Predicting Location of A New Asian Restaurant in Southwest Houston, TX

Applied Data Science Capstone by IBM/Coursera

Deming Wang

Nov. 30th, 2020

1. Business Problem and Background

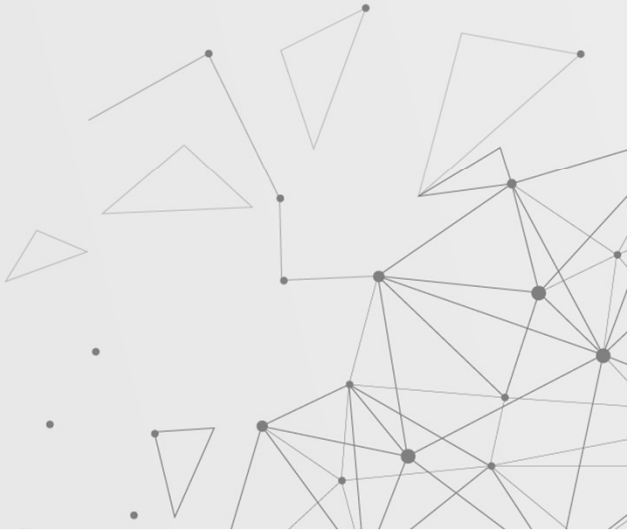
This project aims at analyzing restaurant distribution in West Houston and targets to stakeholders who are interested in opening a new Asian restaurant in Southwest Houston – Sugar Land and Missouri City.

1. Restaurants are one of the most popular business in West Houston. For example, Houston China Town and Katy Asian Town.

2. No similar food town in Southwest Houston (Sugar Land & Missouri City)

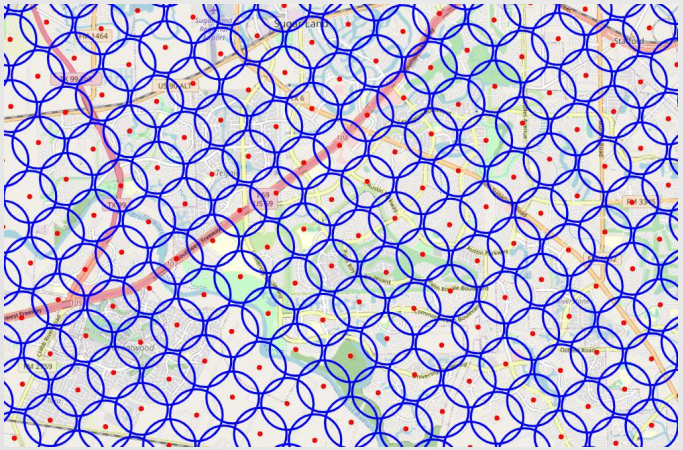
3. Demand for authentic Asian food is expected to remain strong in near future

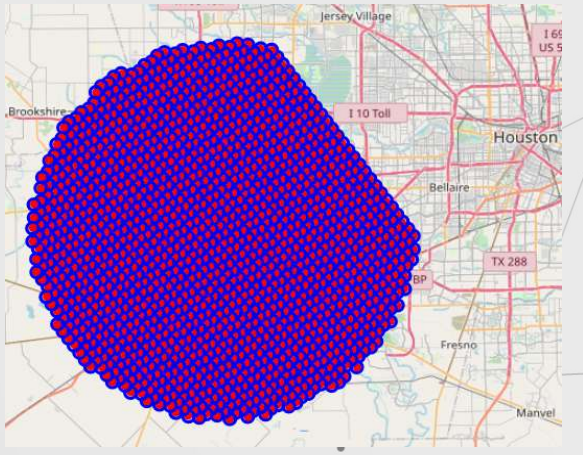
4. Predict potential location for a new Asian restaurant in Southwest Houston area.



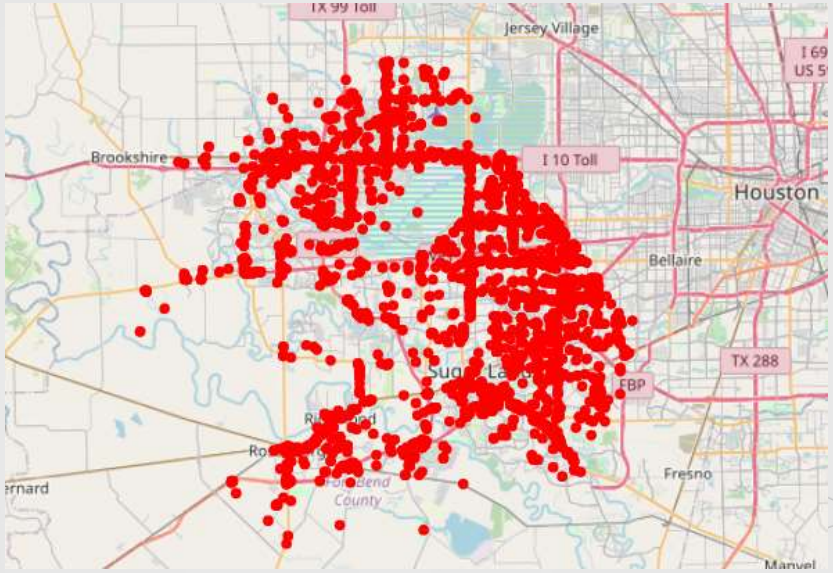
2. Data acquisition and cleaning

- Generate 1534 points around center (Latitude: 29.6589382 Longitude: -95.7276974) in West Houston area
- Use point locations to download all restaurants data using Foursquare API with category under “Food” (id: 4d4b7105d754a06374d81259)



