



Choosing the best neighborhood
to move in with ML.
A practical case

Outline

- Introduction
- Data used / API used
- Methodology
- Conclusion
- Questions
- References

Introduction

- Background
 - People need to move / expatriation
- Problem
 - Where exactly ?
- Interest
 - Good location = good experience
- Scope of this work
 - Finding Best location
 - Given known appreciated neighborhood

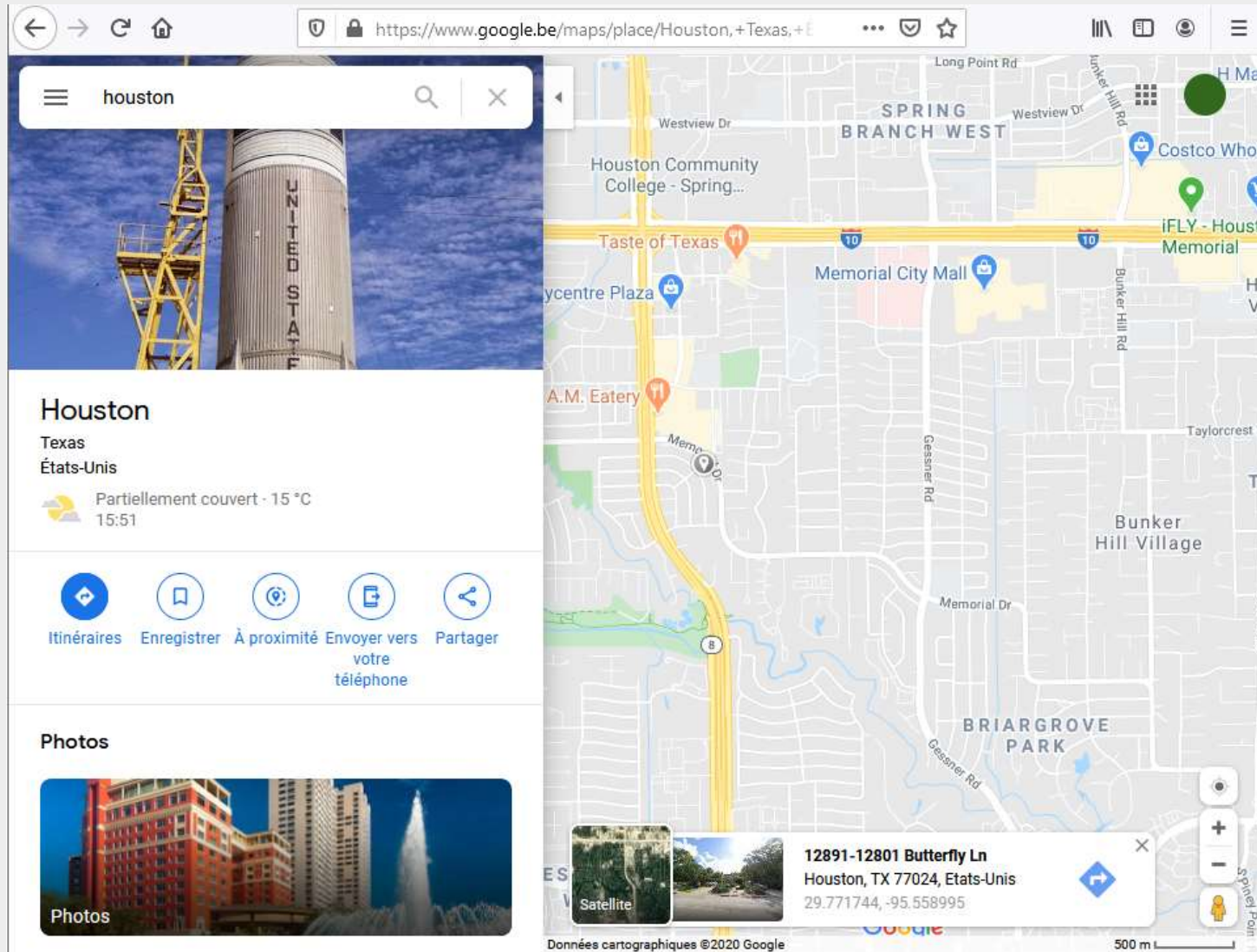
Data used

- Venues information
 - → Foursquare API
- Coordinate / Address
 - → Google Maps
 - → geopy python library
- Criminal record
 - → NIBRS
- Housing prices
 - → RedFin dataset

