

# **Data Science Capstone**

Final Report: Battle of the Neighborhoods

Heston Miles

06.05.2020

<b>INTRODUCTION</b>	<b>3</b>
Business Problem	3
Audience	3
<b>DATA</b>	<b>4</b>
Overview	4
Data Sources	4
<b>METHODOLOGY</b>	<b>5</b>
High Level Goals	5
<b>RESULTS</b>	<b>7</b>
<b>DISCUSSION</b>	<b>9</b>
<b>CONCLUSION</b>	<b>10</b>

## Introduction

### Business Problem

An associate has stated that they would like to open a new, high-end sushi restaurant in Houston but she is new to the area. They have requested advice on the best location(s) to establish their new business venture to give them the best opportunity for success. Using city data, data science and machine learning, we seek to derive an answer to this question and make a recommendation.

### Audience

Our target user population is our associate, an entrepreneur looking to expand their business ventures into the Houston area.

# Data

## Overview

The data that will be used for this analysis/decision-making is based on the following premise: We want to find an area/ZIP code with a high Median Household Income, a sizable population and other area venues (restaurants, entertainment) but little to no direct (high-end sushi restaurant) competition.

The dataset that will need to be created will incorporate Houston (Harris County only) ZIP codes, as well as their corresponding Neighborhood Names, Populations, and Median Household Incomes. This data will be joined with Foursquare API data, which will give us a good idea of venues/competition in the area.

The plan is to use only the top 30 (of 133) ZIP codes in Houston, based on Median Household Income, then reduce that to 20, based on highest Population within those high Median Household Income ZIP Codes. From there, we will apply K-Means Clustering and analyze each Cluster based on Median Household Income, Population and other venues/competition to make the final recommendation.

## Data Sources

The data sources that were used are the following:

- Venue data from Foursquare for selected ZIP codes
- ZIP code data with associated Median Household Income, Location (Lat/Long), Population and County (This data needed significant cleaning to scrape, as it was several nested tables and generally very poorly coded)
  - o <http://zipatlas.com/us/tx/houston/zip-code-comparison/median-household-income.htm>
- ZIP code data with Neighborhood names and County (so that we can isolate only the Harris County ZIP codes and have named neighborhoods)
  - o <https://web.har.com/zipcode>

## Methodology

### High Level Goals

The high level goal for this study was to find the ZIP code in Houston that, based on various factors, will give our associate the highest chance of being successful with their new venture.

### Dataset Creation

I began by, using the BeautifulSoup package, scraping data from HTML tables. This required 2 data scrapes:

- Table1: ZIP Code, Population, Latitude, Longitude, and Median Household Income
- Table 2: ZIP Code, County Name, Neighborhood Name

These tables were then joined by using the common ZIP code. Once the dataset was built, I sorted the dataset (descending) by Median Household Income and then removed all rows past #30.

I subsequently took the reduced list and sorted it again (descending) by Population, and then only kept the top 20 rows.