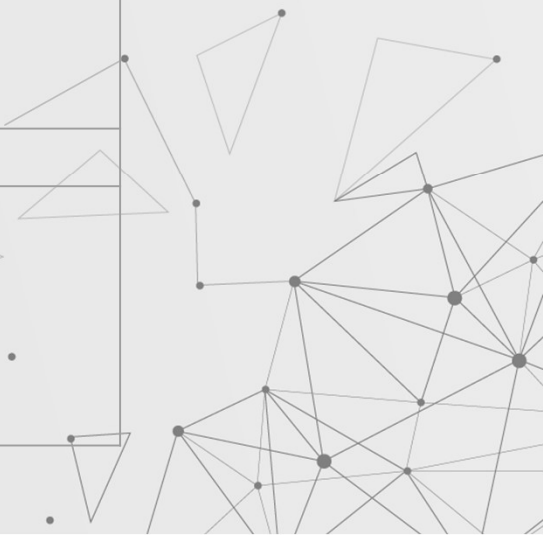


3702 restaurant data downloaded in West Houston

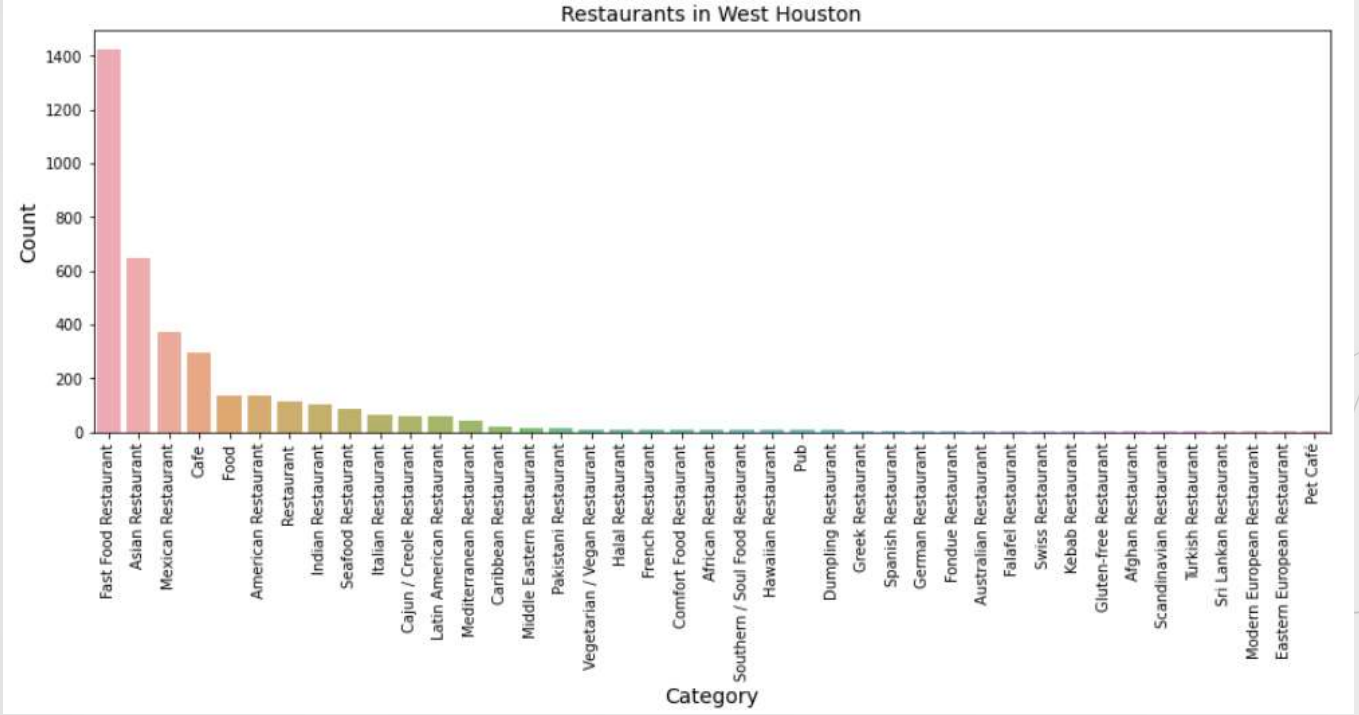


Combine restaurant categories

Original value	New value
[Empty]	Food
Sandwich Place Fried Chicken Joint Pizza Place Fish & Chips Shop Mac & Cheese Joint Breakfast Spot Donut Shop Food Truck Bakery Burger Joint Bistro Wings Joint Bagel Shop Snack Place Hot Dog Joint	Fast Food Restaurant
Irish Pub Gastropub	Pub
Creperie BBQ Joint Buffet Café Deli / Bodega Steakhouse Salad Place Food Court Soup Place	Cafe

