Battle of Neighborhood

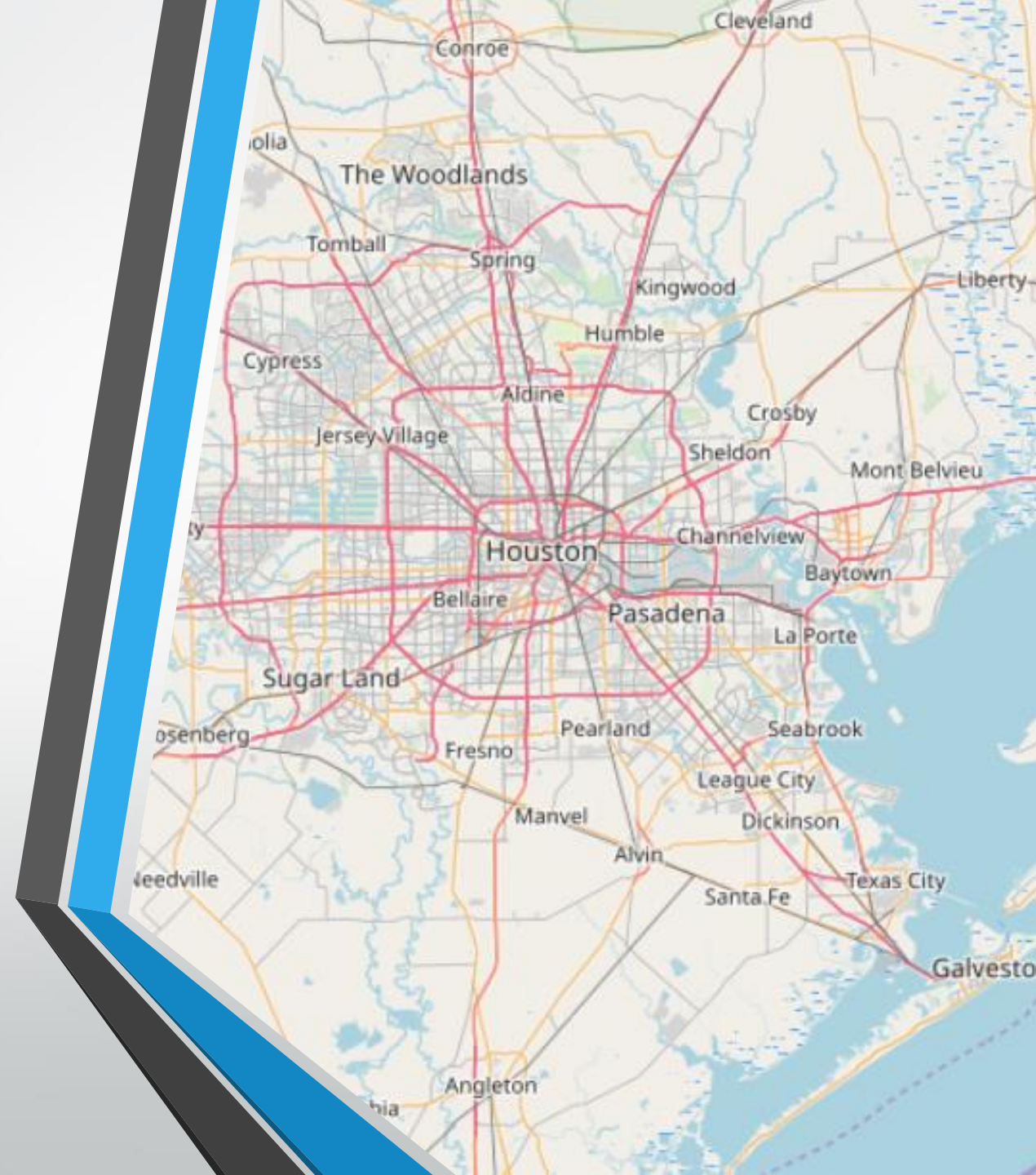## Best Neighborhood for Family in Houston