ZIP Code	Longitude	Latitude	Population	Median HH Income	Neighborhood	County
80	77084	29.82686	-95.65086	63557	53075	Houston - Northwest - ADDICKS BARKER,HOUSTON HARRIS
79	77083	29.693991	-95.64783	52061	52931	Houston - Southwest - HOUSTON HARRIS
73	77077	29.749158	-95.620344	42416	52667	Houston - Southwest - HOUSTON HARRIS
38	77040	29.876755	-95.532143	39520	50183	Houston - Northwest - BAMBEL,HOUSTON,JERSEY VI... HARRIS
91	77095	29.916055	-95.663077	39275	76814	Houston - Northwest - HOUSTON HARRIS
85	77089	29.588469	-95.218012	36450	53123	Houston - Southeast - HOUSTON HARRIS
60	77064	29.924216	-95.537411	35966	60945	Houston - Northwest - HOUSTON HARRIS
53	77057	29.744068	-95.48921	35491	46563	Houston - Southwest - HOUSTON HARRIS
92	77096	29.674133	-95.478354	33987	46140	Houston - Southwest - HOUSTON HARRIS
22	77024	29.771991	-95.515453	32746	82620	Houston - Southwest - BUNKER HILL VILLAGE,HEDW... HARRIS
66	77070	29.97967	-95.572835	32385	60887	Houston - Northwest - HOUSTON HARRIS
75	77079	29.774523	-95.601071	30887	70587	Houston - Southwest - ADDICKS,HOUSTON HARRIS
62	77066	29.958498	-95.500223	27638	55760	Houston - Northwest - HOUSTON HARRIS
61	77065	29.926556	-95.603242	27070	58426	Houston - Northwest - HOUSTON,JERSEY VILLAGE,J... HARRIS
58	77062	29.575781	-95.134334	26978	75689	Houston - Southeast - HOUSTON HARRIS
39	77041	29.857989	-95.585771	25275	56746	Houston - Northwest - HOUSTON,JERSEY VILLAGE,J... HARRIS
23	77025	29.688811	-95.43478	23581	48232	Houston - Inner Loop - ASTRODOME,ASTROWORLD,HO... HARRIS
3	77005	29.718435	-95.423555	23338	104035	Houston - Inner Loop - HOUSTON,SOUTHSIDE PLACE... HARRIS
55	77059	29.615219	-95.13496	16690	104844	Houston - Southeast - HOUSTON HARRIS
17	77019	29.752702	-95.407379	15640	61899	Houston - Inner Loop - HOUSTON HARRIS

By manipulating the dataset in this manner, I was able to focus only on populous areas with high Median Household Incomes, which would make them prime areas for a new restaurant.

Next, I connected to the Foursquare API and pulled data regarding venues for the 20 ZIP codes we focused on. This gave me an idea of competition in each area.

I then joined the Foursquare data with the master dataset, so that it now contained, for each area:

- ZIP code
- Location Data
- Population
- Neighborhood Name
- Median Household Income
- County Name
- Top 10 Venues

ZIP Code	Longitude	Latitude	Population	Median HH Income	Neighborhood	County	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
80	77084	29.82686	-95.65086	63557	53075	Houston - Northwest - ADDICKS BARKER,HOUSTON	HARRIS	2	Brewery	Vietnamese Restaurant	Gift Shop	Farmers Market	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant	Fried Chicken Joint	Frozen Yogurt Shop
79	77083	29.693991	-95.64783	52061	52931	Houston - Southwest - HOUSTON	HARRIS	1	Burger Joint	Gym / Fitness Center	Sandwich Place	Liquor Store	Business Service	Dessert Shop	BBQ Joint	Video Store	Grocery Store	Non-Profit
73	77077	29.749158	-95.620344	42416	52667	Houston - Southwest - HOUSTON	HARRIS	1	Burger Joint	Gym / Fitness Center	Sandwich Place	Liquor Store	Business Service	Dessert Shop	BBQ Joint	Video Store	Grocery Store	Non-Profit
38	77040	29.876755	-95.532143	39520	50183	Houston - Northwest - BAMMEL,HOUSTON,JERSEY VIL...	HARRIS	0	Optical Shop	Vietnamese Restaurant	Furniture / Home Store	Farmers Market	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant	Fried Chicken Joint	Frozen Yogurt Shop
91	77095	29.916055	-95.663077	39275	76814	Houston - Northwest - HOUSTON	HARRIS	1	Fast Food Restaurant	Pharmacy	Video Store	Park	Paper / Office Supplies Store	Gymnastics Gym	Gym	Gun Range	Mobile Phone Shop	Mexican Restaurant
85	77089	29.588469	-95.218012	36450	53123	Houston - Southeast - HOUSTON	HARRIS	3	Shop & Service	Business Service	Trail	Playground	Cajun / Creole Restaurant	Farmers Market	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant
60	77064	29.924216	-95.537411	35966	60945	Houston - Northwest - HOUSTON	HARRIS	1	Fast Food Restaurant	Pharmacy	Video Store	Park	Paper / Office Supplies Store	Gymnastics Gym	Gym	Gun Range	Mobile Phone Shop	Mexican Restaurant
53	77057	29.744068	-95.48921	35491	46563	Houston - Southwest - HOUSTON	HARRIS	1	Burger Joint	Gym / Fitness Center	Sandwich Place	Liquor Store	Business Service	Dessert Shop	BBQ Joint	Video Store	Grocery Store	Non-Profit
92	77096	29.674133	-95.478354	33987	46140	Houston - Southwest - HOUSTON	HARRIS	1	Burger Joint	Gym / Fitness Center	Sandwich Place	Liquor Store	Business Service	Dessert Shop	BBQ Joint	Video Store	Grocery Store	Non-Profit
22	77024	29.771991	-95.515453	32746	82620	Houston - Southwest - BUNKER HILL VILLAGE,HEDW...	HARRIS	4	Basketball Court	Clothing Store	Vietnamese Restaurant	Gift Shop	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant	Fried Chicken Joint	Frozen Yogurt Shop

