

Houston Food Delivery App

Shanshan Bai

July 21, 2020

1. Introduction

1.1 Background

In challenging time of covid-19 crisis, it is safe to avoid Dine-in in restaurants, and also important to support our Houston, TX local businesses. Now, more and more people choose to find the food and order from restaurants easily with apps and deliver to home. Although there are couple of apps support this service like Uber Eats, Doordash, and GrubHub, it is far short of such platforms to satisfy the market's needs. To develop restaurants delivery platforms and start the business is not easy. Let's start step by step. For this project, we will only cover the first step, which is to cluster restaurants and to find potential customers by income in different zip codes area.

1.2 Problem

During pandemic, the demand of online meal order and food delivery is dramatically increase in United States especially in population dense cities like Houston. Such service is developed very well in other countries like China, however, US. is just in beginning stage. Current providers as Uber Eats, Doordash, and GrubHub fill gap of this area, but service and price are still needed to be improved by involving in more competitors.

1.3 Interest

With more and more food delivery platforms setting up, American will feel their life be easy and enjoy the convenience and excellent service just by clicking their phone other than drive all the way to restaurants during their busy time. Restaurants owners would also expect profit of such platforms because it will save their money to advertise. The platform includes all information of the restaurants with rating, tips, categories, promotion, etc. If they had delivery service before, it will save their labor to deliver food as well.

2. Data

2.1 Data sources

Data will be collected and described to set up the restaurant's delivery platform. First, we will choose our potential customers. In the initial stage of the new business, we don't want to cover all residents with short of hands. The goal is to provide excellent service with reasonable price. One option is to collect the income data by different zip codes in Houston. And we can focus on residents with a certain range of income. Median Household income in Houston by zip code can be find from

<http://www.houstonstateofhealth.com/indicators/index/indicatorcsv?indicatorId=315&localeType=3&localeFilterId=0>. And the .CSV can be downloaded. After getting the data, we will clean, group by zip code and calculate the median income in each zip code. By sorting certain of range income, we can target our potential customers.

To create map from data, we also need to get the latitude and the longitude coordinates of each zip code. Data can be downloaded from <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/?refine.state=TX&q=houston>.

And it can also be saved .csv file.

To get details georgical information, population and population density of each zip code area are collected from <http://zipatlas.com/us/tx/houston/zip-code-comparison/population-density.htm>

After target potential customers in those zip codes, we will have better idea of their nearby restaurants that they would be familiar with. To collect restaurant data, we will choose from data providers. There are some available location data providers such as Foursquare, Google Places, and Yelp to provide local restaurants features. We can find the rate, cost, tips, etc. in their API. Here, I will use Foursquare API that can provide comprehensive and straight-forward data. This API contains massive dataset of accurate location data and being used widely by developers. To achieve our goal to set up the platform, we will use Foursquare Houston restaurants data.

<https://foursquare.com>. We can find restaurants and get venue category, rating, and other features.

3. Methodology

3.1 Queue Data

Database are from multiple sources. First, Great Houston Area Household Income with zip code data is collected from .csv file.

Indicator Name	What Is This Indicator	Location Type	Location	Indicator Rate Value	Indicator Rate Value Units	Rate Lower Confidence Interval	Rate Upper Confidence Interval	Indicator Count Value	Indicator Count Units	Breakout Rate Value	Breakout Rate Value Units	Breakout Rate Lower Confidence Interval	Breakout Rate Upper Confidence Interval
0 Median Household Income	This indicator shows the median household income.	Zip Code	77002	67043	dollars	59927	74159	NaN	NaN	...	NaN	NaN	NaN
1 Median Household Income	This indicator shows the median household income.	Zip Code	77002	72306	dollars	65002	79610	NaN	NaN	...	NaN	NaN	NaN
2 Median Household Income	This indicator shows the median household income.	Zip Code	77002	68229	dollars	58198	78280	NaN	NaN	...	NaN	NaN	NaN
3 Median Household Income	This indicator shows the median household income.	Zip Code	77002	62089	dollars	61992	72186	NaN	NaN	...	NaN	NaN	NaN
4 Median Household Income	This indicator shows the median household income.	Zip Code	77002	71319	dollars	57900	84738	NaN	NaN	...	NaN	NaN	NaN

5 rows × 29 columns

To apply the above data on a map, the latitude and the longitude coordinated of each zip code need to be collected.