Data used - Foursquare

	Cell_index	Cell_Latitude	Cell_Longitude	Venue	Venue_Category	Venue_Latitude	Venue_Longitude
0	0	29.601804	-95.561416	Walgreens	Pharmacy	29.600119	-95.563614
1	0	29.601804	-95.561416	Starbucks	Coffee Shop	29.599570	-95.563985
2	0	29.601804	-95.561416	Tornado Burger	Burger Joint	29.611505	-95.564204
3	0	29.601804	-95.561416	El Vaquero	Mexican Restaurant	29.589083	-95.564442
4	0	29.601804	-95.561416	CVS pharmacy	Pharmacy	29.600460	-95.565012

Data used – Google maps

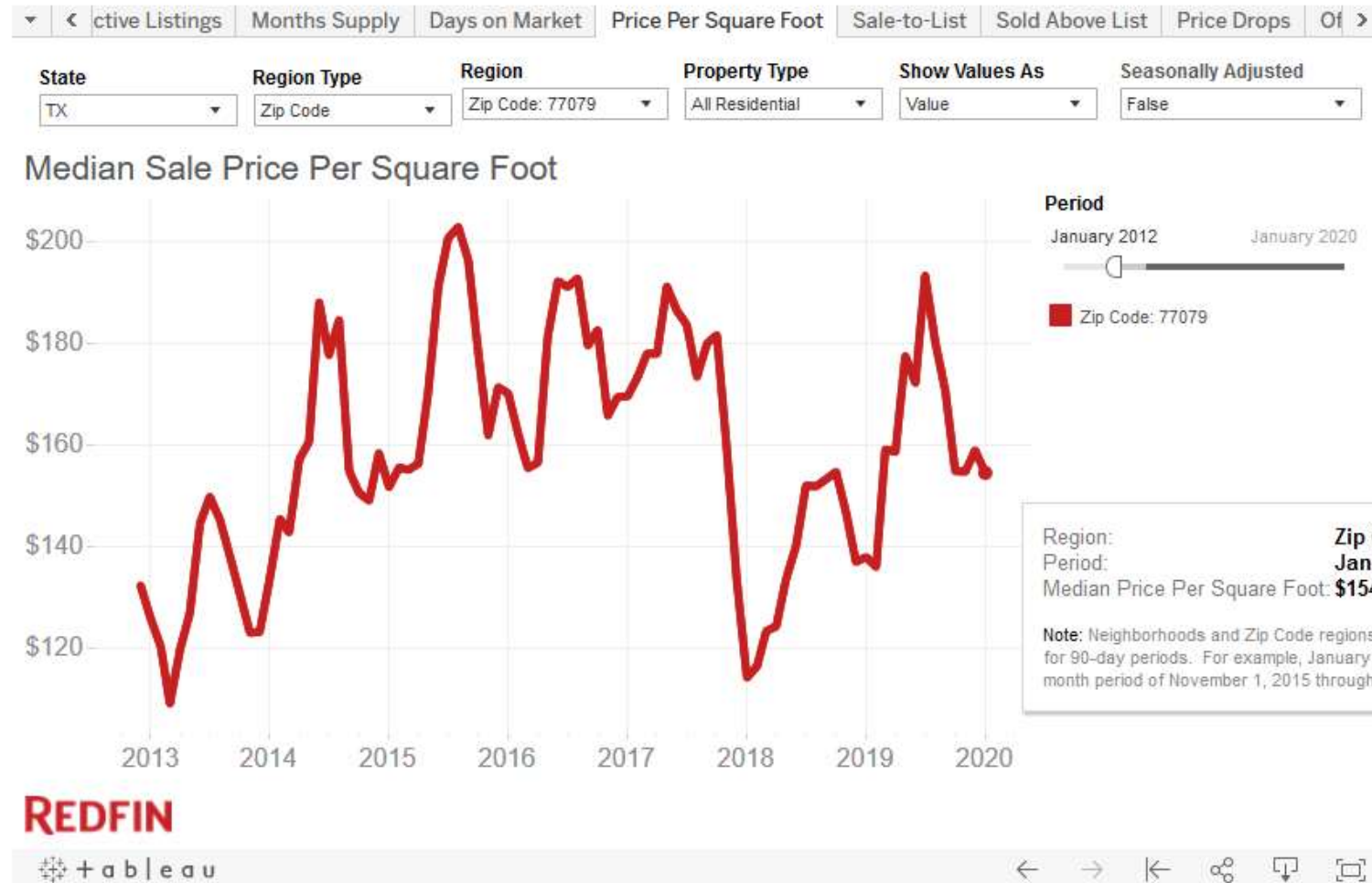


Data used - records

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Incident	Occurrence Date	Occurrence Hour	NIBRS Class	NIBRS Description	Offense Count	Beat	Premise	Block Range	Street Name	Street Type	Suffix	ZIP Code
2	5619	01/01/2019	0	290	Destruction, damage, vandalism		19C30	Residence, Home (Includes Apartment)	9622	SAN CARLOS			77013
3	17319	01/01/2019	0	35A	Drug, narcotic violations		17C10	Highway, Road, Street, Alley	0	EAST	FWY		77020
4	18119	01/01/2019	0	290	Destruction, damage, vandalism		116E40	Residence, Home (Includes Apartment)	16718	LONE QUAIL	CT		77489
5	19019	01/01/2019	0	520	Weapon law violations		1	Residence, Home (Includes Apartment)	1909	MELBOURNE			77026-00
6	20519	01/01/2019	0	13A	Aggravated Assault		115E30	Residence, Home (Includes Apartment)	4034	OSBY	DR		77025
7	20519	01/01/2019	0	23H	All other larceny		115E30	Residence, Home (Includes Apartment)	4034	OSBY	DR		77025
8	20519	01/01/2019	0	290	Destruction, damage, vandalism		115E30	Residence, Home (Includes Apartment)	4034	OSBY	DR		77025
9	20519	01/01/2019	0	35A	Drug, narcotic violations		115E30	Residence, Home (Includes Apartment)	4034	OSBY	DR		77025