## Machine Learning

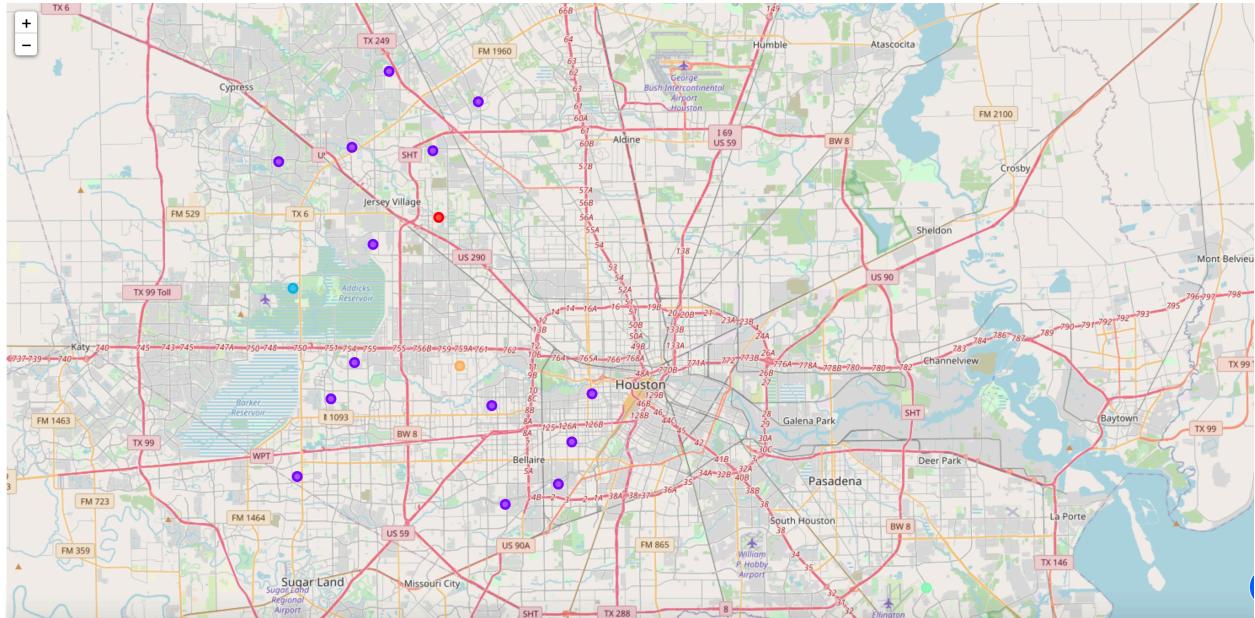
With the dataset now complete, I applied K-Means Clustering to the areas to see how the ZIP codes could be grouped, based on similarities in the data. 5 clusters were created.

