Study of Asian Food Restaurants in Houston, Texas
for a Potential Expansion of a Vietnamese Restaurant Chain

I. Introduction

A Vietnamese restaurant chain is expanding its business and is planning to bring its restaurants to Texas. As a first step, the company wants to open two restaurants and has selected Houston as the first Texas city to enjoy its delicious cuisine. However, the Greater Houston area is a vast geographical area and management is uncertain on what are the best locations for the initial two restaurants. Moreover, the city has plenty of other type of Asian food options to offer which would represent a significant competition for a new Vietnamese restaurant chain.

The goal of this study is to determine the best locations for the proposed restaurants and make a recommendation to the restaurant chain management. This study investigates how many Vietnamese restaurants the city already offers, how many restaurants of other type of Asian food are in the city, and how concentrated they are located in the city. It analyzes different areas in the Greater Houston area and compares them. The Houston areas selected are Cinco Ranch, Sugarland, Galleria, Pearland, Cypress, Jersey Village, Greenspoint, The Woodlands, Kingwood, and Channelview.

II. Data used in this study

Data has been collected from two sources. The latitude and longitude coordinates of each Houston area studied was obtained from Latitude (https://latitude.to), and the information and location of the Asian food restaurants in each area was obtained from Foursquare.

Latitude
Latitude does not provide data by an API but offers a search tool that returns the coordinates for addresses, popular areas, and landmarks. All ten selected areas of Houston are included in the Latitude database, and since the number of areas in this study is small, the coordinates were obtained manually using the search feature and stored in a .csv file which was read into a Pandas data frame in the study.

Foursquare
To collect the information about Asian food restaurants in each area, an API call to Foursquare for each was made, passing the coordinates obtained from Latitude. The API calls used the explore endpoint and asked Foursquare to provide venues in the Vietnamese, Chinese, Thai, Japanese, and Korean food categories only. This filtering was applied by passing the category ID for each of these types of Asian food. The data returned by Foursquare was placed in a Pandas data frame for analysis.

III. Methodology

The study focuses on finding the answers to the following questions:
- How many Vietnamese food restaurants are in each area
- How many restaurants of other major types of Asian food restaurants are in each area
- How each area compares to the other areas
- How are the restaurants geographically located in each area

Special consideration was put on how many Vietnamese restaurants currently operate in each area, but the study also investigated how many restaurants of other type of Asian food exist in each area. The other types of Asian food restaurants that were analyzed are Chinese, Thai, Korean, and Japanese.

Competition Index

Not all these types of Asian food restaurants pose the same competition for a new Vietnamese restaurant, so each one was assigned a different weight. The study refers to this weigh as the Competition Index or CI for short. Chinese and Thai food restaurants are considered to pose more competition to a Vietnamese restaurant than Korean and Japanese restaurants and were assigned a CI of 2. Japanese and Korean food restaurants were assigned a CI of 1, and existing Vietnamese restaurants which pose a direct competition to the proposed new venues were assigned a CI of 3.

Data Collection and Preprocessing

The first step was to collect the coordinates of the Houston areas selected for this study. The data was saved in a .csv file and read into a Pandas data frame. Below is a map of Houston showing the ten areas.
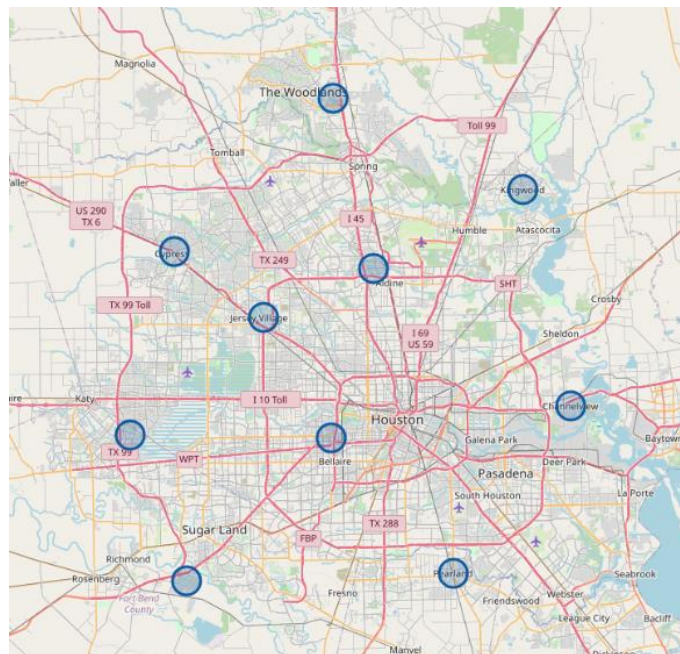

Figure 1. The ten Houston areas subject of this study.

