IBM DATA SCIENCE CERTICATION CAPSTONE PROJECT

IDENTIFY BEST NEIGHBERHOOD IN HOUSTON, TX TO OPEN INDIAN RESTAURANT

THE BATTLE OF NEIGHBORHOODS

OUR BIG IDEA

Houston is the fourth largest populous city in the United Status. Identify the optimal and best neighborhood to start an Indian restaurant by using data analysis techniques.

REASONS TO FIND BEST LOCATION TO OPEN INDIAN RESTAURANT IN HOUSTON,TX

✓ Energy Capital:

Oil price is always full of highs, lows. For the past few years, lower oil prices and increased competition have put significant pressure on Houston's more than 12,000 restaurants.

✓ Natural Disasters:

Houston restaurant community faces many challenges such as damages, including lost homes, cars and businesses due to recent hurricanes such as Hurricane Harvey. These unfortunate circumstances highlight the precarious nature of the restaurant business, where margins are slim, and most operators do not have significant sums saved for emergencies.

✓ Optimal Restaurant Neighborhood:

Due to above scenarios opening a new Indian restaurant requires serious consideration and is a lot more complicated than it seems. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question to find an optimal neighborhood in the city of Houston.

Data Acquisition

- ➢List of neighborhoods in Houston, TX
- ➢Latitude and longitude coordinates of the neighborhoods
- ➢Venue data, particularly data related to restaurants

Sources of data

- ➢Wikipedia page for neighborhoods

  https://en.wikipdia.org/wiki/Category:Greater_Houston

- ➢Geocoder package for latitude and longitude coordinates
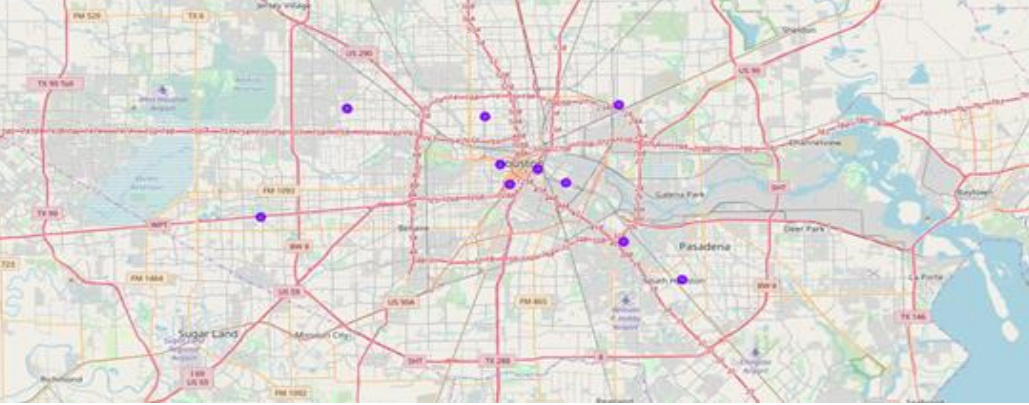- ➢Foursquare API for venue data

Suspendisse sit amet ipsum finibus justo

Methodology

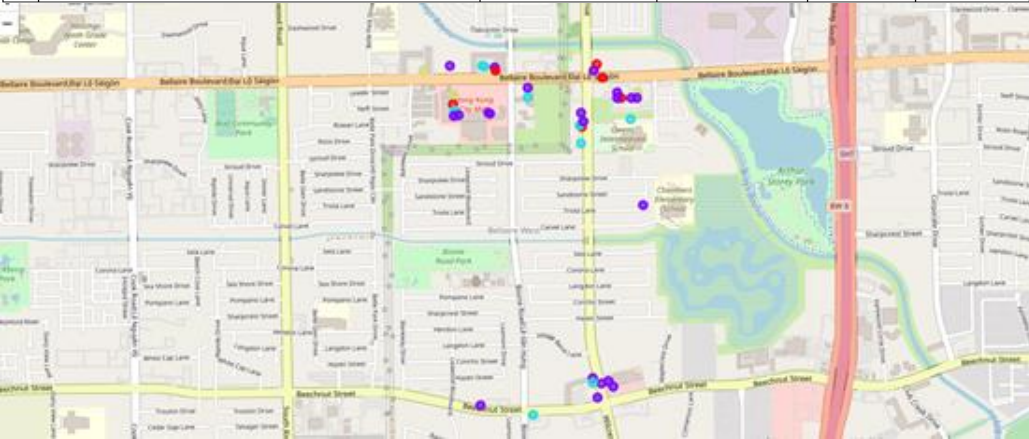✓ Web scraping Wikipedia page for neighborhoods list

✓ Get latitude and longitude coordinates using Geocoder

✓ Use Foursquare API to get venue data

✓ Group data by neighborhood and taking the mean of the frequency of occurrence of each venue category