## Analysis

Upon completion of the K-Means Clustering, the ZIP code points with their associated Cluster Colors were applied to the map, and analysis was conducted on each cluster to select an ideal location. This analysis included looking at the Population and Median Household Incomes for each Cluster, as well as the 10 most common venues for each area.

## Results

The clusters were generated and plotted to the map as seen below:



The results for each cluster is as follows:

### Cluster 1:

- ZIP Code(s): 77040
- Median Annual Household: \$50,183.00
- Average Population: 39,520

Longitude	Neighborhood	County	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
38 29.876755	Houston - Northwest - BAMBEL,HOUSTON,JERSEY VIL...	HARRIS	0	Optical Shop	Vietnamese Restaurant	Furniture / Home Store	Farmers Market	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant	Fried Chicken Joint	Frozen Yogurt Shop

### Cluster 2:

- ZIP Code(s): 77083, 77077, 77095, 77064, 77057, 77096, 77070, 77079, 77066, 77065, 77041, 77025, 77005, 77019
- Median Annual Household Income: \$60,902.29
- Average Population: 31,786

Longitude	Neighborhood	County	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
79	29.693991	Houston - Southwest - HOUSTON	HARRIS	1 Burger Joint	Gym / Fitness Center	Sandwich Place	Liquor Store	Business Service	Dessert Shop	BBQ Joint	Video Store	Grocery Store	Non-Profit
73	29.749158	Houston - Southwest - HOUSTON	HARRIS	1 Burger Joint	Gym / Fitness Center	Sandwich Place	Liquor Store	Business Service	Dessert Shop	BBQ Joint	Video Store	Grocery Store	Non-Profit
91	29.916055	Houston - Northwest - HOUSTON	HARRIS	1 Fast Food Restaurant	Pharmacy	Video Store	Park	Paper / Office Supplies Store	Gymnastics Gym	Gym	Gun Range	Mobile Phone Shop	Mexican Restaurant
60	29.924216	Houston - Northwest - HOUSTON	HARRIS	1 Fast Food Restaurant	Pharmacy	Video Store	Park	Paper / Office Supplies Store	Gymnastics Gym	Gym	Gun Range	Mobile Phone Shop	Mexican Restaurant
53	29.744068	Houston - Southwest - HOUSTON	HARRIS	1 Burger Joint	Gym / Fitness Center	Sandwich Place	Liquor Store	Business Service	Dessert Shop	BBQ Joint	Video Store	Grocery Store	Non-Profit
92	29.674133	Houston - Southwest - HOUSTON	HARRIS	1 Burger Joint	Gym / Fitness Center	Sandwich Place	Liquor Store	Business Service	Dessert Shop	BBQ Joint	Video Store	Grocery Store	Non-Profit
66	29.97967	Houston - Northwest - HOUSTON	HARRIS	1 Fast Food Restaurant	Pharmacy	Video Store	Park	Paper / Office Supplies Store	Gymnastics Gym	Gym	Gun Range	Mobile Phone Shop	Mexican Restaurant
75	29.774523	Houston - Southwest - ADDICKS,HOUSTON	HARRIS	1 Salon / Barbershop	Fast Food Restaurant	Park	Hookah Bar	Bakery	Supermarket	Bank	Fried Chicken Joint	Pizza Place	Juice Bar
62	29.958498	Houston - Northwest - HOUSTON	HARRIS	1 Fast Food Restaurant	Pharmacy	Video Store	Park	Paper / Office Supplies Store	Gymnastics Gym	Gym	Gun Range	Mobile Phone Shop	Mexican Restaurant
61	29.926556	Houston - Northwest - HOUSTON,JERSEY VILLAGE,J...	HARRIS	1 Fast Food Restaurant	Spa	Restaurant	Mobile Phone Shop	Bank	Sandwich Place	Burger Joint	Department Store	Liquor Store	Cupcake Shop
39	29.857989	Houston - Northwest - HOUSTON,JERSEY VILLAGE,J...	HARRIS	1 Fast Food Restaurant	Spa	Restaurant	Mobile Phone Shop	Bank	Sandwich Place	Burger Joint	Department Store	Liquor Store	Cupcake Shop
23	29.688811	Houston - Inner Loop - ASTRODOME,ASTROWORLD,HO...	HARRIS	1 Park	Recreation Center	Gym / Fitness Center	Gym	Brewery	Frozen Yogurt Shop	Eastern European Restaurant	Farmers Market	Fast Food Restaurant	Food & Drink Shop
3	29.718435	Houston - Inner Loop - HOUSTON,SOUTHSIDE PLACE...	HARRIS	1 Outdoors & Recreation	Optical Shop	Bakery	Beach	Eastern European Restaurant	Vietnamese Restaurant	Fruit & Vegetable Store	Fast Food Restaurant	Food & Drink Shop	Food Truck
17	29.752702	Houston - Inner Loop - HOUSTON	HARRIS	1 Men's Store	Coffee Shop	Cosmetics Shop	Gym	Kids Store	Café	Clothing Store	French Restaurant	Vietnamese Restaurant	Kitchen Supply Store

