained the word "restaurant". However, there were discovered to be venues that were sports bars, steakhouses, or burger joints. While these venues are also places where people can eat, they did not show up in our filtered column because they did not contain the phrase "restaurant". Factoring for these possibilities in the future could alter the location data and potentially change the results.

Another way to improve this model is to also obtain tips from the Foursquare API data. It would be a good idea to open a restaurant around venues that had good tips and very positive ratings. These venues are very popular and attract a lot of potential customers to these areas.

- ## Conclusion

In conclusion, it appears that Cypress would be the best choice to open a restaurant. This choice is driven by location data of various suburban areas in West Houston. Categories of venues in these areas are extracted by key words and analyzed in order to determine which area had higher chance of success. Not only were the numbers of these venues considered, but also the location of them with respect to each other. This analysis could be improved by including other types of eateries other than explicitly "restaurants". However, this analysis of location data provides very valuable insight to potential stakeholders, investors, and business owners.