Next, using the coordinates for each area, the venue category IDs for each type of Asian restaurant, and a radius of 8000 meters, API calls were made to Foursquare and the results were placed in a data frame and some processing was done to the data. The following tasks were performed during this processing:

- Foursquare returned some sub-categories and these were merged with the categories selected. Hunan restaurants were changed to Chinese restaurants and Sushi and Ramen restaurants were changed to Japanese restaurants.

- All the category names include the word "restaurant" which seemed redundant, so the word "restaurant" was stripped from all category names.

- A column for the Competition Index (CI) was added with a default value of 0.

- The CI value was changed for each type of restaurant. Vietnamese restaurants were assigned a value of 3, Chinese and Thai restaurants a value of 2, and Japanese and Korean restaurants a value of 1.

- After changing the CI values, the remaining rows that still had the default value of 0 were deleted.

The resulting data frame contains 566 restaurants and its first 5 rows are shown below.

	Area	Area Lat	Area Lng	Restaurant Name	RestLat	RestLng	Category	CI
0	Channelview	29.77606	-95.11465	Bamboo China Cafe	29.806960	-95.168218	Chinese	2
1	Channelview	29.77606	-95.11465	Bibo's Cafe	29.824889	-95.167574	Vietnamese	3
2	Channelview	29.77606	-95.11465	Panda Express	29.771835	-95.186176	Chinese	2
3	Channelview	29.77606	-95.11465	Osaka Japanese Steakhouse	29.806678	-95.167649	Japanese	1
4	Channelview	29.77606	-95.11465	Tam's Vietnamese Sandwich & Noodle Shop	29.801011	-95.162196	Vietnamese	3

Figure 2. First five rows of the data frame with all 566 results returned by Foursquare.

Grouping and Clustering

The next step was to group the restaurants by area and two new data frames were created to better make sense of the data. The first one is a pivot data frame showing the number of restaurants in each category for each area. The second data frame shows the total number of restaurants and the total CI for each area.

	Area	Chinese	Japanese	Korean	Thai	Vietnamese
0	Channelview	7	4	0	0	3
1	Cinco Ranch	44	29	9	4	14
2	Cypress	17	16	0	5	6
3	Galleria	29	35	0	11	20
4	Greenspoint	16	9	3	1	15
5	Jersey Village	33	12	2	11	41
6	Kingwood	15	11	3	2	5
7	Pearland	13	3	0	0	15
8	Sugarland	17	22	2	5	8
9	The Woodlands	16	20	1	5	7

	Area	Restaurants	CI
5	Jersey Village	99	225
1	Cinco Ranch	100	176
3	Galleria	95	175
8	Sugarland	54	92
4	Greenspoint	44	91
9	The Woodlands	49	84
2	Cypress	44	78
7	Pearland	31	74
6	Kingwood	36	63
0	Channelview	14	27

