Study of Asian Food Restaurants in Houston

for a Potential Expansion of a Vietnamese Restaurant Chain

Applied Data Science Capstone Project

Andres Lopez, May 2020

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.

The goal of this study is to determine the best locations for the proposed restaurants and make a recommendation to the restaurant chain management.

Data and Methodology

The study focuses in ten areas in Houston and five types of Asian food restaurants:

Cinco Ranch, Sugarland, Galleria, Pearland, Cypress, Jersey Village, Greenspoint, The Woodlands, Kingwood, and Channelview.

Types of Asian food restaurants:

Vietnamese, Chinese, Thai, Japanese, and Korean

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.

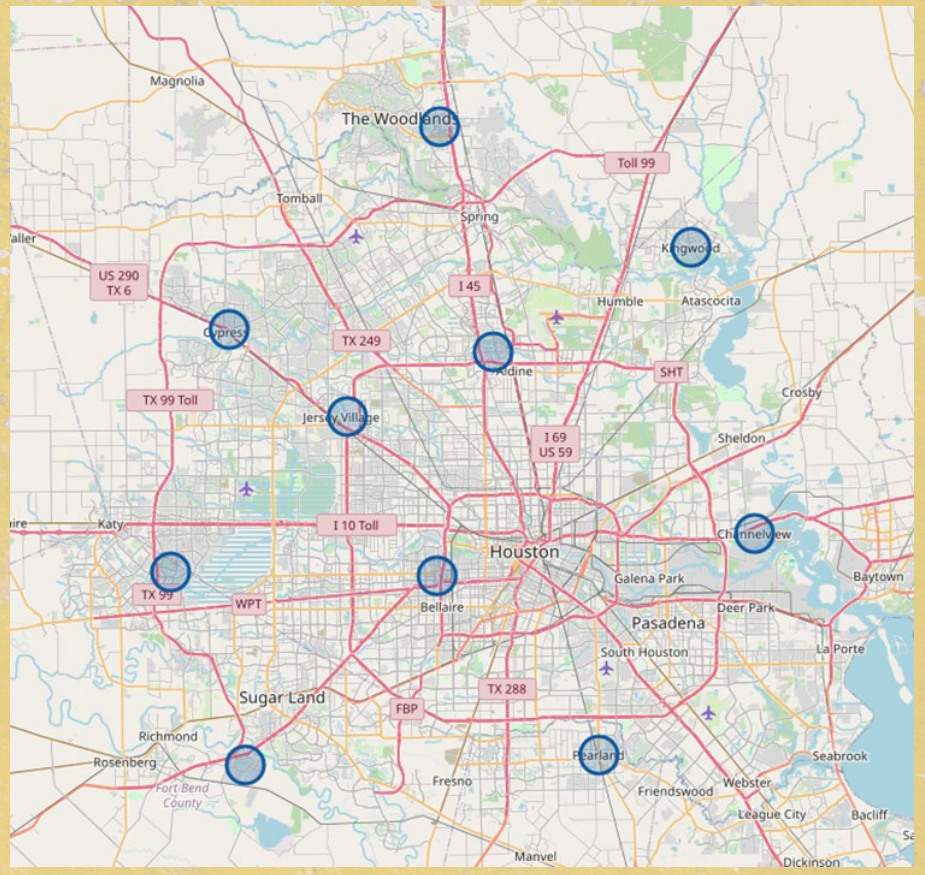

Figure 1. The ten Houston areas subject of this study.

Data and 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

Competition Index

Not all these types of Asian food restaurants pose the same competition for a new Vietnamese restaurant, so each type was assigned a different weight from 1 to 3, called 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 and Methodology

Preprocessing

The data returned by Foursquare was placed in a Pandas data frame and several tasks were performed:

- 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.

Data and Methodology

Foursquare returned a total of 566 restaurants in all ten Houston areas.
The first 5 venues 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.

Data and Methodology

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.