### Cluster 3:

- ZIP Code(s): 77084
- Median Annual Household Income: \$53,075.00
- Average Population: 63,557

Longitude	Neighborhood	County	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
80	29.82686	Houston - Northwest - ADDICKS,BARKER,HOUSTON	HARRIS	2 Brewery	Vietnamese Restaurant	Gift Shop	Farmers Market	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant	Fried Chicken Joint	Frozen Yogurt Shop

### Cluster 4:

- ZIP Code(s): 77089, 77062, 77059
- Median Annual Household: \$77,885.33
- Average Population: 26,706

Longitude	Neighborhood	County	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
85	29.588469	Houston - Southeast - HOUSTON	HARRIS	3 Shop & Service	Business Service	Trail	Playground	Cajun / Creole Restaurant	Farmers Market	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant
58	29.575781	Houston - Southeast - HOUSTON	HARRIS	3 Shop & Service	Business Service	Trail	Playground	Cajun / Creole Restaurant	Farmers Market	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant
55	29.615219	Houston - Southeast - HOUSTON	HARRIS	3 Shop & Service	Business Service	Trail	Playground	Cajun / Creole Restaurant	Farmers Market	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant

### Cluster 5:

- ZIP Code(s): 77024
- Median Annual Household Income: \$82,620.00
- Average Population: 32,746

Longitude	Neighborhood	County	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
22	29.771991	Houston - Southwest - BUNKER HILL VILLAGE,EDW...	HARRIS	4 Basketball Court	Clothing Store	Vietnamese Restaurant	Gift Shop	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant	Fried Chicken Joint	Frozen Yogurt Shop

## Discussion

Based on the data analysis conducted, my recommendation for an area to create a new upscale sushi restaurant in Houston would be:

Cluster 5 (77024 – Bunker Hill/Hedwig Village Area)

This recommendation is based on:

1. High Median Household Income (#1 out of 5 clusters)
2. Area Population (#3 out of 5 clusters)
3. Lack of Competition in the area (Sushi restaurants are not in the top 10 venues in the area)

One observation/thought that I had after working on this project is that other data that I would look into (but do not currently have access to) would be restaurant fail rates and property prices in these ZIP codes.

## Conclusion

In conclusion, based on several factors stated in the previous section, the best location to start a new high-end sushi restaurant in Houston would be:

### Cluster 5:

- ZIP Codes: 77024
- Median Annual Household Income: \$82,620.00
- Average Population: 32,746

Longitude	Neighborhood	County	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
22 29.771991	Houston - Southwest - BUNKER HILL VILLAGE,HEDW...	HARRIS	4	Basketball Court	Clothing Store	Vietnamese Restaurant	Gift Shop	Fast Food Restaurant	Food & Drink Shop	Food Truck	French Restaurant	Fried Chicken Joint	Frozen Yogurt Shop