Capstone Presentation

Michael Mehlman
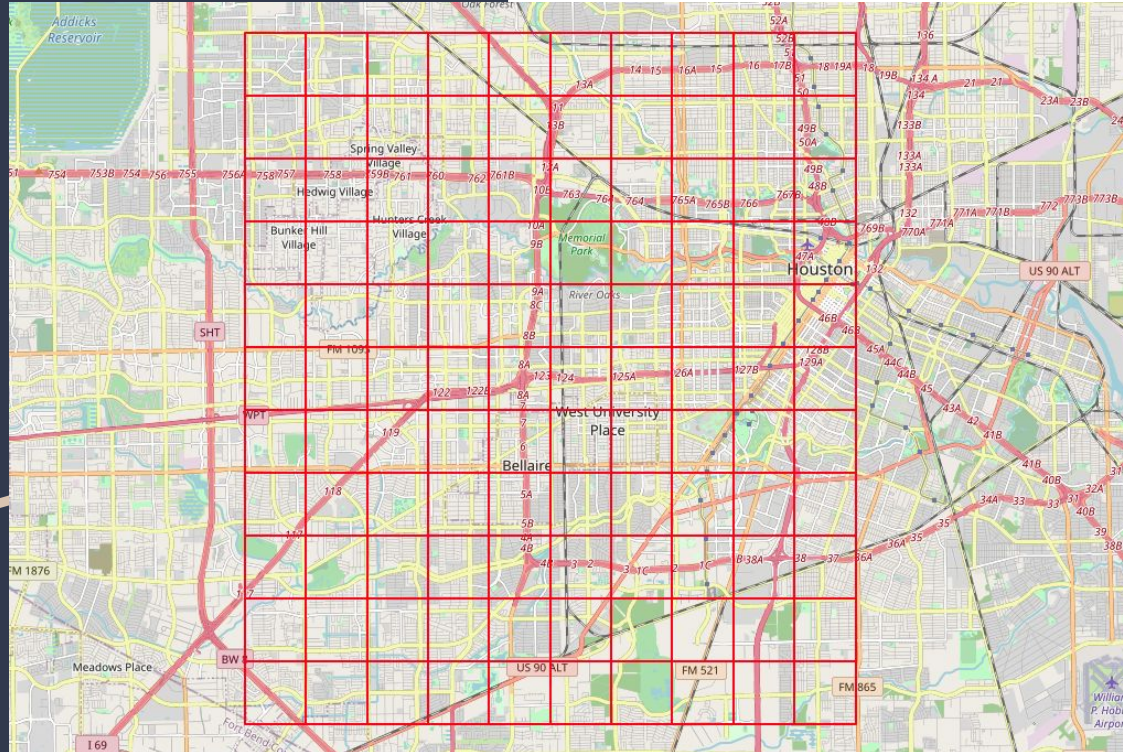
Introduction

Goal: Identify regions of Houston that have fewer restaurants than their peer group indicating a potential market opportunity.

Break Houston Down into Regions

10x10 grid of 1.8km square rectangles centered on my brother's house.

Get Data from Foursquare

I used Foursquare data for both the region clustering (finding similar regions) and also for comparing density of venues in those regions.

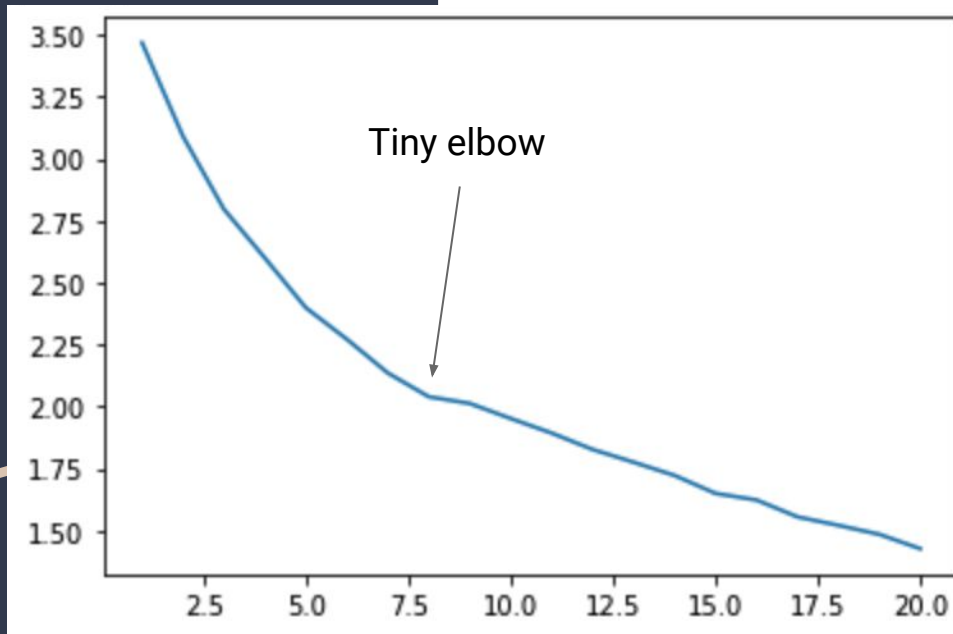| venues							
	reg	lat	lon	venue	venue_lat	venue_lon	venue_category
0	0	29.644244	-95.537192	FedEx Ship Center	29.639955	-95.535449	Shipping Store
1	0	29.644244	-95.537192	Walgreens	29.637378	-95.529751	Pharmacy
2	0	29.644244	-95.537192	Redbox	29.644398	-95.538065	Video Store
3	0	29.644244	-95.537192	MicroSearch	29.649152	-95.544295	Camera Store
4	0	29.644244	-95.537192	CVS pharmacy	29.638534	-95.523906	Pharmacy
...	...	...	...	...	...	...	...