Data and Methodology

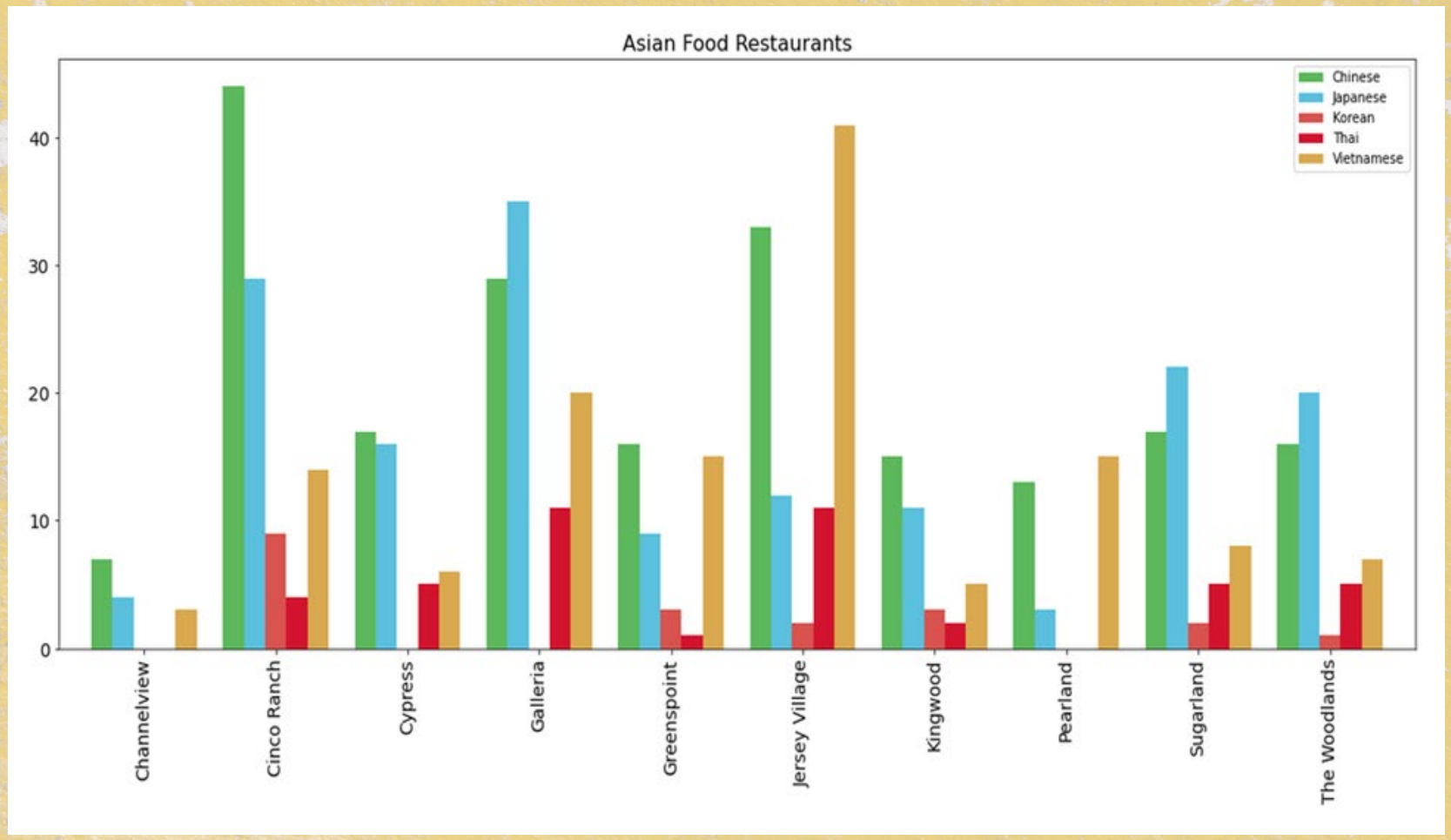

Figure 5. Graph showing the distribution of different types of Asian food restaurants in each area.

Data and Methodology

	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.

Results and Observations

By looking at the data and comparing the areas, the study started to rule areas out until only the bests areas remained. Aras were ruled out in a series of disqualification rounds.

First round – Channelview, Cinco Ranch, Jersey Village, and Galleria

Channelview has too few Asian food restaurants compared to all other areas and was considered an outlier.

Cinco Ranch, Jersey Village, and Galleria have too many Asian food restaurants compared to other areas. The market here seems saturated, especially in Jersey Village were the percentage of Vietnamese restaurants is very high (41%).

Second Round - Pearland and Greenspoint were ruled out because of the six areas left, their percentage of Vietnamese restaurants, 48% and 34% respectively, is much higher than in the other four areas where this percentage is between 13% and 14%.

Third Round - Sugarland and The Woodlands were ruled out because of the four areas left, even though they are very similar, Sugarland and The Woodlands have slighter higher percentage of Vietnamese restaurants than the other two areas.

Conclusion

After going through a process of eliminating the least suitable areas for a new Vietnamese restaurant, based on the data collected and the corresponding analysis, this study found that **Kingwood** and **Cypress** are the best places to open the new venues.

In these two areas Asian food seems to be popular and the Vietnamese food share of this market is very low which indicates that there is room for more Vietnamese restaurants.

It is the recommendation of this study that the Vietnamese food restaurant chain open its first restaurant in the **Kingwood** area and its second restaurant in the **Cypress** area.

Note: 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.