✓ Filter venue category by Restaurants

✓ Perform clustering on the data by using k-means clustering

✓ Visualize the clusters in a map using Folium

Results

	Neighborhood	Asian Restaurant	Cluster Labels	Latitude	Longitude
0	► Baytown, Texas (1 C, 16 P)	0.000000	0	29.731875	-94.967078
1	► Bellaire, Texas (1 C, 7 P)	0.070000	0	29.696550	-95.575290
2	► Brazoria County, Texas (6 C, 16 P)	0.000000	0	29.179619	-95.410481
3	► Chambers County, Texas (7 C, 4 P)	0.000000	0	29.775887	-94.678311
4	► Clear Creek Independent School District (1...	0.000000	0	29.812750	-95.561750
5	► Conroe, Texas (1 C, 31 P)	0.010000	0	29.678734	-95.405491
6	► Fort Bend County, Texas (6 C, 11 P)	0.000000	0	29.538752	-95.448635
7	► Galveston Bay Area (8 C, 96 P)	0.000000	0	29.656876	-95.244964
8	► Galveston County, Texas (8 C, 15 P)	0.000000	0	29.507951	-95.093515
9	► Harris County, Texas (12 C, 53 P)	0.000000	0	29.658619	-95.228339
10	► Houston (25 C, 15 P, 3 F)	0.000000	0	29.760580	-95.369680
11	► Katy, Texas (2 C, 9 P)	0.000000	0	29.588706	-95.297429
12	► La Porte, Texas (8 P)	0.028571	0	29.713948	-95.278350
13	► Montgomery County, Texas (6 C, 7 P)	0.000000	0	30.300205	-95.503049
14	► Pasadena, Texas (3 C, 11 P)	0.000000	0	29.771910	-95.455933
15	► Pearland, Texas (3 C, 15 P)	0.000000	0	29.564527	-95.285006
16	► People from the Houston metropolitan area ...	0.000000	0	29.760580	-95.369680
17	► Rosenberg, Texas (1 C, 9 P)	0.000000	0	29.562258	-95.810106
18	► Santa Fe, Texas (8 P)	0.000000	0	29.661902	-95.301964
19	► Sports in the Houston metropolitan area (2...	0.000000	0	39.049830	-84.895540
20	► Sugar Land, Texas (5 C, 15 P)	0.020000	0	29.596949	-95.620896

✓Bellaire, TX neighborhood has greater number of Asian restaurants

✓Market place is great as one fourth of restaurants in this neighborhood are Asian

✓The completion is very intense

name	id	categories	lat	lng	total likes	total likes_cat	categories_new
Crawfish & Noodles	4bae9302f964a52051c23be3	Vietnamese Restaurant	29.70423	-95.5802	59	avg avg	Asian Food
Tony Thai Restaurant	4b2dab2ef964a52034da24e3	Thai Restaurant	29.70292	-95.56938	34	below avg	Asian Food
Com Tam Kieu Giang	4c66c9f9e75ac928aa77f8da	Vietnamese Restaurant	29.7023	-95.57845	4	poor	Asian Food
Pho Danh 2	52a4bf1511d278ee52857b20	Vietnamese Restaurant	29.70234	-95.57849	8	poor	Asian Food
Kim Son	4b8077a1f964a520077530e3	Vietnamese Restaurant	29.70317	-95.5685	65	avg avg	Asian Food
Pho Danh	4b916eadf964a5200ebc33e3	Vietnamese Restaurant	29.7023	-95.57849	11	poor	Asian Food
Nam Giao	4b744122f964a52035d02de3	Vietnamese Restaurant	29.70179	-95.5715	7	poor	Asian Food
Nguyễn-ngộ	4bc0e1f4461576b010f97a32	Vietnamese Restaurant	29.70441	-95.57685	20	poor	Asian Food
Thien Thanh	4b6daa97f964a52093842ce3	Vietnamese Restaurant	29.70446	-95.57697	17	poor	Asian Food
Pho Duy	51a3aa7e498e80da50870908	Vietnamese Restaurant	29.70076	-95.57158	9	poor	Asian Food
Lee's Sandwiches	4b6db576f964a52072882ce3	Vietnamese Restaurant	29.70424	-95.57623	40	below avg	Asian Food
Si Mê Restaurant	4d9ac3e73f785481881fd1d1	Vietnamese Restaurant	29.70163	-95.57166	0	poor	Asian Food
Bo Ne Houston	4bb6480af562ef3b0ce92f97	Vietnamese Restaurant	29.68782	-95.57418	9	poor	Asian Food
Xuan Huong Restaurant	590c28080f013c66ce1aa1a6	Vietnamese Restaurant	29.70296	-95.57448	3	poor	Asian Food
Mai's Baguettes	5bce720add8442002c7e4bf4	Vietnamese Restaurant	29.70191	-95.56886	0	poor	Asian Food
Pho Nguyen	4c3b5a7616cb2d7fd7ba02a9	Vietnamese Restaurant	29.68935	-95.57092	2	poor	Asian Food

Results

✓Popularity of Asian restaurants in this neighborhood is below average based on customer ratings.

✓Shortage of good quality restaurants

Recommendations

✓ Optimal Restaurant Location

Bellaire, TX neighborhood is the most preferred location to open a new Indian restaurant.

✓ Quality & Popularity

Quality and great customer satisfaction are critical

to the success of the new Indian restaurant

✓ Required More Analysis

This analysis is performed on limited data. Good amount of data and analysis by adding other factors such as transportation, demographics of inhabitants is required to get better results.

THANK YOU!