XYZ Company Ltd

Date – 19 Jan 2020

Business Problem

Recommend Best Neighborhood in Houston for families based on below criteria –

- Low Crime Rate

- Reasonable Home Prices

- Good Public School Ratings

Data

- All Zip Codes/Neighborhood in Houston
- Home Price data for each Zip Code
- Crime data in every Zip Code
- Average School Rating in every Zip Code
- Latitude and Longitude of each Zip Code
- The Number of venues in each Zip Code from Foursquare APIs based on Latitude and Longitude of Zip Code

Home Prices Data

	Neighborhood	Zip	HomePrice
0	1960/Cypress	77065	$179,000
1	Aldine Area	77039	$133,500
2	Alief	77072	$164,000
3	Alvin North	77511	$227,000
4	Alvin South	77511	$163,900

- **Data Source** –

  Website of Houstonia Magazine

  https://www.houstoniamag.com/articles/2017/3/24/neighborhoods-by-the-numbers-real-estate-data-2017

- **Method**

  Scrape the page using BeautifulSuop4 Library

Crime Data

- **Data Source** –

  Excel file from Website of Houston Police Department

  http://www.houstontx.gov/police/cs/xls/06-2019.NIBRS_Public_Data_Group_A&B.xlsx

- **Method**

  Read Excel

School Accountability Ratings

	Zip	Crimes
0	75248	1
1	77002	547
2	77003	183
3	77004	520
4	77005	91

- **Data Source**

  On ArcGIS website uploaded by Texas Education Agency

  https://opendata.arcgis.com/datasets/6cf4436417ff43d0a6e741dc83339ae2_0.csv

- **Method**

  Read CSV then filter for Houston City

  Average the ratings for each zip code

Latitude and Longitudes of Zip Codes

	Zip	Latitude	Longitude
0	71937	34.398483	-94.39398
1	72044	35.624351	-92.16056
2	56171	43.660847	-94.74357
3	49430	43.010337	-85.89754
4	52585	41.194129	-91.98027

- **Data Source**

  In CSV format on website of OpenDataSoft

  https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/download/?format=csv&timezone=America/Chicago&use_labels_for_header=true&csv_separator=%3B

- **Method**

  Read CSV then filter for Houston City

Venues

- **Data Source**

  Foursquare API

  https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}

  Method

  API call for each Zip Code using its latitude and longitude

  Count the venues in the zip code
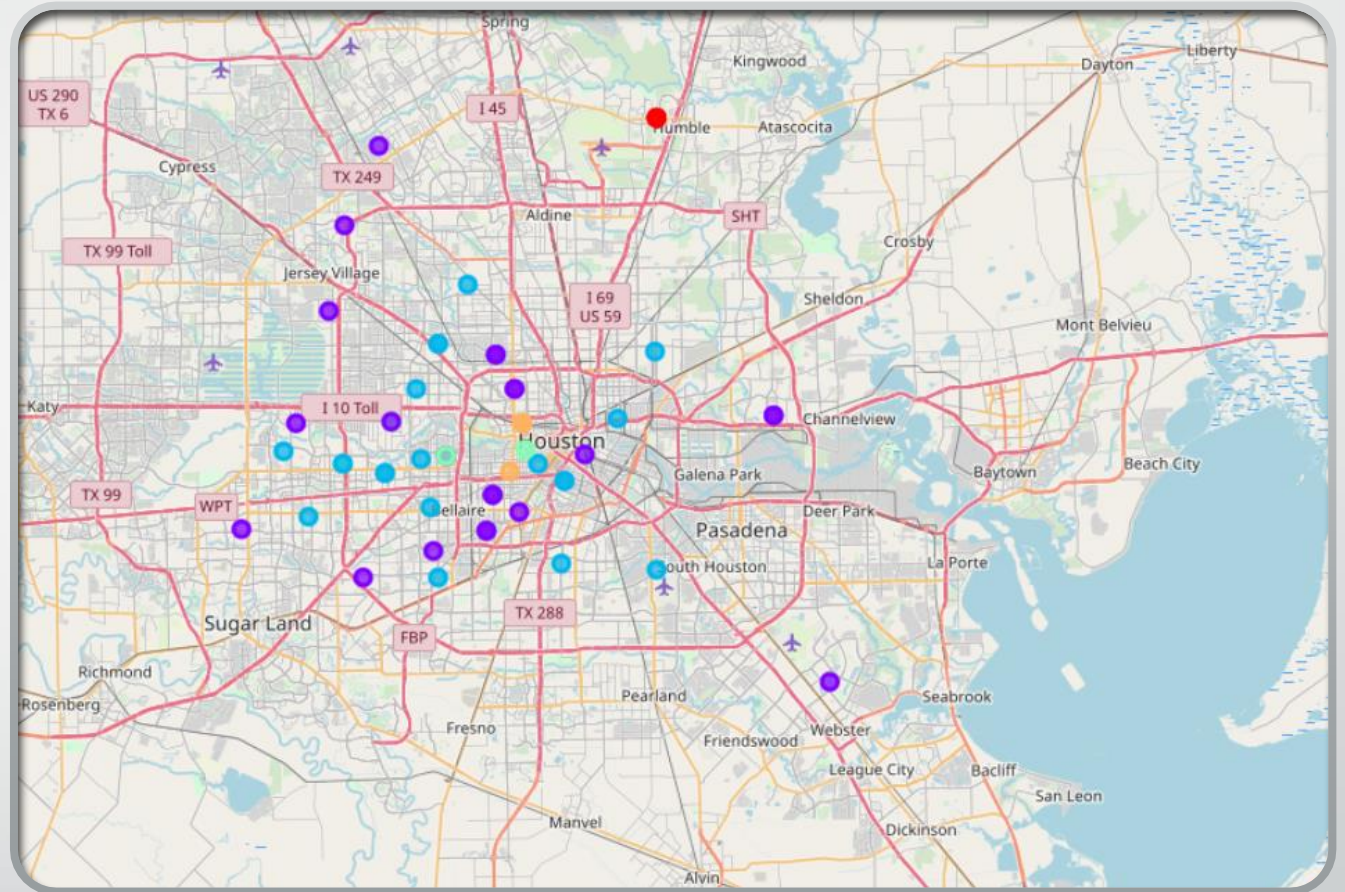
Final Dataset

