Final Capstone Project Report

1. Introduction to Business Problem

No one would disagree that restaurant is one of our favorite places of interest. They play such an important role in our daily life as places our family gathering around and where we hangout with friends, or enjoy the gourmets.

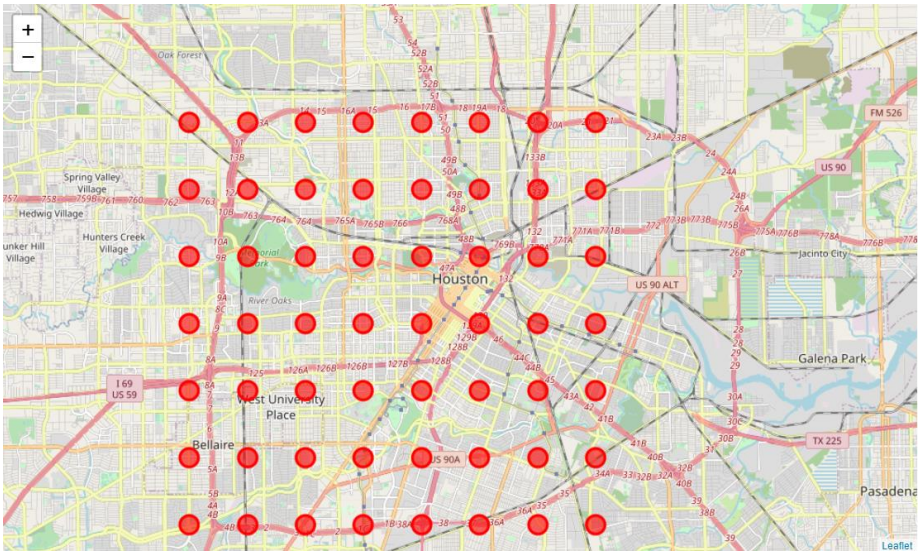
We all have experiences that we are most likely to find groups of restaurants surrounding a place with high population density, such as outlet shopping malls or subway station. But in our neighborhood areas, what will be the “hubs” of restaurant?

In this project, we are trying to look for the answers that compare to popular point of interest, such as shopping malls, how restaurants gather around our neighborhood venues, like Groceries stores and cinemas. What’s is the minimum distance one restaurant closes to its hub? The research objective area we picked is greater Houston are in Texas.

2. Data Requirement

The analysis of this project mostly relies on the geographical location data of each venues, simply speaking, the latitude and longitude data of point of interest. Thanks to the data serve from FourSquare, we are able to pull required data for each restaurant. However, there’s a limit of 50 results from each data pulling and we need much more restaurant geographical data for this project. We have to developed a search matrix as below, creating a grid contains 64 base point centers in searching restaurants around each point within a 3 km radius.

Other than the FourSquare database, we also use Geolocator to pullout geographic coordinates from address input. Also, we imported python package of sklearn for later data manipulation and analysis.



3. Data Acquiring and cleaning

From the above searching grid, we managed to acquire massive of restaurants location data. Data engineering includes:

- Creating a function to input search criteria, such as search point latitude/longitude, searching radius and key word; then transpose the Jason structure file to a clean dataframe.
- Removing duplicate search results when restaurants shown up as result between different search centers.
- Removing unrelated data when putting “food” as key word searching on the database, some irrelevant results would come up, e.g. food companies and gas station.

Finally, we have a clean DataFrame of ***262 restaurants*** with wanted geographical data. Moreover, we pulled out data for Groceries stores, shopping malls, and cinema in the area.

Restaurants Data

	id	name	categories	lat	lng
2	5b82fb81123a19002c4cd06b	Heavenly Choices Soul Food	Southern / Soul Food Restaurant	29.799557	-95.332985
8	534aa891498eb5bc1146720f	Helen T's Soul Food Cafe'	Café	29.830334	-95.319173
11	4f325999619836c91c7cf1297	King's Tasty Foods	Food	29.813857	-95.317654
14	5ade261993bd637585162402	El Venado Foods	Food Service	29.803803	-95.340890
7	4f324cb519836c91c7ca12c9	Casa Blanca Mexican Food	Food	29.798847	-95.354043
...	...	...	...	...	...
12	4f43bf2d19834bc91f58a353	Foodarama	Bakery	29.662364	-95.464889
20	553ed75b498e43a4ea90ced2	Whole Foods	Miscellaneous Shop	29.661458	-95.459679
11	55e050f7498e7faa023d0351	Coffee Shop Food Truck	Coffee Shop	29.657413	-95.480209
5	4d4af74ab480a143e92835ec	Tacos D' Mundo Food Truck	Mexican Restaurant	29.657864	-95.288424
4	4e4287c3d1645b30b56b7ed1	Food Store	Shop & Service	29.630280	-95.339554