Zip	City	State	Latitude	Longitude	Timezone	Daylight savings time flag	geopoint
0 77046	Houston	TX	29.733181	-95.431310	-6		1 29.733181,-95.43131
1 77015	Houston	TX	29.778526	-95.181180	-6		1 29.778526,-95.18118
2 77289	Houston	TX	29.833990	-95.434241	-6		1 29.83399,-95.434241
3 77072	Houston	TX	29.700898	-95.590020	-6		1 29.700898,-95.59002
4 77216	Houston	TX	29.833990	-95.434241	-6		1 29.83399,-95.434241

The population density is another feature we should consider, and the data is collected as below,

Zip	Location	City	Population	People / Sq. Mile	National Rank	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	Unnamed: 10	Unnamed: 11	Unnamed: 12	Unnamed: 13
0 77046	29.733084,-95.430659	Houston, Texas	471.0	32343.09	#108	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1 77081	29.712099,-95.480935	Houston, Texas	49691.0	15758.06	#308	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2 77036	29.699048,-95.536507	Houston, Texas	76146.0	9851.11	#609	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3 77006	29.741003,-95.391271	Houston, Texas	18875.0	8443.44	#754	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4 77057	29.744068,-95.489210	Houston, Texas	35491.0	8037.23	#825	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

We can collect venues by search restaurants from Foursquare API.

id	name	categories	referralId	hasPerk	location.address	location.crossStreet	location.lat	location.lng	l
0 4cc379573d7fa1cd4f08a65f	Spindletop Restaurant at Hyatt Regency Houston	[{"id": "4bf58dd8d48988d1ce941735", "name": "S..."}]	v-1595241921	False	1200 Louisiana St		Polk	29.756949	-95.369097
1 4b6a250ef964a52018ca2be3	Andalucia Tapas Restaurant and Bar	[{"id": "4bf58dd8d48988d150941735", "name": "S..."}]	v-1595241921	False	1204 Caroline Street		NaN	29.753929	-95.364056
2 4ae7fd79f964a520d1d821e3	Kim Son Restaurant Downtown	[{"id": "4bf58dd8d48988d14a941735", "name": "V..."}]	v-1595241921	False	2001 Jefferson St		Chartres	29.745905	-95.360474
3 4ac8af68f964a52045bc20e3	Harry's Restaurant	[{"id": "4bf58dd8d48988d1c0941735", "name": "M..."}]	v-1595241921	False	318 Tuam St		Bagby St.	29.746533	-95.381130
4 4d3f2b86557d6dcbb9bfd544	Mail's Restaurant	[{"id": "4bf58dd8d48988d14a941735", "name": "V..."}]	v-1595241921	False	3403 Milam St		Francis St	29.741242	-95.379769

3.2 Data cleaning

Data downloaded from websites are saved as .csv format. The data is read by Pandas .read_csv function. Columns are not usefully for this project are dropped. Groupby function is applied to group the same zip code information into one row, meantime the different income from the same zip code is calculated by median. In this project, multiple data are merged by using pandas.merge function for inner joint.

As starting business, we will focus on the top 10 income residents as our potential customers. After drop tedious info from the income date frame, we only select top 10n income data containing income, zip code in this dataset.

income_med	
Zip	
77010	209063.0
77005	165002.5
77401	158378.0
77094	158279.0
77059	131796.5
77345	123835.5
77024	114279.5
77007	106812.0
77389	103821.5
77379	100253.5

The latitude and the longitude coordinates of each zip code can be found in open source as well. Here, we need to get the zip code related geographic info.

	Zip	Latitude	Longitude
0	77046	29.733181	-95.431310
1	77015	29.778526	-95.181180
2	77289	29.833990	-95.434241
3	77072	29.700898	-95.590020
4	77216	29.833990	-95.434241

The density of area is cleaned and merged with income and geographic data.

	Zip	income_med	Latitude	Longitude	Population	People / Sq. Mile
0	77002	66909.5	29.755578	-95.36531	13289.0	6245.28
1	77003	39843.0	29.749278	-95.34741	9195.0	3533.01
2	77004	46791.0	29.728779	-95.36570	30379.0	4980.71
3	77005	165002.5	29.717529	-95.42821	23338.0	6009.48
4	77006	70333.0	29.741878	-95.38944	18875.0	8443.44

We can clean and sort restaurants rating getting from Foursquare API to get all useful information.