- **Merge all data**

	Neighborhood	Zip	HomePrice	Latitude	Longitude	Crimes	Rating	Venues
2	Alief	77072	$164,000	29.700898	-95.59002	426.0	4.181818	9.0
13	Braeswood Place	77025	$715,000	29.690230	-95.43474	225.0	5.166667	5.0
14	Brays Oaks	77031	$225,000	29.654132	-95.54311	146.0	5.000000	2.0
15	Briargrove	77057	$824,000	29.745129	-95.49131	432.0	4.666667	4.0
16	Briargrove Park/Walnut Bend	77042	$460,000	29.741565	-95.55996	421.0	4.200000	12.0

- **Normalize using Min Max**

	Neighborhood	Zip	HomePrice	Latitude	Longitude	Crimes	Rating	Venues
2	Alief	77072	0.026221	29.700898	-95.59002	0.816764	0.480519	0.106667
13	Braeswood Place	77025	0.224138	29.690230	-95.43474	0.424951	0.761905	0.053333
14	Brays Oaks	77031	0.048132	29.654132	-95.54311	0.270955	0.714286	0.013333
15	Briargrove	77057	0.263290	29.745129	-95.49131	0.828460	0.619048	0.040000
16	Briargrove Park/Walnut Bend	77042	0.132543	29.741565	-95.55996	0.807018	0.485714	0.146667

Methodology

- Run *k*-means to cluster the neighborhood into 5 clusters

- Analyze each cluster

- Recommend best 4-5 neighborhoods/zip codes

Analyze Cluster 1

- Neighborhoods in this cluster have low home prices and low crime rates.

- Public school ratings are not available.

- So we are unable to recommend them.

	Neighborhood	Zip	HomePrice	Latitude	Longitude	Crimes	Rating	Venues
60	Humble Area East	77338	0.025144	30.005691	-95.28488	0.033138	0.0	0.426667
61	Humble Area South	77338	0.006214	30.005691	-95.28488	0.033138	0.0	0.426667
62	Humble Area West	77338	0.021552	30.005691	-95.28488	0.033138	0.0	0.426667
70	Kingwood South	77338	0.053520	30.005691	-95.28488	0.033138	0.0	0.426667

Analyze Cluster 2

- When we examine cluster 2, we find that in this cluster, the home price is on lower side except Memorial Villages, Crime rate is lower except Sharpstown area and Number of Venues are also lower. Some of the neighborhoods seems interesting -
  - Medical Center Area
  - Memorial Villages
  - West University
  - Rice/Museum District
- These neighborhoods have higher school ratings, lower crimes but have lower venues for other activities. Further Medical center area has lower home prices while others have higher home prices.
- We would recommend Zip Code - 77030 and 77005 for the families who are okay with little less venues for activities in their own neighborhood.