262 rows × 5 columns

Grocery Store Data

	id	name	categories	lat	lng
4	5e3b4b2f9b92a90008cbfb35	H-E-B Curbside Pickup & Grocery Delivery	Grocery Store	29.768968	-95.396583
6	5e00dc17898a6a0008661fb1b	H-E-B Curbside Pickup & Grocery Delivery	Grocery Store	29.714046	-95.377260
8	59dc14196fa81f4616cc9aa1	H-E-B Curbside Pickup & Grocery Delivery	Grocery Store	29.738034	-95.402704
13	5e3ad71c57fcde0008a94157	H-E-B Curbside Pickup & Grocery Delivery	Grocery Store	29.726999	-95.427071
16	5b802d6fbd4009002c144081	H-E-B Curbside Pickup & Grocery Delivery	Grocery Store	29.707609	-95.468950
18	5e382f3b0d366a000803ca9d	H-E-B Curbside Pickup & Grocery Delivery	Grocery Store	29.688878	-95.464268
25	5c48863ef62e09002c7ef750	The Heights H-E-B	Grocery Store	29.807098	-95.409275
29	5d15b7e31822230026f11900	Kroger Grocery Pickup and Delivery	Grocery Store	29.804629	-95.400582
33	5e5eb9809f640c0007d027e4	H-E-B Curbside Pickup & Grocery Delivery	Grocery Store	29.807351	-95.408957
1	4db85cbef7b15ca52ce8b875	Whole Foods Market	Grocery Store	29.757994	-95.397643
12	4ac0bec7f964a520799420e3	Whole Foods Market	Grocery Store	29.739152	-95.418184
13	544aac19498eead34ecc457e	Whole Foods Market	Grocery Store	29.749480	-95.461635
20	5dbdb8546c2a1c00084168fb	Whole Foods Market	Grocery Store	29.744379	-95.381412
25	589ccbb603e29a788c4a0327	Whole Foods Market 365	Grocery Store	29.813679	-95.398328

Cinema Data

	id	name	categories	lat	lng
7	4b3feb8af964a52013b225e3	Regal Edwards Greenway Grand Palace ScreenX & RPX	Movie Theater	29.730839	-95.440551
8	4a9045c1f964a520f31620e3	Cinemark	Movie Theater	29.768997	-95.229342
9	465978f4f964a5201b471fe3	Regal Edwards Houston Marq*E Screenx, 4DX, IMA...	Movie Theater	29.785704	-95.463748