10	31119	01/01/2019	0	13B	Simple assault		119G20	Residence, Home (Includes Apartment)	4700	KIRKWOOD	RD	S	77072
11	31119	01/01/2019	0	290	Destruction, damage, vandalism		119G20	Residence, Home (Includes Apartment)	4700	KIRKWOOD	RD	S	77072
12	31119	01/01/2019	0	13A	Aggravated Assault		219G20	Residence, Home (Includes Apartment)	4700	KIRKWOOD	RD	S	77072
13	34419	01/01/2019	0	13B	Simple assault		15F10	Residence, Home (Includes Apartment)	7429	LONG POINT	RD		77055
14	34819	01/01/2019	0	290	Destruction, damage, vandalism		115E30	Residence, Home (Includes Apartment)	4065	SILVERWOOD	DR		77025
15	34819	01/01/2019	0	13B	Simple assault		215E30	Residence, Home (Includes Apartment)	4065	SILVERWOOD	DR		77025
16	39919	01/01/2019	0	13B	Simple assault		120G10	Residence, Home (Includes Apartment)	3500	WOODCHASE	DR		77042
17	40219	01/01/2019	0	13A	Aggravated Assault		113D20	Residence, Home (Includes Apartment)	6502	ROXBURY	RD		77087
18	40519	01/01/2019	0	13B	Simple assault		119G20	Residence, Home (Includes Apartment)	6819	COOK	RD		77072
19	41419	01/01/2019	0	290	Destruction, damage, vandalism		17C20	Residence, Home (Includes Apartment)	4521	KASHMERE	ST		77026
20	43019	01/01/2019	0	35A	Drug, narcotic violations		110H40	Highway, Road, Street, Alley	1160	GRAY	ST		77002
21	43019	01/01/2019	0	35B	Drug equipment violations		110H40	Highway, Road, Street, Alley	1160	GRAY	ST		77002
22	43719	01/01/2019	0	13B	Simple assault		211H40	Residence, Home (Includes Apartment)	1114	CHOATE	CIR		77017
23	44419	01/01/2019	0	13B	Simple assault		210H40	Bar, Nightclub	2600	TRAVIS	ST		77006
24	45919	01/01/2019	0	120	Robbery		11A30	Highway, Road, Street, Alley	2200	SOUTHWEST	FWY		77098