	Neighborhood	Zip	HomePrice	Latitude	Longitude	Crimes	Rating	Venues
13	Braeswood Place	77025	0.224138	29.690230	-95.43474	0.424951	0.761905	0.053333
14	Brays Oaks	77031	0.048132	29.654132	-95.54311	0.270955	0.714286	0.013333
21	Champions Area	77069	0.051435	29.984672	-95.52887	0.000000	0.714286	0.053333
23	Clear Lake Area	77062	0.051006	29.574930	-95.13238	0.097466	0.771429	0.013333
40	East End Revitalized	77003	0.058190	29.749278	-95.34741	0.343080	0.857143	0.080000
41	Eldridge North	77041	0.043642	29.858730	-95.57243	0.120858	0.785714	0.053333
49	Galleria	77056	0.209770	29.747328	-95.46931	0.637427	0.809524	0.186667
50	Garden Oaks	77015	0.127155	29.778526	-95.18118	0.294347	0.688312	0.013333
53	Heights/Greater Heights	77008	0.132507	29.798777	-95.40951	0.481481	0.714286	0.226667
72	Knollwood/Woodside Area	77025	0.121767	29.690230	-95.43474	0.424951	0.761905	0.053333
81	Medical Center Area	77030	0.084052	29.704584	-95.40466	0.249513	1.000000	0.146667
85	Memorial Villages	77024	0.458724	29.773994	-95.51771	0.368421	0.959184	0.000000
86	Memorial West	77079	0.209770	29.773018	-95.60125	0.360624	0.714286	0.293333
87	Meyerland Area	77096	0.106501	29.674336	-95.48123	0.539961	0.771429	0.160000
90	Mission Bend Area	77083	0.028736	29.691714	-95.64978	0.017544	0.587302	0.106667
94	North Channel	77015	0.016613	29.778526	-95.18118	0.294347	0.688312	0.013333
98	Oak Forest East Area	77018	0.106681	29.825476	-95.42619	0.346979	0.510204	0.173333
107	Rice/Museum District	77005	0.261566	29.717529	-95.42821	0.163743	1.000000	0.026667
114	Shepherd Park Plaza Area	77018	0.120690	29.825476	-95.42619	0.346979	0.510204	0.173333
133	Timbergrove/Lazybrook	77008	0.119971	29.798777	-95.40951	0.481481	0.714286	0.226667
142	West University/Southside Area	77005	0.395474	29.717529	-95.42821	0.163743	1.000000	0.026667
146	Willowbrook	77064	0.022450	29.923638	-95.55919	0.116959	0.673469	0.026667

Analyze Cluster 3

- When we examine cluster 3, we find that in this cluster, the home prices are on relatively reasonable side, Crime rate is average. School ratings are high except Upper Kirby. These neighborhoods have many venues for different activities.