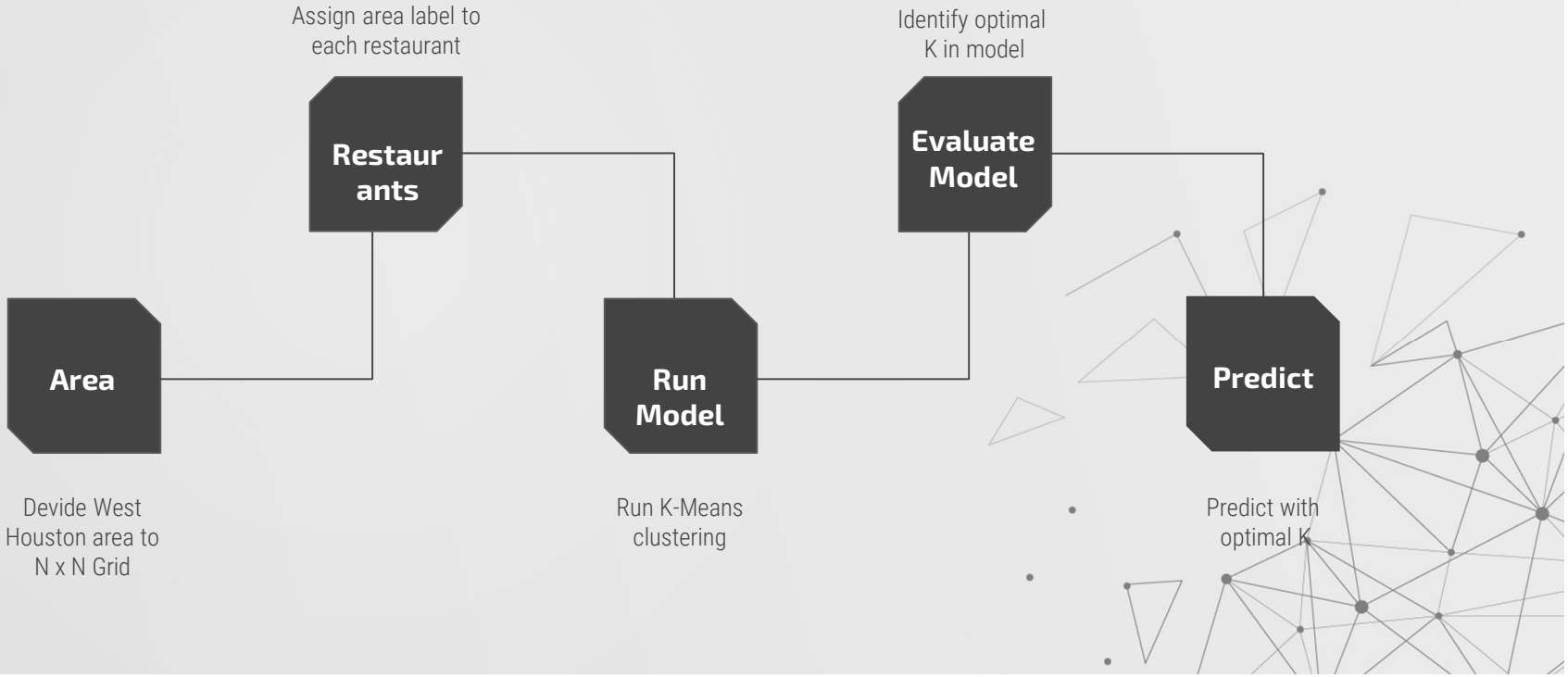


Restaurant distribution in West Houston



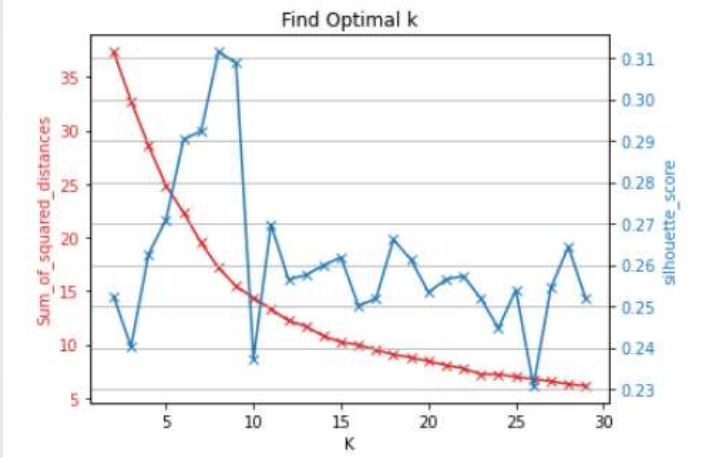
3. Methodology

Divide West Houston area to N x N grid and run K-Means to cluster restaurants based on category. Repeat this if no optimal K parameter found.