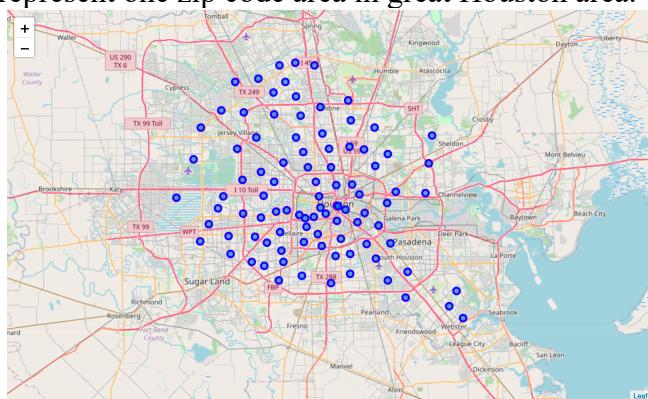
	name	categories	lat	lng	Zip	id
0	Spindletop Restaurant at Hyatt Regency Houston	Seafood Restaurant	29.756949	-95.369097	77002	4cc379573d7fa1cd4f08a65f
1	Andalucia Tapas Restaurant and Bar	Spanish Restaurant	29.753929	-95.364056	77002	4b6a250ef964a52018ca2be3
2	Kim Son Restaurant - Downtown	Vietnamese Restaurant	29.745905	-95.360474	77003	4ae7d79f964a520d1d821e3
3	Harry's Restaurant	Mediterranean Restaurant	29.746533	-95.381130	77006	4ac8af68f964a52045bc20e3
4	Mai's Restaurant	Vietnamese Restaurant	29.741242	-95.379769	77002	4d3f2b86557d6dcb9bfd5543

3.3 Data visualization

Folium is imported to visualize map containing geographic data. In this project, we select the highest household income zip code as the center on map to search restaurants around. Population density and median income in top10 income zip code is plotted by bubble plot. Top rating restaurants queued from Foursquare API are merged with top10 income zip code residents data by folium to the last map.

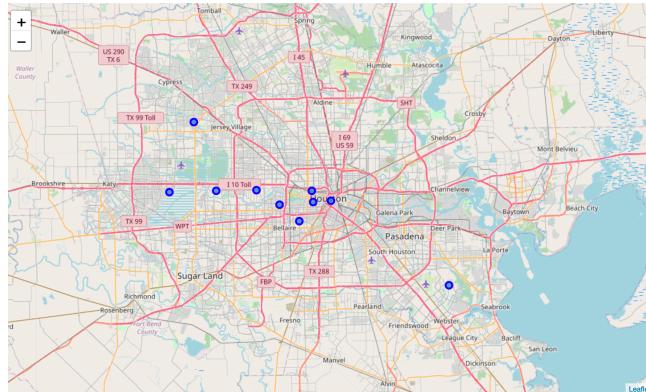
3.3.1 Visualization of all zip code in great Houston area with folium

One blue spot in this map represent one zip code area in great Houston area.



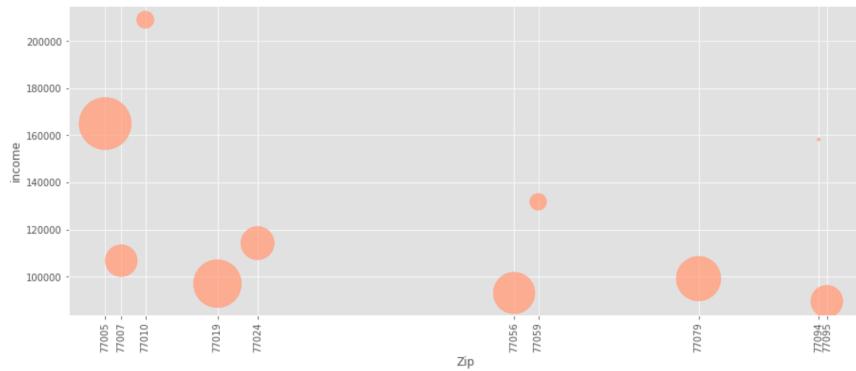
3.3.2 Map income top10 zip code with folium

In the beginning of a business, we will only focus some resident but not all. It is better to select the top income group. The map below shows the top10 highest income area.