- We would like to recommend all neighborhoods in Zip Code - 77007

	Neighborhood	Zip	HomePrice	Latitude	Longitude	Crimes	Rating	Venues
2	Alief	77072	0.026221	29.700898	-95.59002	0.816764	0.480519	0.106667
15	Briargrove	77057	0.263290	29.745129	-95.49131	0.828460	0.619048	0.040000
16	Briargrove Park/Walnut Bend	77042	0.132543	29.741565	-95.55996	0.807018	0.485714	0.146667
17	Briarmeadow/Tanglewilde	77063	0.072198	29.734379	-95.52269	0.771930	0.673469	0.293333
22	Charnwood/Briarbend	77063	0.156789	29.734379	-95.52269	0.771930	0.673469	0.293333
37	Denver Harbor	77020	0.005029	29.775927	-95.31836	0.526316	0.532468	0.146667
42	Energy Corridor	77077	0.087644	29.750897	-95.61255	0.851852	0.510204	0.040000
52	Gulfton	77081	0.063398	29.708280	-95.48361	0.697856	0.693878	0.226667
57	Hobby Area	77061	0.012033	29.660280	-95.28446	0.703704	0.542857	0.120000
82	Medical Center South	77051	0.011135	29.665430	-95.36871	0.530214	0.142857	0.040000
89	Midtown-Houston	77004	0.109914	29.728779	-95.36570	1.000000	0.480519	0.066667
92	Montrose	77006	0.188218	29.741878	-95.38944	0.686160	0.428571	0.440000
95	Northeast Houston	77028	0.000000	29.827315	-95.28631	0.426901	0.357143	0.026667
96	Northside	77092	0.005029	29.833326	-95.47644	0.912281	0.357143	0.000000
97	Northwest Houston	77088	0.017241	29.879213	-95.45028	0.789474	0.357143	0.000000
99	Oak Forest West Area	77092	0.057112	29.833326	-95.47644	0.912281	0.357143	0.000000
110	Riverside	77004	0.078305	29.728779	-95.36570	1.000000	0.480519	0.066667
118	Spring Branch	77055	0.084052	29.798877	-95.49629	0.643275	0.514286	0.053333
143	Westchase Area	77042	0.184896	29.741565	-95.55996	0.807018	0.485714	0.146667
145	Willow Meadows Area	77035	0.078664	29.654108	-95.47692	0.508772	0.200000	0.146667

Analyze Cluster 4

- Neighborhoods in this cluster have highest home prices. Public school ratings are not the best except 77024. Also number of venues in the neighborhoods are not that many.

- We would recommend Zip Code - 77024 for affluent families who are okay with little less venues for activities in their own neighborhood.

	Neighborhood	Zip	HomePrice	Latitude	Longitude	Crimes	Rating	Venues
83	Memorial Close In	77024	0.810776	29.773994	-95.51771	0.368421	0.959184	0.000000
108	River Oaks Area	77019	0.755388	29.752528	-95.39923	0.401559	0.714286	0.346667
109	Rivercrest	77042	1.000000	29.741565	-95.55996	0.807018	0.485714	0.146667
129	Tanglewood Area	77056	0.557651	29.747328	-95.46931	0.637427	0.809524	0.186667

Analyze Cluster 5

- Neighborhoods in this cluster have highest home prices. Public school ratings are not the best except 77024. Also number of venues in the neighborhoods are not that many.

- We would recommend Zip Code - 77024 for affluent families who are okay with little less venues for activities in their own neighborhood.

	Neighborhood	Zip	HomePrice	Latitude	Longitude	Crimes	Rating	Venues
83	Memorial Close In	77024	0.810776	29.773994	-95.51771	0.368421	0.959184	0.000000
108	River Oaks Area	77019	0.755388	29.752528	-95.39923	0.401559	0.714286	0.346667
109	Rivercrest	77042	1.000000	29.741565	-95.55996	0.807018	0.485714	0.146667
129	Tanglewood Area	77056	0.557651	29.747328	-95.46931	0.637427	0.809524	0.186667

Results and Discussion

- Going by the results we see it is very difficult to recommend one single zip code where we get all 4 things together – Low Home Price, School Ratings, Lot of venues and Lower Crimes. It is generally observed that home prices would be higher in those area where schools are good, crimes are low and lot of venues to explore. But again, good schools and low crimes are very important so we will not compromise with these two attributes. Home prices as optional choices for family depending on the affluency.

- Considering all of this, we will finally recommend 4 zip codes. Families can choose based on their priorities –

  - **77007** -We would like to recommend all neighborhoods in Zip Code. This zip code has all favorable attributes except Memorial Park are in this zip code may have costly homes.

  - **77024** - We would recommend this zip code for affluent families who are okay with little less venues for activities in their own neighborhood.

  - **77030 and 77005** - We would recommend zip code for the families who are okay with little less venues for activities in their own neighborhood.

Conclusion

- This analysis has been done with limited set of data and conserving the data which is available is correct.

- There are many other things which impact suitability of a neighborhood for a family like distance from office etc. Also, affluent facilities put their kids in private schools also.

- So before recommending a neighborhood for a family this individual, preferences should also be considered. But certainly, the above analysis would form a base for recommendation.