Data Used – House prices

Home Prices, Sales & Inventory

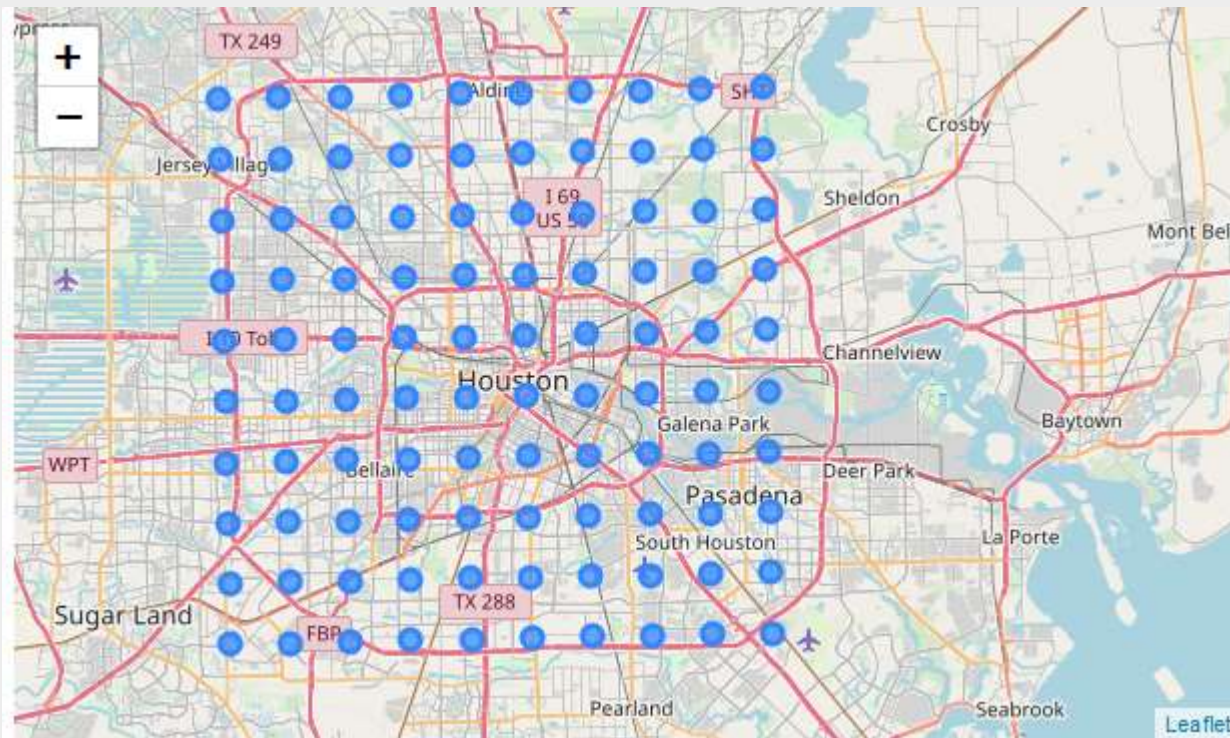


Methodology

- Find best location
 - Based on similar venues between
 - Destination town
 - Origin address or address of reference
 - Based on House price and criminality
 - → safer place for a given budget