Prepare for Kmeans

- Perform onehot encoding on all venues
- Group venues by region (mean)
- Return the 10 most common venues for kmeans

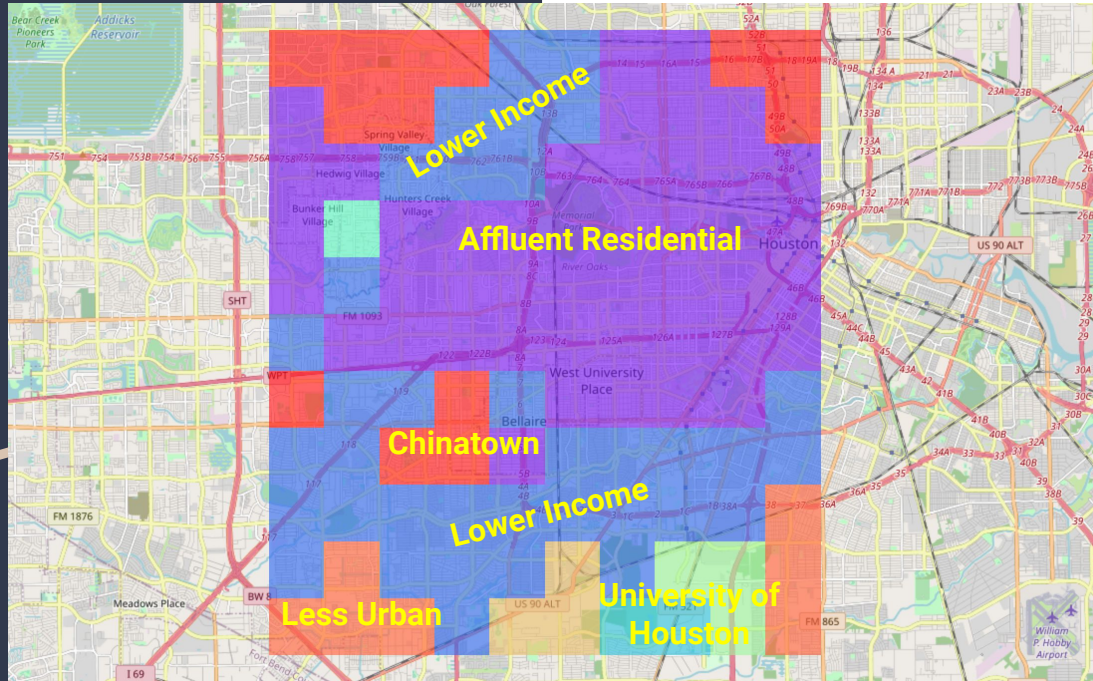| Auto Workshop	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Baseball Stadium
0.030303	0.000000	0.000000	0.0	0.030303	0.030303	0.000000	0.000000	0.0
0.000000	0.000000	0.000000	0.0	0.022727	0.022727	0.000000	0.000000	0.0
0.000000	0.000000	0.000000	0.0	0.000000	0.028571	0.028571	0.000000	0.0
0.000000	0.000000	0.000000	0.0	0.000000	0.095238	0.000000	0.095238	0.0
0.000000	0.000000	0.000000	0.0	0.000000	0.055556	0.055556	0.111111	0.0

Look for the elbow in inertia!