3.3.3 Bubble plots of income, population density in top10 income zip code

We also need to visualize the relationship of income, population density and zip code distribution. Bubble plot will give a clear answer



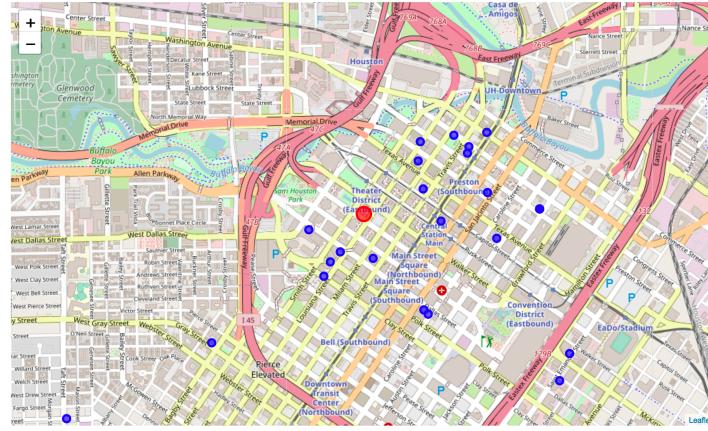
4. Results

4.1.1 To make a platform, we will first select top rated restaurants from the API.

	name	categories	latitude	longitude	Zip	id	rating
0	Korea Garden Restaurant	Korean Restaurant	29.797101	-95.526260	77055	4bb389b2715eef3bc27986bb	8.9
1	Cuchara Restaurant	Mexican Restaurant	29.748292	-95.385546	77006	4fcfc59108174aac005f032	8.8
2	Huynh Restaurant	Vietnamese Restaurant	29.751692	-95.355321	77003	4b42a9dcf964a5206fd825e3	8.7
3	The Fish Restaurant & Sushi Bar	Sushi Restaurant	29.752249	-95.376820	77002	4ad75ab2f964a520c40921e3	8.6
4	Harry's Restaurant	Mediterranean Restaurant	29.746533	-95.381130	77006	4ac8af68f964a52045bc20e3	8.6

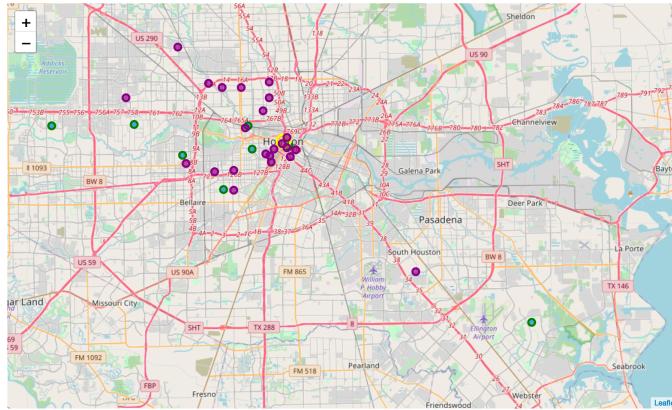
4.1.2 Map restaurants with folium

In Foursquare API, restaurants can be searched and queued on folium map by its regular venue search and we will map the top rating restaurants on folium map.



4.1.3 Merge Household Data in Top10 Income Zip Code Area with Top Rating Restaurants

Finally, household data and restaurant data are merged into one map. In this map, purple dots represent top rating restaurants and green dots are top 10 income residents neighborhood.



5. Discussion

Houston is a multi-culture city. It is not easy to find all top rating restaurants from Foursquare API. To set up food delivery app, more complex data from different sources will be needed. With enlarging dataset, all data can be clustered and sorted to make multiple categories on the platform.

A competitive feature will be added into the platform will be search restaurants by map. People sometimes like to search by map to get such category food like Mexican food, Japanese cuisine, Chinese food, Italian restaurant, etc. In Houston and other big cities, such categories are also related to their location. So map will help a lot to find what people like to order.

Daily premium call from Foursquare API is limited to 50. So, there are lots of missing data by such limitation. Besides, lots of data are out of date. I found there are permanent close business are still listed there. So, we will need more and better sources to collect data.

6. Conclusion

Restaurants list of neighborhoods of top income would be get by selecting top rating and popularity. Multiple sources will be needed for further analysis. We can select the top10 income zip code residents as our potential customers when start the business.

7. References:

<http://www.houstonstateofhealth.com/indicators/index/indicatorcsv?indicatorId=315&localeTypeId=3&localeFilterId=0>.

<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/?refine.state=TX&q=houston>

<http://zipatlas.com/us/tx/houston/zip-code-comparison/population-density.htm>

<https://foursquare.com>.