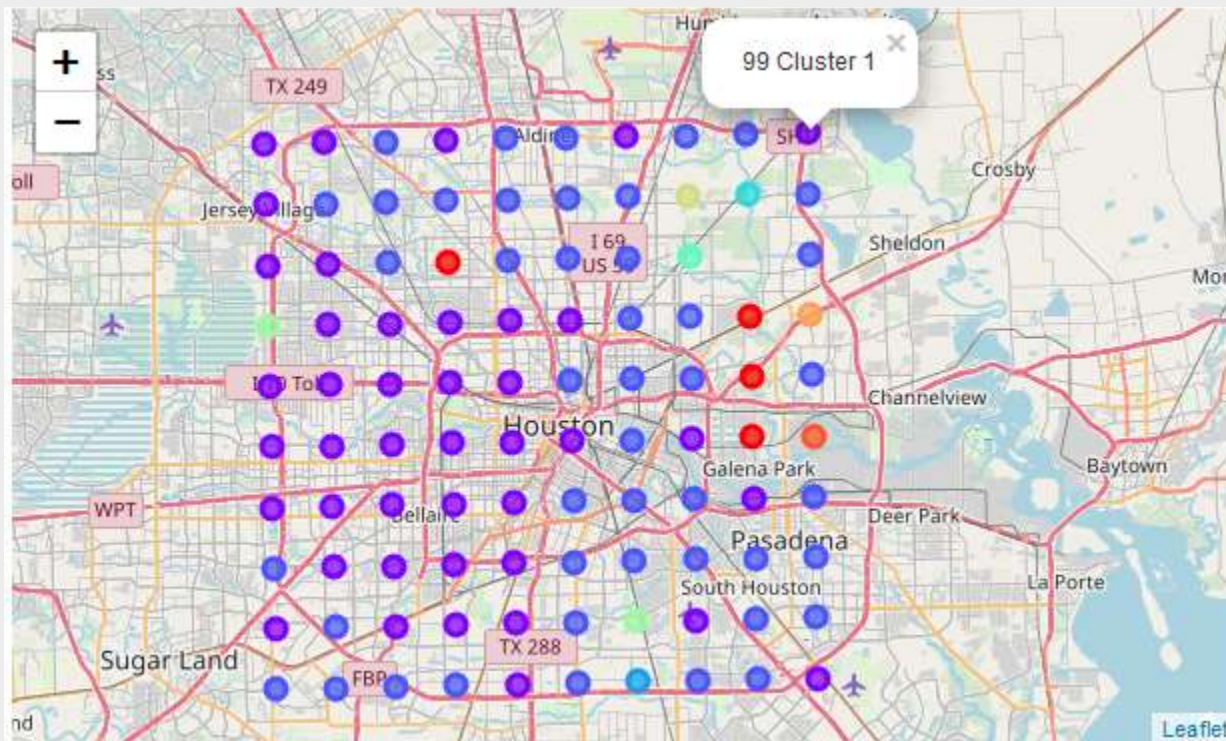
Methodology

- Divide Destination town in sub cells
 - 100 cells
 - 4x4km wide



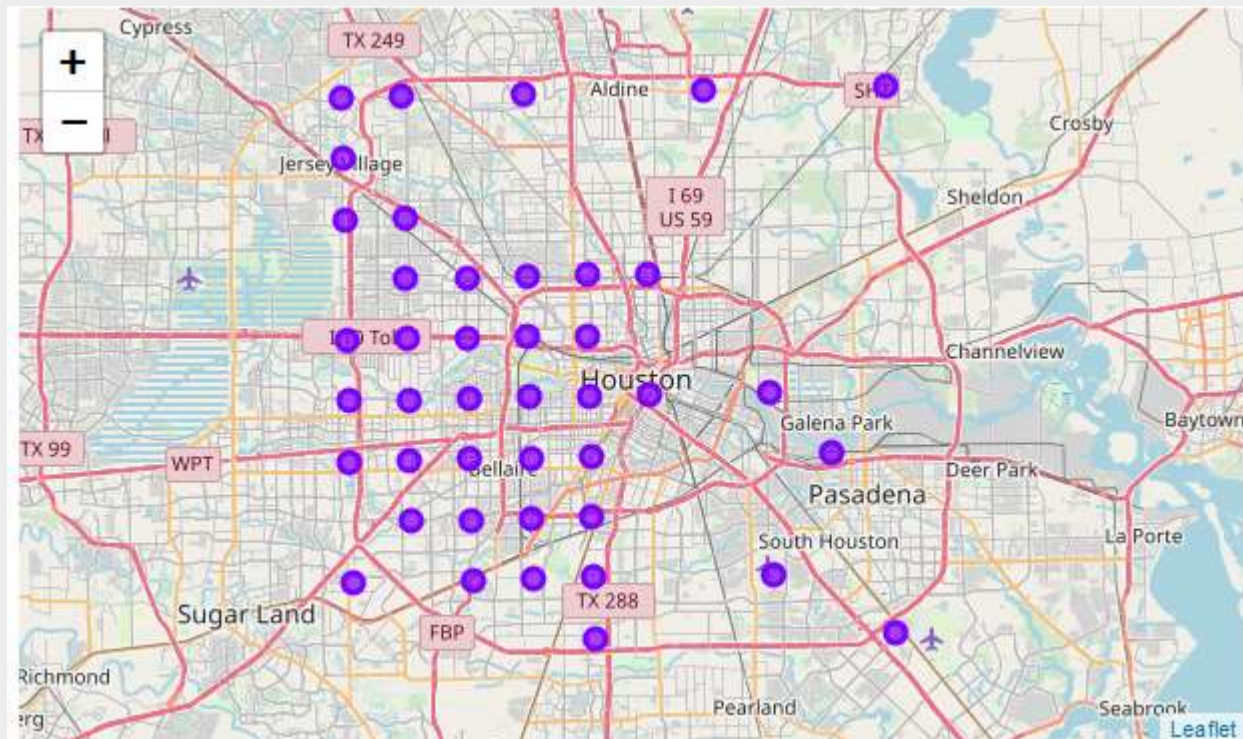
Methodology

- Clusterize according to nearby venues
 - K-means ; $K = 10$



Methodology

- Filter on what cluster is reference origin address



Methodology

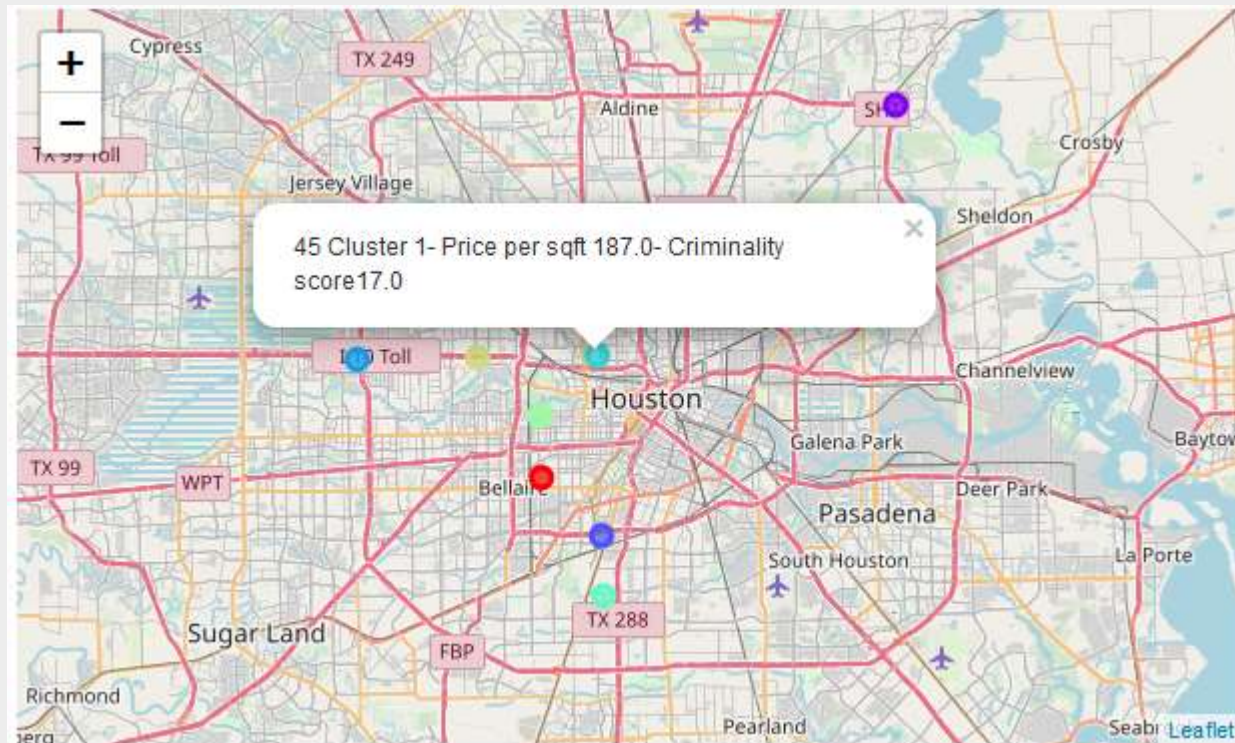
- Split on price range



Purple		: cat_0 :	87.0	-	114.3\$
Blue		: cat_1 :	114.3	-	141.6\$
Blue2		: cat_2 :	141.6	-	168.9\$
Blue3		: cat_3 :	168.9	-	196.2\$
Green		: cat_4 :	196.2	-	223.5\$
Green2		: cat_5 :	223.5	-	250.8\$
Yellow		: cat_6 :	250.8	-	278.1\$
(None in that cat):		cat_7 :	278.1	-	305.4\$