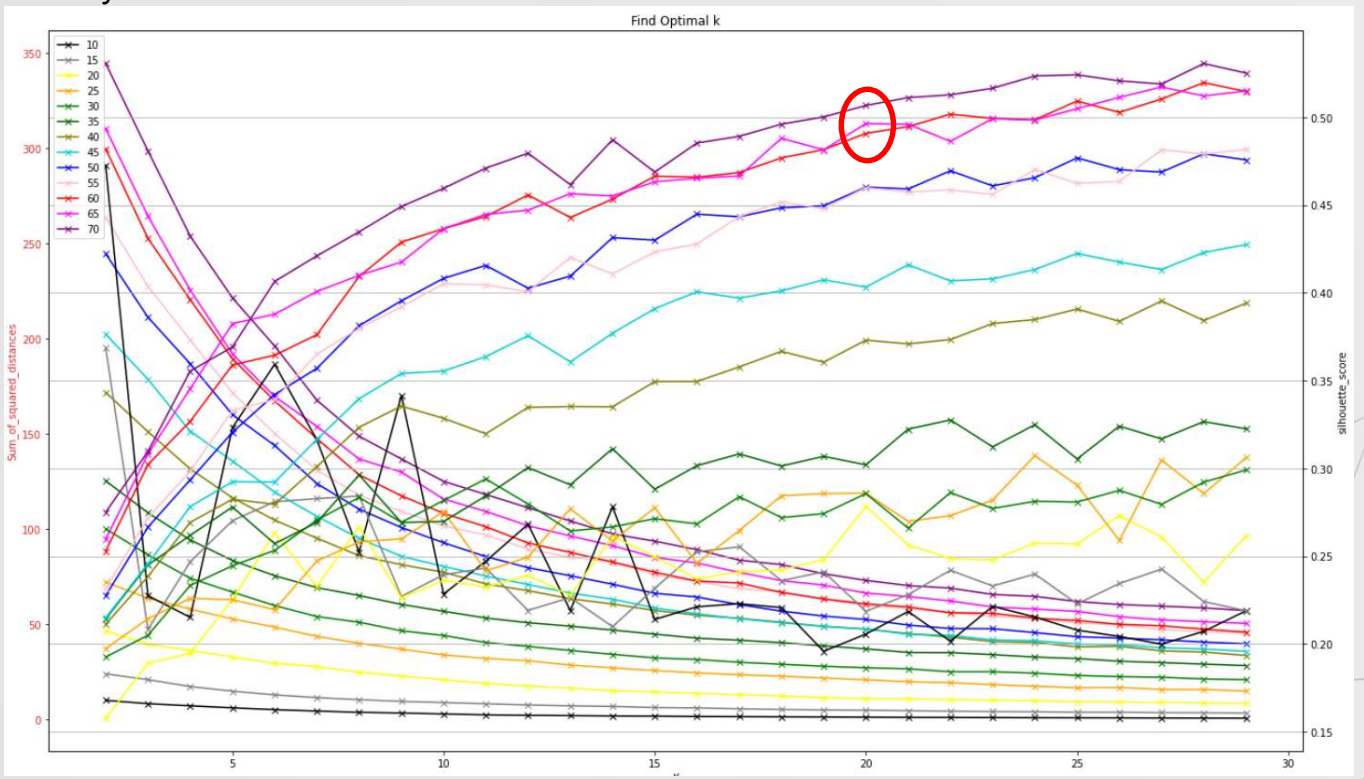


Quick test with 18 x 18 grid.

Result: no clear trend to identify K using elbow method
silhouette score (0.31) is too small to be acceptable.