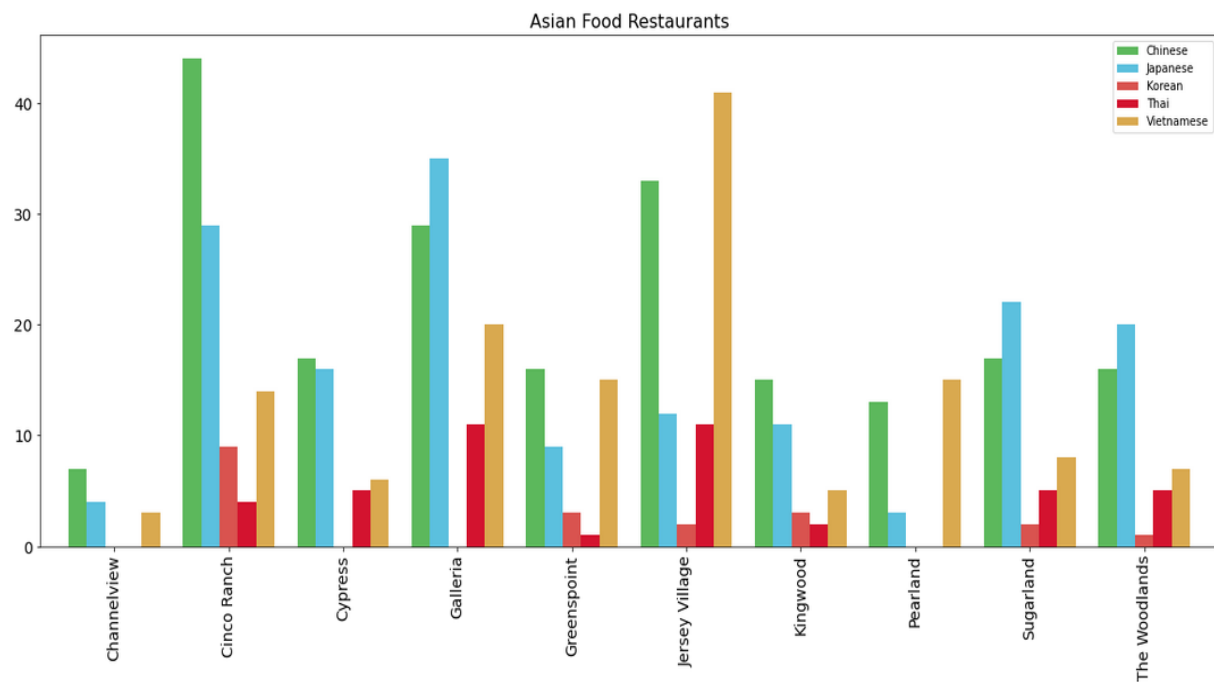Figure 3 and 4. Tables showing the total number of restaurants in each area.

Figure 5. Graph showing the distribution of different types of Asian food restaurants in each area.

Finally, to further compare each area, segmentation was performed using K-Means Clustering. The results are shown in the data frame below:

	Cluster Labels	Area	Chinese	Japanese	Korean	Thai	Vietnamese
0	0	Cypress	0.386364	0.363636	0.000000	0.113636	0.136364
1	0	Galleria	0.305263	0.368421	0.000000	0.115789	0.210526
2	0	Sugarland	0.314815	0.407407	0.037037	0.092593	0.148148
3	0	The Woodlands	0.326531	0.408163	0.020408	0.102041	0.142857
4	1	Greenspoint	0.363636	0.204545	0.068182	0.022727	0.340909
5	1	Jersey Village	0.333333	0.121212	0.020202	0.111111	0.414141
6	1	Pearland	0.419355	0.096774	0.000000	0.000000	0.483871
7	2	Channelview	0.500000	0.285714	0.000000	0.000000	0.214286
8	2	Cinco Ranch	0.440000	0.290000	0.090000	0.040000	0.140000
9	2	Kingwood	0.416667	0.305556	0.083333	0.055556	0.138889

Figure 6. Data frame showing the clustering.

IV. Results and Observations

At a first glance, after looking at the data, Channelview seems to be the best place where to open a new Vietnamese restaurant. This is the area with the lowest number of Asian food restaurants and with the lowest CI which seems to indicate that a new Vietnamese restaurant will face little competition. However, the very low number of Asian foods restaurants compared to other areas could mean that Channelview is not an area where Asian food restaurants are very successful. It could be that Asian food in general is not that popular here or there could be some demographic characteristics articular to this area that makes it unattractive to Asian food restaurant ventures.

On the other end of the spectrum, Cinco Ranch, Jersey Village, and Galleria are the areas with the most Asian food restaurants. Moreover, the number of Asian food restaurants in these areas is disproportionately higher than in other areas, except for Galleria. With 100, 99, and 95 Asian food restaurants respectively, Cinco Ranch, Jersey Village, and Galleria have about double the number of restaurants than the seven other areas. Jersey Village stands out for the high percentage of Vietnamese restaurants. Here, Vietnamese restaurants account for 41% of all Asian food restaurants. Jersey Village is the place where a new Vietnamese restaurant would face the highest competition. Cinco Ranch is in a distant second place in terms of high competition and closely behind in third place is Galleria. In all these three areas the competition for a new Vietnamese restaurant is significantly higher than in the rest of the pack.