Unfortunately, I did not find a good elbow. There was a tiny elbow at 8, so I went with that...
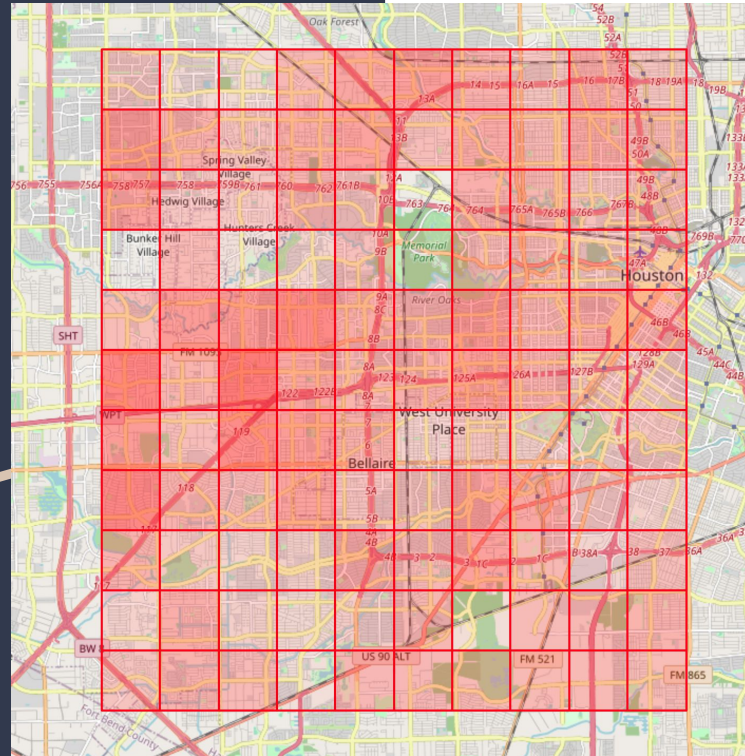
Resulting Clusters

The proof is in the pudding (clustering is consistent with my experience).
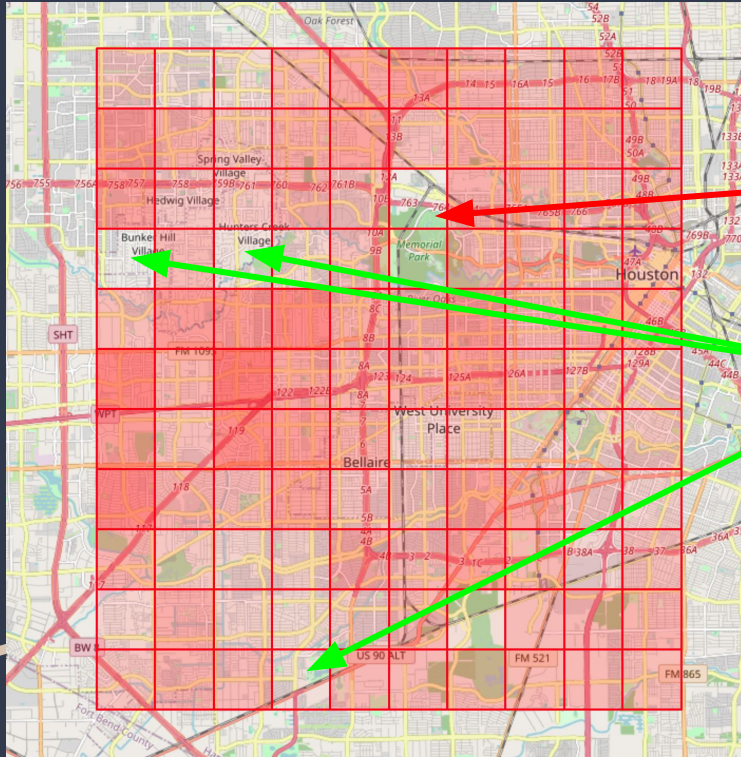
Restaurants versus peers

Map indicates quantity of restaurants relative to peer group (dark is more than peers, light is less).

Conclusions

The memorial park area doesn't have many restaurants relative to peers… but it's also a large park and does not have room for restaurants.

Some other areas (see arrows) are worth further investigation!

Other Insights

Most underrepresented venues (value is delta to peer mean)

```
Mexican Restaurant        -3.622222
Coffee Shop               -3.222222
Pizza Place               -3.153846
Sandwich Place            -2.861111
Fast Food Restaurant      -2.583333
Bar                       -2.488889
Grocery Store             -2.384615
American Restaurant       -2.355556
Italian Restaurant        -2.111111
Burger Joint              -1.977778
Name: min, dtype: float64
```

Where to open a coffee shop (value is delta to peer mean)

reg	lat	lon	Coffee Shop
85	29.780244	-95.442192	-3.222222
66	29.746244	-95.423192	-3.222222
73	29.763244	-95.480192	-2.222222
96	29.797244	-95.423192	-2.222222
90	29.797244	-95.537192	-2.222222

I also looked at what other venues are particularly underrepresented in the search area.

There are also a few regions lacking coffee shops, which may be worth considering as a business opportunity.