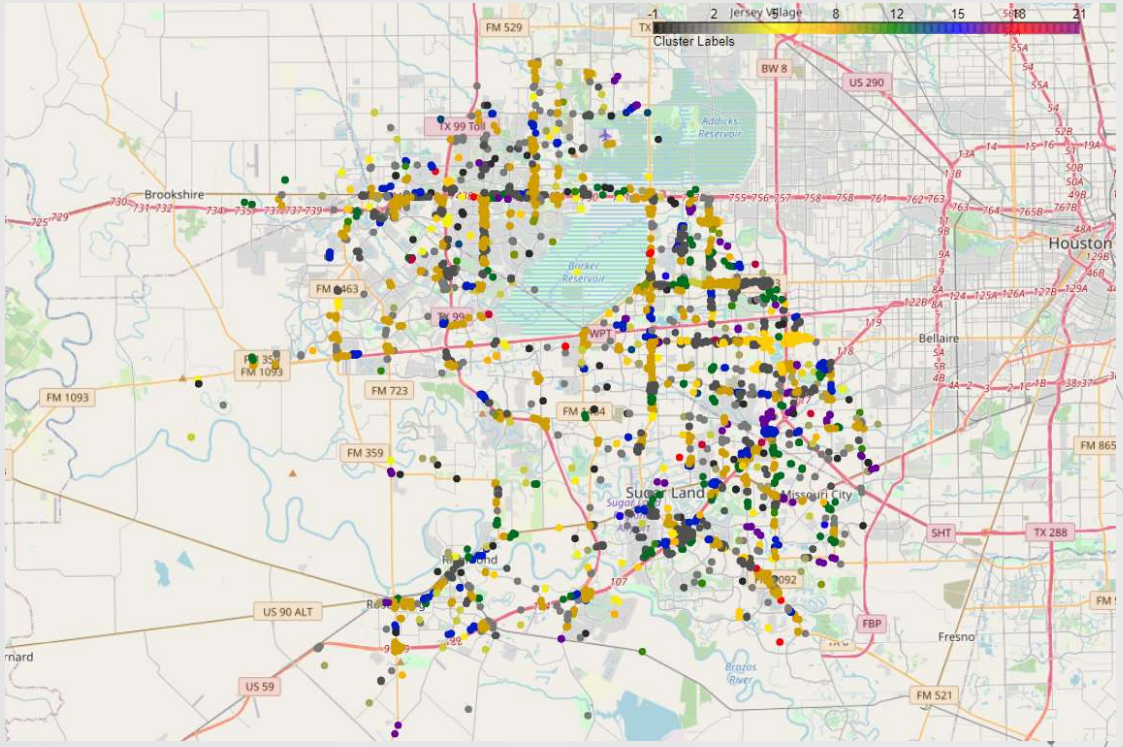


Uncertainty test: test N from 10 to 70



Run uncertainty test and select high silhouette score while keeping K small. N=65/K=20 seems acceptable

4. Clustering restaurants in 65x65 grid and 20 clusters



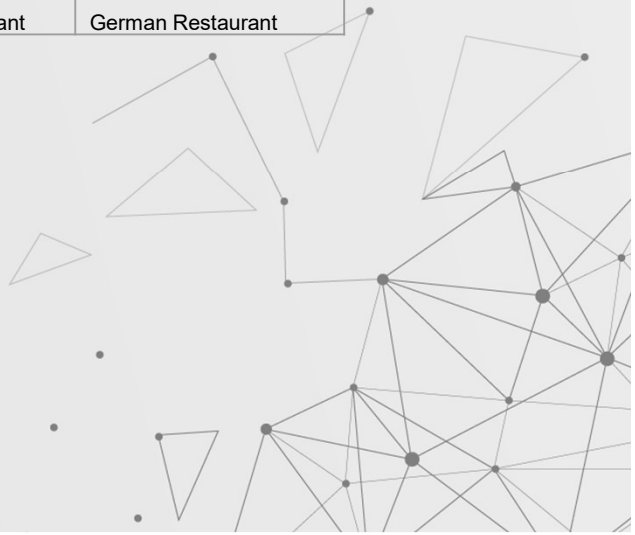
Top 5 Most Common Restaurants in each cluster

Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cafe	Vegetarian / Vegan Restaurant	Eastern European Restaurant	Gluten-free Restaurant	German Restaurant
1	**Asian Restaurant**	Fast Food Restaurant	Mexican Restaurant	Cafe	American Restaurant
2	Fast Food Restaurant	Mexican Restaurant	Cafe	Pakistani Restaurant	American Restaurant
3	Food	Vegetarian / Vegan Restaurant	Halal Restaurant	Gluten-free Restaurant	German Restaurant
4	Mexican Restaurant	Italian Restaurant	**Asian Restaurant**	Food	Fast Food Restaurant
5	American Restaurant	Fast Food Restaurant	Mexican Restaurant	Spanish Restaurant	Cafe
6	Indian Restaurant	Fast Food Restaurant	Middle Eastern Restaurant	**Asian Restaurant**	Cajun / Creole Restaurant
7	**Asian Restaurant**	Fast Food Restaurant	Cajun / Creole Restaurant	Food	Cafe
8	Cajun / Creole Restaurant	**Asian Restaurant**	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Eastern European Restaurant
9	Fast Food Restaurant	**Asian Restaurant**	Mexican Restaurant	Cafe	American Restaurant
10	Italian Restaurant	Restaurant	American Restaurant	Indian Restaurant	Fast Food Restaurant
11	Restaurant	Indian Restaurant	American Restaurant	Vegetarian / Vegan Restaurant	Eastern European Restaurant
12	Cafe	Fast Food Restaurant	Restaurant	Food	Caribbean Restaurant
13	Caribbean Restaurant	Vegetarian / Vegan Restaurant	Eastern European Restaurant	Gluten-free Restaurant	German Restaurant

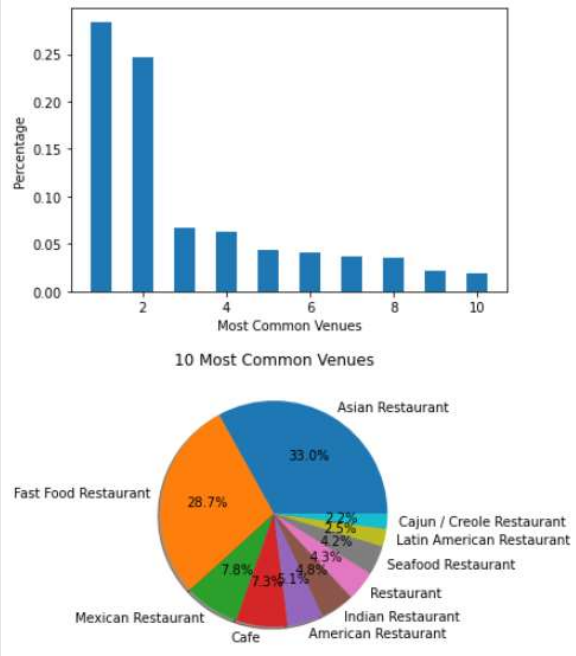
Top 5 Most Common Restaurants in clusters - continued