Therefore, Channelview, Cinco Ranch, and Jersey Village are probably not good places where to open a new Vietnamese restaurant. In Channelview there seems not to be much interest in Asian food in general and in Cinco Ranch, Jersey Village, and Galleria the market is saturated and there is too much competition.

That leaves six areas to chose from. The best place should be the area where Asian Food is popular, yet the percentage of Vietnamese restaurants is low. All six areas are similar in terms of number of Asian food restaurants, but the percentage of Vietnamese restaurants over the number of Asian food restaurants in these areas varies significantly. Therefore, for these areas the study focused on this percentage. In Pearland 48% of Asian food restaurants are Vietnamese restaurants, while in Greenspoint this percentage is 34% . In all the remaining four areas this percentage is between 13% and 14%, significantly lower than in Pearland and Greenspoint. Cypress, Kingwood, Sugarland, and The Woodlands are remarkably similar, and they are the best laces for a new Vietnamese restaurant.

V. Conclusion

Of all the ten areas studied, Cypress, Kingwood, Sugarland, and The Woodlands were found to be the best places where to open a new Vietnamese restaurant. Since these four areas are so similar, selecting the best area of the four is not easy. In Cypress and Kingwood, the percentage of Vietnamese restaurants of the total number of Asian food restaurant is slightly lower than in Sugarland and the Woodlands. For this reason, this study found that Cypress and Kingwood are the best locations for a new Vietnamese restaurant. Since the restaurant chain is planning to open two restaurants, it will be recommended to the chain leadership that one restaurant should be open in Cypress and one in Kingwood.

This study assigned different weights in terms of competition to competing types of Asian foods and called this weight the competition index or CI. This parameter did not help much in the end. It did not add any new information to the analysis. Of much better use resulted to be the percentage of Vietnamese restaurants from all Asian food restaurants in each area. This data was calculated during clustering and proved to be very useful.

The maps below show the Asian food restaurants in the two recommended areas. The blue dots indicate Vietnamese restaurants and the red dots indicate all the other types Asian food restaurants studied.
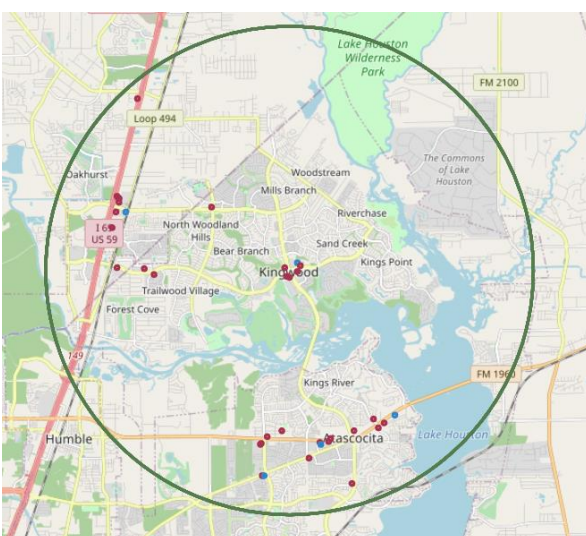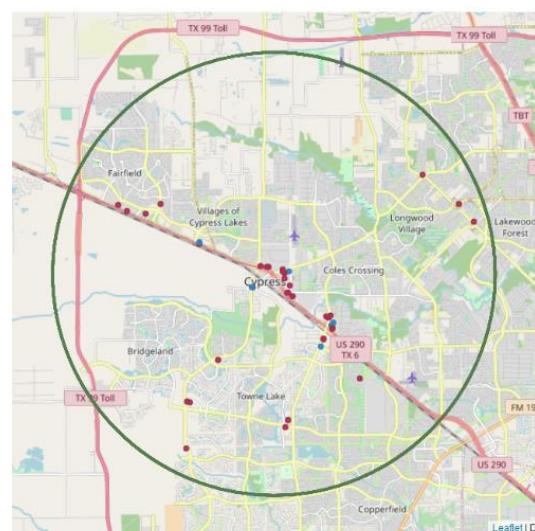

Figure 7. Asian food restaurants in Kingwood.

Figure 8. Asian food restaurants in Cypress.