Final Capstone Project

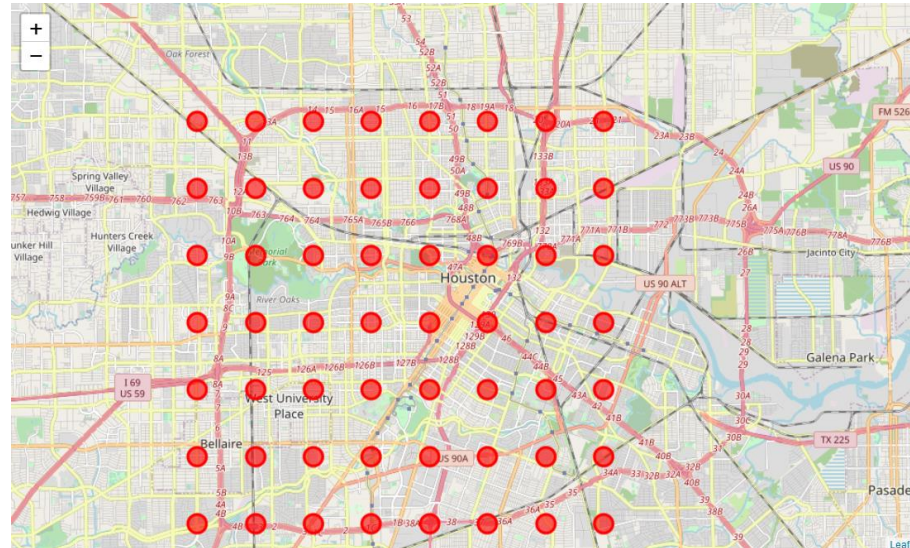A LOCAL RESTAURANT CLUSTERING METHOD

Introduction to Business Problem

In this project, we are trying to map out all restaurants in a defined area then applied the methodology of clustering, finding out centroids of each cluster. By comparing the geographical location of each centroid and other high population destinations such as shopping malls, cinema and groceries stores, we would get a general idea if restaurant business are built around specific categories of hubs.

We don't make certain business proposition in this project. Instead, the following analysis aims at proceeding a preliminary research for business plan.
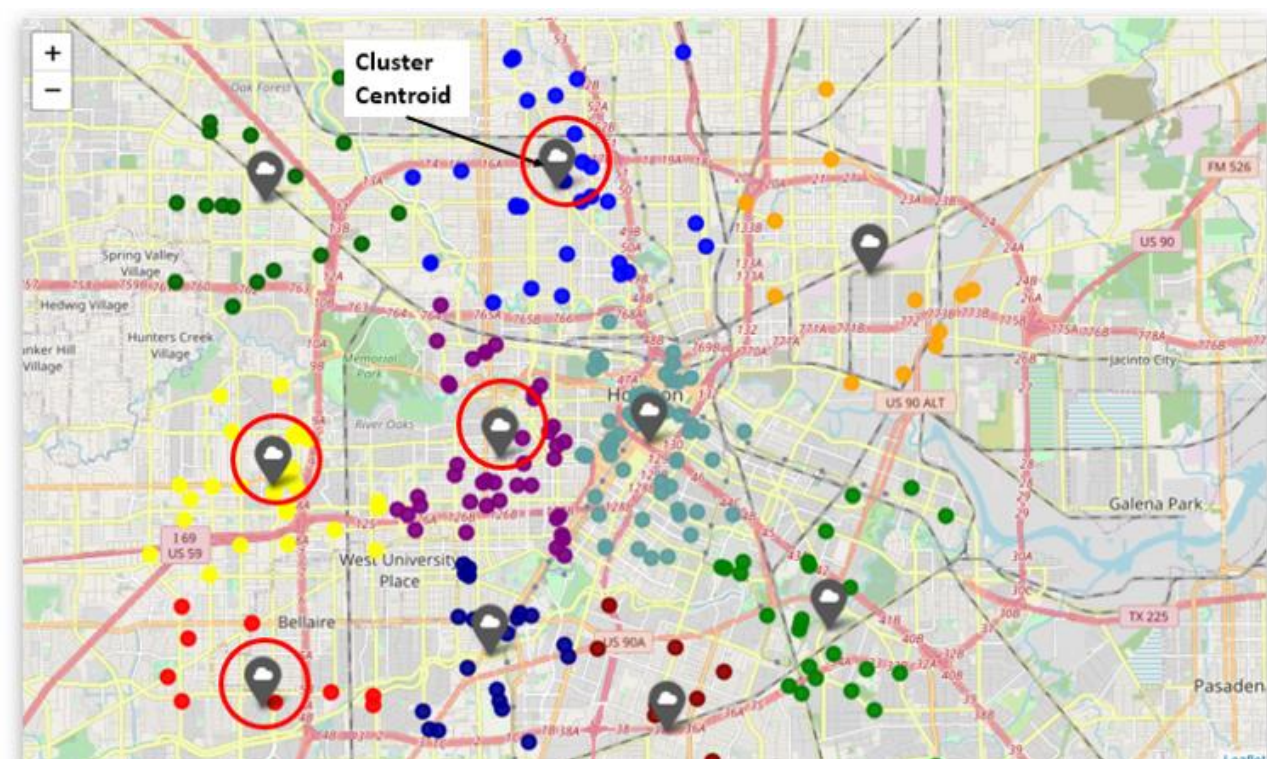
Data Requirement