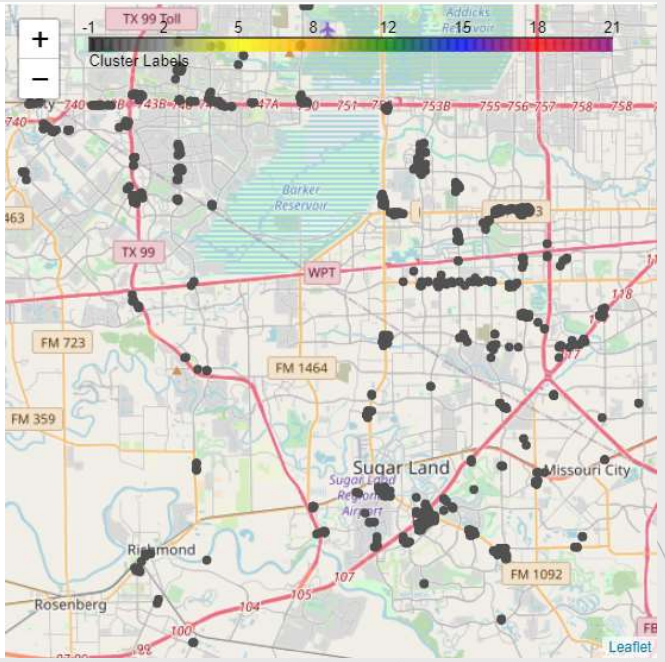
Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
14	Mexican Restaurant	Fast Food Restaurant	**Asian Restaurant**	Cafe	American Restaurant
15	Middle Eastern Restaurant	Vegetarian / Vegan Restaurant	Eastern European Restaurant	Gluten-free Restaurant	German Restaurant
16	Food	Fast Food Restaurant	Mexican Restaurant	Seafood Restaurant	American Restaurant
17	Latin American Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Eastern European Restaurant	Gluten-free Restaurant
18	Seafood Restaurant	Vegetarian / Vegan Restaurant	Dumpling Restaurant	Gluten-free Restaurant	German Restaurant
19	Mediterranean Restaurant	Afghan Restaurant	Eastern European Restaurant	Gluten-free Restaurant	German Restaurant

- Asian restaurants are popular in west Houston area.
- Cluster 1 and 7 with Asian restaurants as 1st most common venues
- Cluster 8 and 9 with Asian restaurants as 2nd most common venues