Conclusion

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

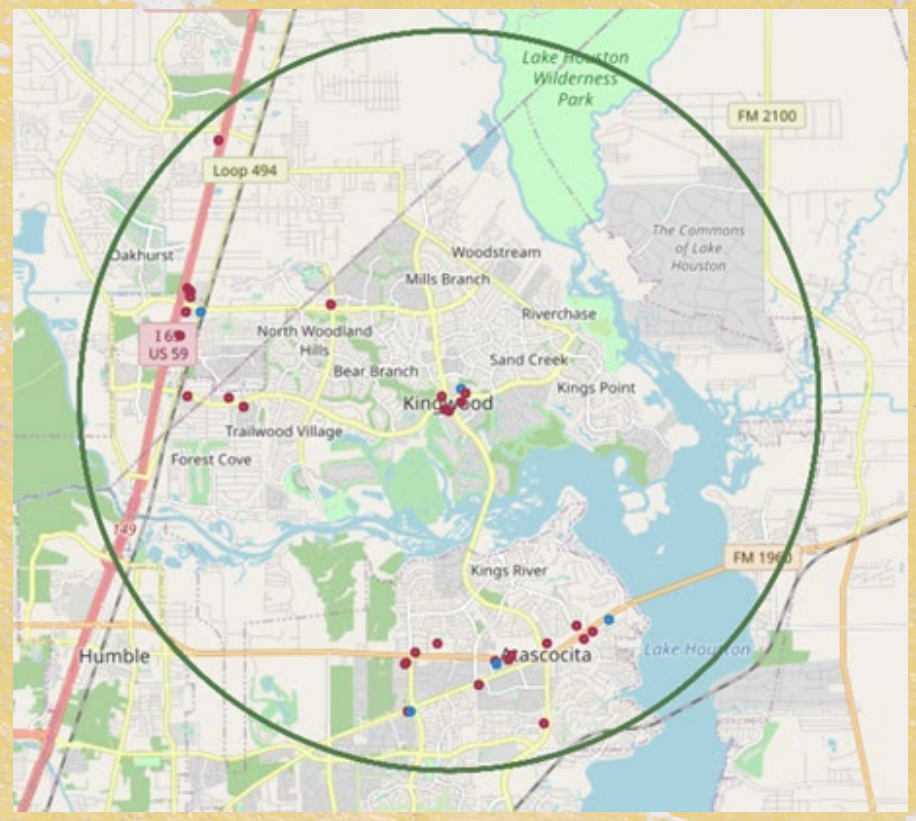

Figure 7. Asian food restaurants in Kingwood.

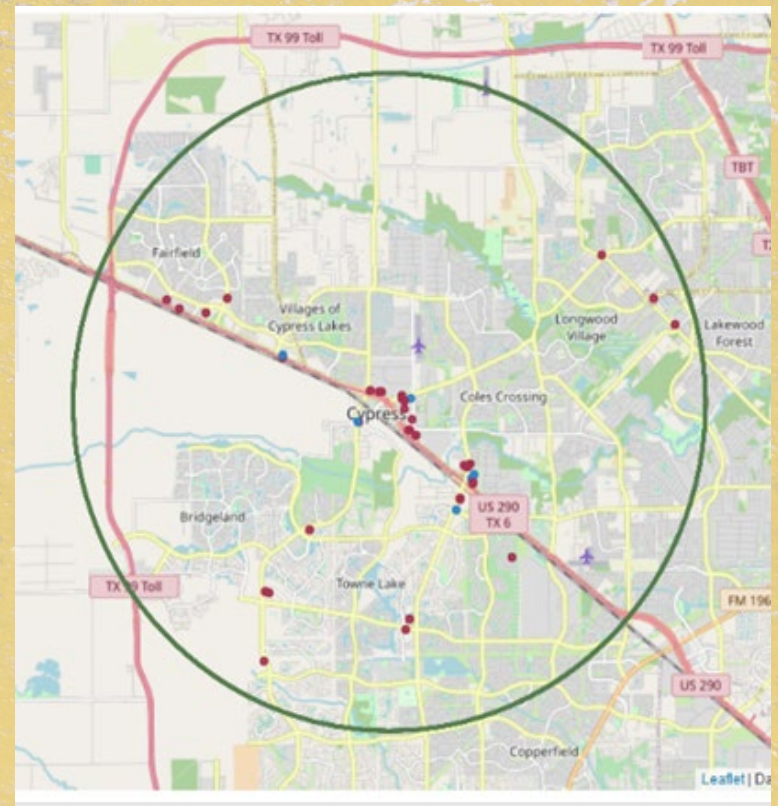

Figure 8. Asian food restaurants in Cypress.

THE END

Applied Data Science Capstone Project

Andres Lopez, May 2020