(None in that cat):		cat_8 :	305.4	-	332.7\$
Red		: cat_9 :	332.7	-	361.0\$

Methodology

- Refine according to Criminal data
 - Keep safest by price range



Methodology

- Assigned committed crime to cell
- Give score to cell by committed crime
 - 1. no physical damage (theft, fraud) = 1 point
 - 2. simple assault = 2 points
 - 3. aggravated assault = 3 points
 - 4. Rape = 5 points
 - 5. Homicide = 10 points

Conclusion

- ML can help people to make choice
- Some trigger to choice / emotion cannot be modeled
- Study here only take into account
 - Venues
 - Price
 - Criminality
- There is other features to take account

Conclusion

- Next steps
 - Divide in smaller cells → need higher API privilege (\$\$)
 - Extend venue radius for analysis (Foursquare limitation toward number of request)
 - Take venue distances from cell into account.
 - The closer the better (or worse depending of type of venue)
 - Criminality, use a larger time frame
 - Criminality instead of assigning to closest cell, account for each crime committed in a certain radius
 - Review criminality scoring
 - Adds other type of features to better pin point user interest.

Questions / Answers

References

<https://developer.foursquare.com>

https://www.houstontx.gov/police/cs/Monthly_Crime_Data_by_Street_and_Police_Beat.htm

<https://www.google.be/maps/>

<https://www.redfin.com/blog/data-center/>