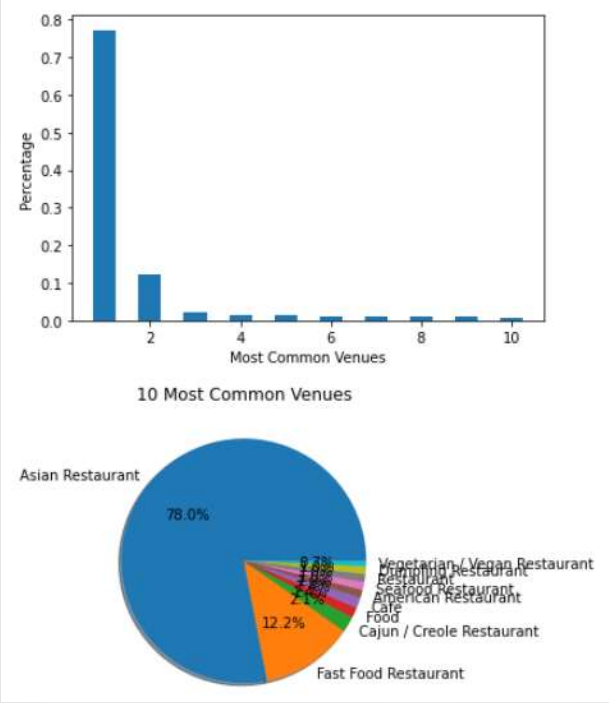


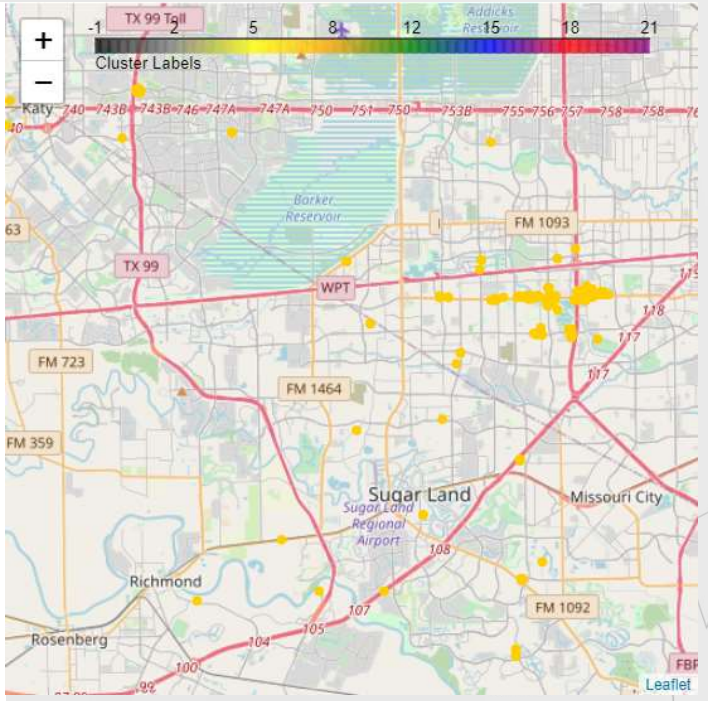
5. Discussion



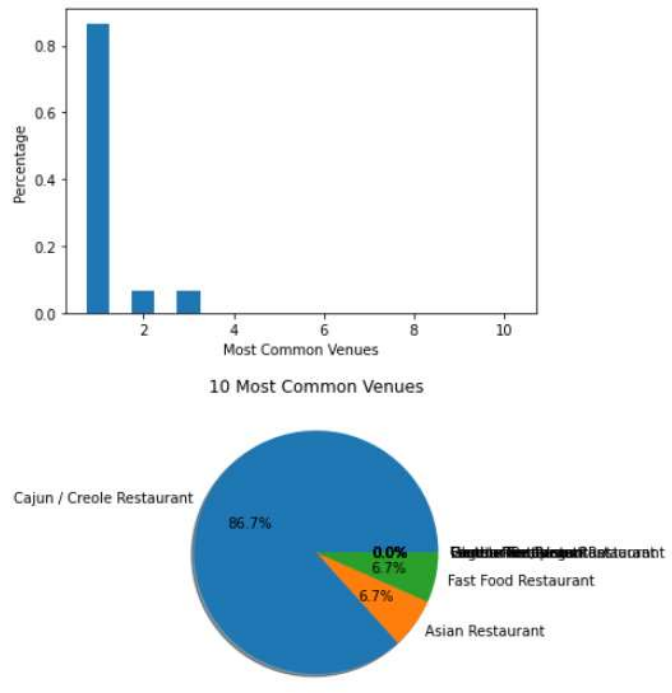


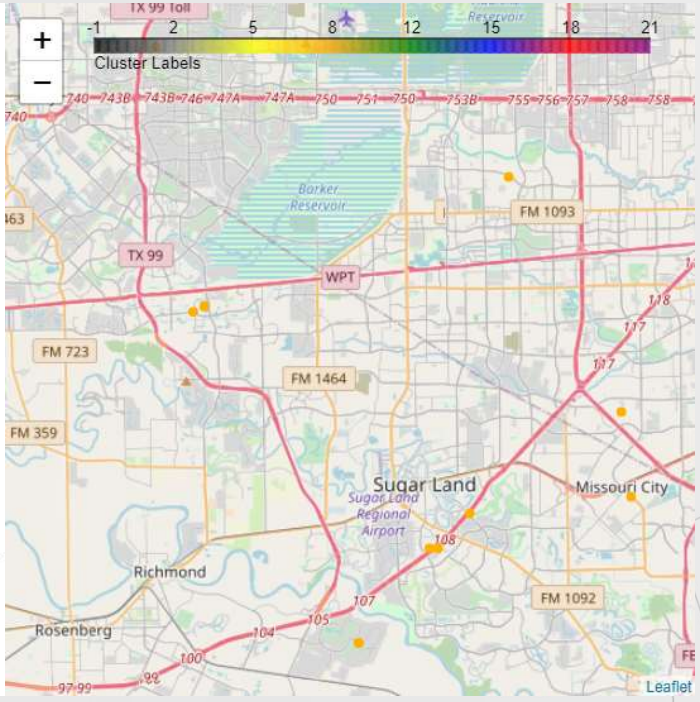
- Restaurants in cluster 1
- Popular categories with 33% Asian restaurants and 28.7% Fast food restaurant in west Houston.
- Most restaurants in Sugar Land/Missouri City area are located along highway 59 and 6.