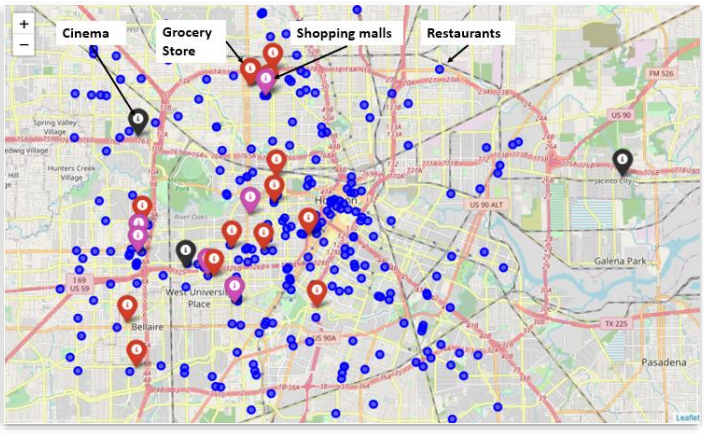
Shopping Mall Data

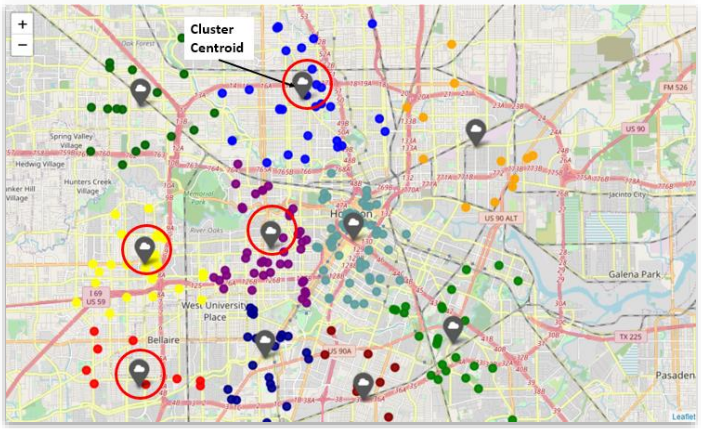
	id	name	categories	lat	lng
1	4e04fb3eb61cf74861301723	River Oaks Shopping Center	Shopping Mall	29.753011	-95.409400
3	50b3d9d9e4b0036cfc3b9d46	West University Shopping Center	Shopping Mall	29.727532	-95.430148
4	42829c80f964a52064221fe3	The Galleria	Shopping Mall	29.736836	-95.463692
5	4eda75737ee5e8e3e6efc62f	19th Street Shopping Strip	Shopping Mall	29.802941	-95.401842
9	4c6963098d22c9288a0cb745	Rice Village	Shopping Mall	29.715885	-95.416903
10	4fabd893e4b034d5347d2871	The Centre At Post Oak	Shopping Mall	29.741921	-95.463199

4. Data Analysis – Clustering Methodology

Now we have all the data needed, it's time to run some analysis. With the help of sklearn package of clustering, we can group the 269 restaurants in greater Houston area into multiple clusters using the K-Mean clustering method.

When we put the cluster map and a venue map with locations of groceries stores, cinema and shopping mall side by side, it is not a coincidence that some of the cluster centroids (red circle) are very close to our presumptive point of places like groceries store.



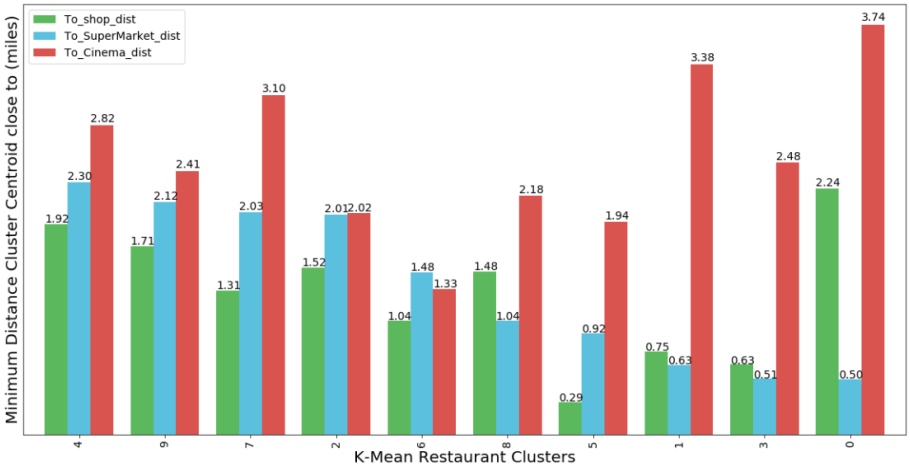


5. Results and Discussion

In our observation, we want to investigate further that how much overlap between the cluster centroid and venues locations we picked. At some clusters, there are multiple venues around its centroid. The below bar chart shows the ***Closest Venue*** to the centroid, distance reported in miles.

Because FourSquare doesn't have enough data of cinema at our designated location, the result of cinema distance has large variances. However, for those cinemas around Cluster "8" and "5", we still consider them as "hubs" for restaurants.

From the result, what surprise us is that our hypothesis of Groceries Shop will be "hubs" of restaurants are preliminary confirmed, as their minimum average distance are within 1.5 miles, very close to a shopping mall distance of 1.3 miles.



	Cluster	To_shop_dist	To_SuperMarket_dist	To_Cinema_dist
count	10.00000	10.000000	10.000000	10.000000
mean	5.50000	1.295941	1.357972	2.544310
std	3.02765	0.610288	0.717208	0.729784
min	1.00000	0.298286	0.508656	1.330073
25%	3.25000	0.828320	0.709160	2.063036
50%	5.50000	1.402927	1.262026	2.449488
75%	7.75000	1.671107	2.028576	3.032687
max	10.00000	2.249473	2.303671	3.746713

6. Conclusion

Our above analysis suggests that grocery store/super market are usually surrounded by restaurants. This result corresponds to our life experience. Most people would love to go out for dinning after grocery shopping at the weekend. We assume the result could be more promising if the database of FourSquare included more small restaurants like we saw inside a super market plaza.

The next step of this project extension would be to explore customer flows of a restaurant based on a distance gradient from its hub. But we will make it as a side project so this is out of the topic of this assignment.

Thank you for your review.