FourSquare database is used to fetched location data of restaurants and other point of places. A searching grid is created so we can bypass the search limit of 50 results. In this method, we fetched location data of 262 restaurnats.
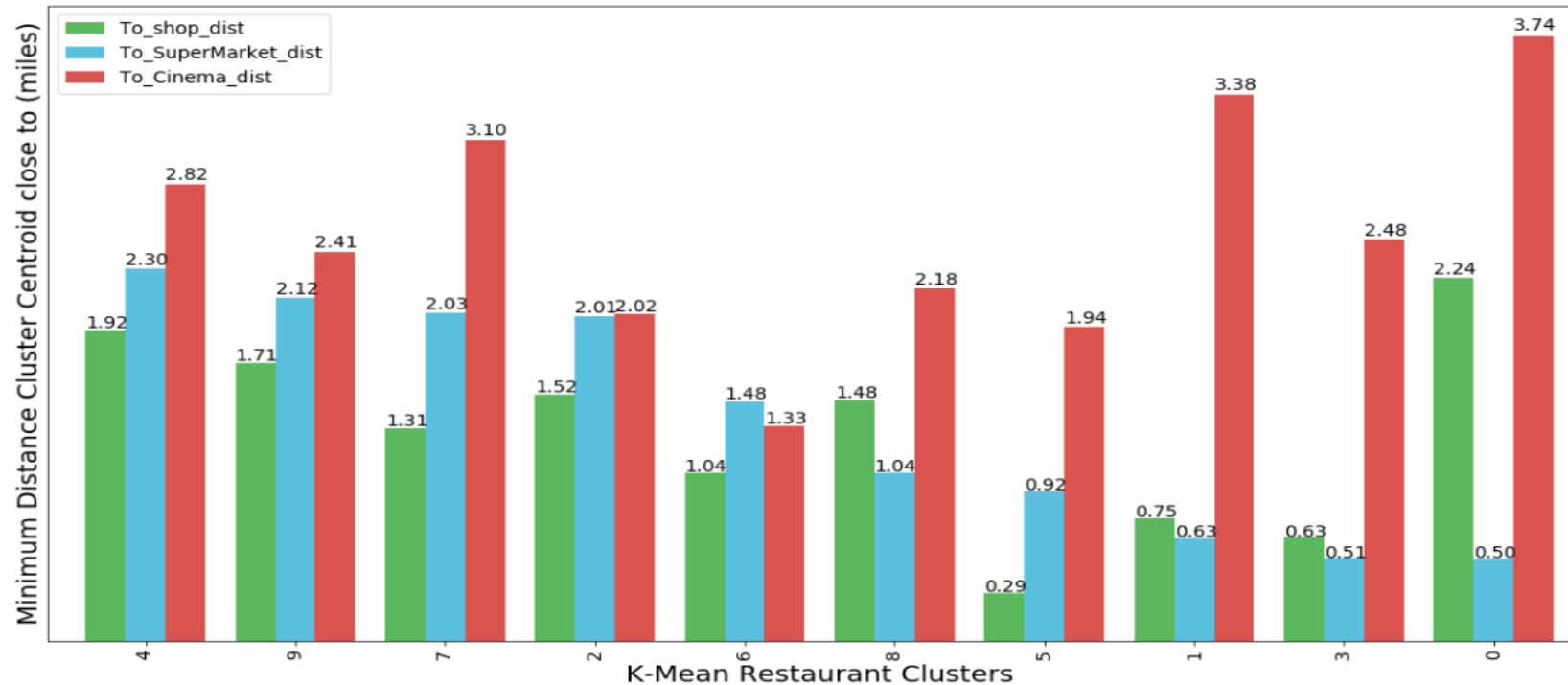
Data Analysis – Clustering Methodology

- Left Figure- Plots out all the restaurants and point of places

- Right Figure – Clustered restaurants and plot out its centroid of each cluster.

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.

Conclusion

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.

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.

	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