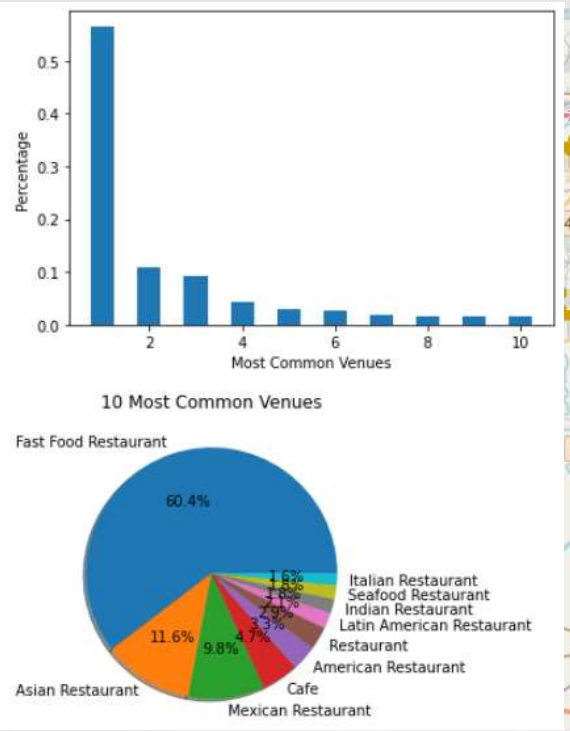


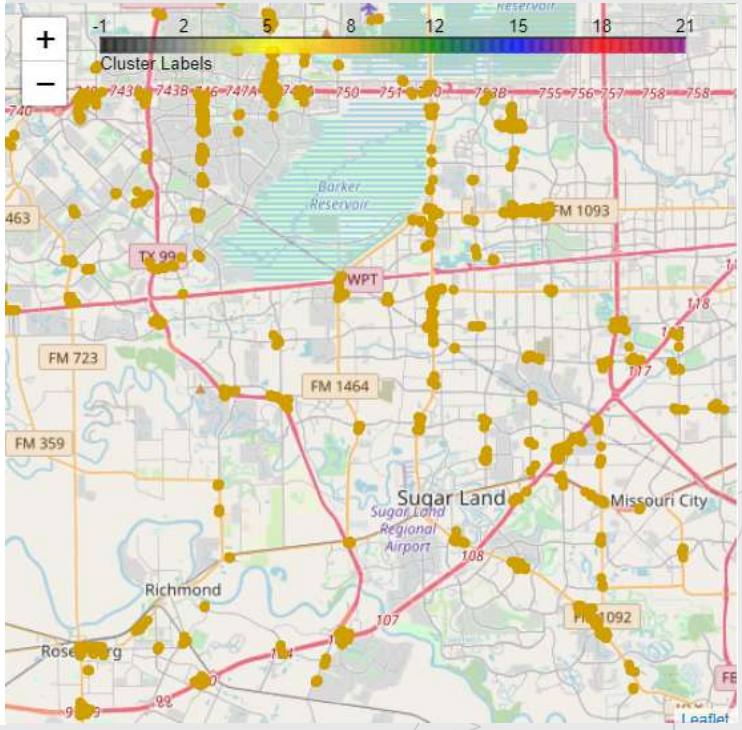
- Restaurants in cluster 7
- 7, 78% restaurants are Asian restaurants: Hot spots for Asian food. Both Houston China town and Katy Asian town fall in this cluster
- 5 Asian restaurants in the plaza near Hight 6 and Austin Parkway in Sugar Land/Missouri City area. This is probably the best location for future Asian restaurant with the best traffic.





- Restaurants in cluster 8
- Only 6.7% Asian restaurants, and not many restaurants are observed from map.





- Restaurants in cluster 9
- Popular in West Houston.
- 60.4% Fast food restaurants and 11.6% Asian restaurants
- All these restaurants are mainly located along highway 6 in Sugar Land/Missouri area.

6. Conclusion

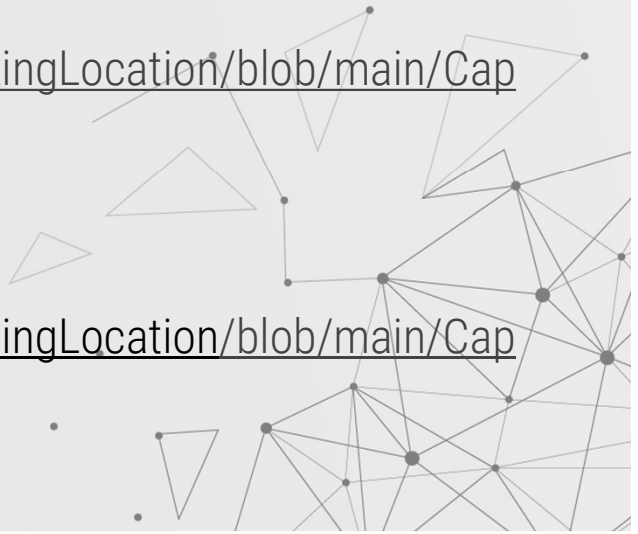
- Considering the percentage/number of restaurants in each cluster, cluster 1,7,and 9 seem to be the best clusters for a new Asian restaurant in Sugar Land/Missouri City area.
- Most of restaurants in cluster 1,7 and 9 are opened along highway 59 and 6.
- Cluster 7 has the highest percentage of Asian restaurants and covers Houston China town, Katy Asian Town and a plaza with 5 Asian restaurants near Highway 6 and Austin Pkwy in Sugar Land. Therefore that plaza or a nearby location in Sugar Land might be a perfect place for opening a new Asian restaurant
- Cluster 1 and 9 areas are also suitable for a new Asian restaurant business and more challenges will be expected with high percentage of fast-food restaurants.

References

1. Sugar Land, Wikipedia https://en.wikipedia.org/wiki/Sugar_Land,_Texas
2. Missouri City, Wikipedia https://en.wikipedia.org/wiki/Missouri_City,_Texas
3. Foursquare Venue Category Hierarchy https://developer.foursquare.com/docs/build-with-foursquare/categories/
4. Selecting the number of clusters with silhouette analysis on K-Means clustering https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html

Github Link

1. Project implementation using python:
   https://github.com/wdmhouston/CapestoneProject_PredictingLocation/blob/main/CapestoneProject_PredictingLocation.ipynb
2. Project report:
   https://github.com/wdmhouston/CapestoneProject_PredictingLocation/blob/main/CapestoneProject_PredictingLocationon_Report.pdf
3. Project presentation:
   https://github.com/wdmhouston/CapestoneProject_PredictingLocation/blob/main/CapestoneProject_PredictingLocationon_Presentation.pdf



THANKS

For any questions?

wangdeming@gmail.com