Finding similar neighborhoods in Raleigh, NC

by Anton Bezuglov

The Problem

- Average American moves 11.4 times in their lifetime
- How to find the best neighborhood at the new address?
- Let us find one similar to the previous residence!
- Use socioeconomic and nearby venues data to decide similarity

Murraywood and Raleigh, NC

- Use Murraywood (Columbia, SC) as the previous residence
- Neighborhoods in Raleigh, NC -- the new residence area

The Data: List of neighborhoods in Raleigh

- Scrape from City-data.com
- Choose those with 'Raleigh' in their name
- Save

The Data: Socioeconomic data
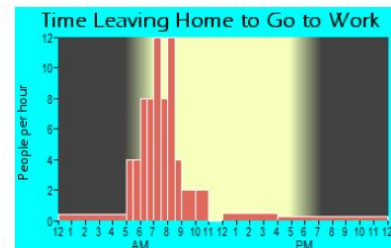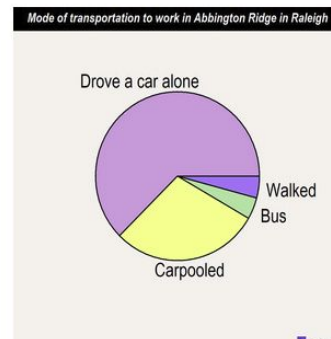
- Scrape from city-data.com
- Save

The Data: Venues

- Convert neighborhood name to coordinates
- Use Foursquare API to pull venues data

FOURSQUARE DEVELOPERS

Personal Account

- 99,500 Regular Calls / Day
- 500 Premium Calls / Day
- 1 Photo per Venue
- 1 Tip per Venue

Looking to build something for a commercial application?
Click below:

Upgrade Now ›

The Data: Joined dataset (the first 5 rows shown)

|	area	population	avg household size	avg number of cars apts	avg number of cars houses	males	med age females	med age males	med household income	med rent	...
Neighborhood											
Alanda, Raleigh, NC	0.019	4	21.0	1.7	1.7	2.0	36.7	30.7	65332.0	853.0	...
Ashebrook, Raleigh, NC	0.056	16	39.5	1.0	2.3	7.0	25.7	46.0	51273.0	619.0	...
Ashebrook, Raleigh, NC	0.056	16	39.5	1.0	2.3	7.0	25.7	46.0	51273.0	619.0	...
Aztec Park, Raleigh, NC	0.039	9	21.0	1.7	1.7	5.0	36.7	30.7	65332.0	853.0	...
Biltmore Hills, Raleigh, NC	0.164	2111	23.8	1.9	3.2	942.0	43.2	37.2	51299.0	734.0	...

5 rows × 29 columns

Exploratory Analysis

- Find errors/irregularities in the data
- Detect outliers
- See dependencies
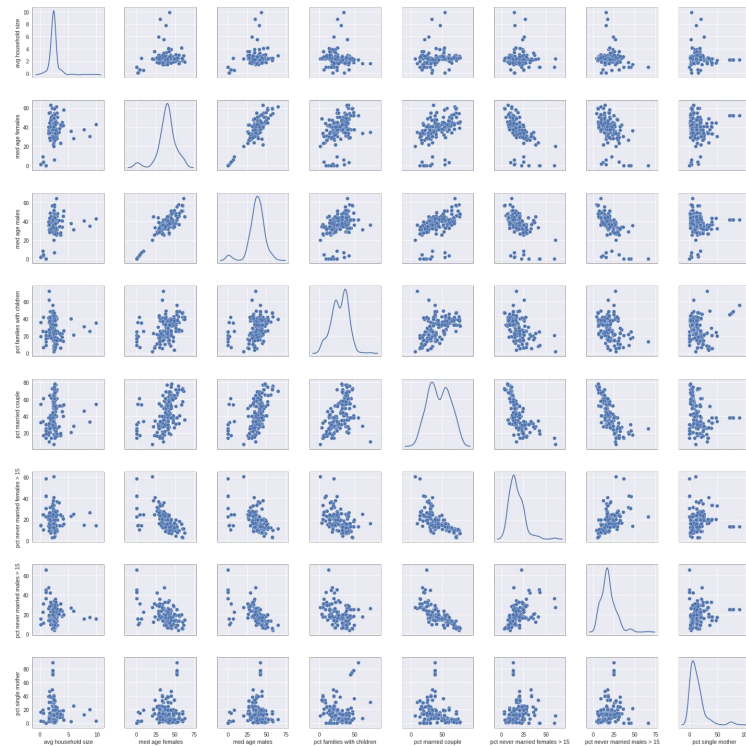
Machine Learning: K-Mean Clustering

- Find clusters of neighborhoods
- See what cluster Murraywood belongs to
- See other neighborhoods from this cluster
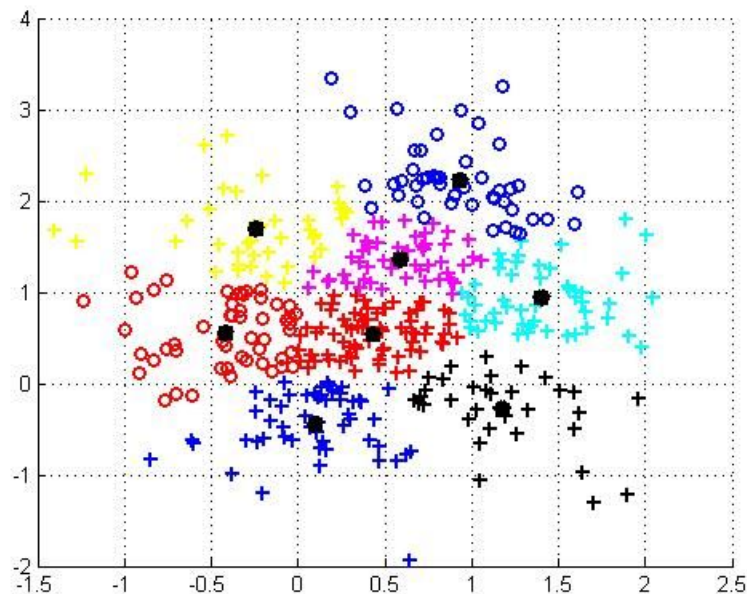
What is the best number of clusters?

- Elbow method -- see where the chart has 'elbow'
- ~15 clusters

Average distances from points to their cluster centers

Clusters (#11 -- Murraywood)

cluster	avg household size	med age females	med age males	pct families with children	pct married couple	pct never married females > 15	pct never married males > 15	pct single mother	med household income	med rent	pct units mortgage
6	2.500000	49.450000	54.750000	40.600000	75.400000	10.400000	5.450000	4.300000	198252.500000	1125.000000	67.050000
3	2.766667	50.400000	52.133333	34.900000	72.900000	8.200000	7.833333	2.900000	162000.333333	1663.000000	67.033333
10	2.550000	50.662500	45.650000	37.075000	65.250000	9.537500	10.700000	2.812500	147450.000000	1829.250000	59.875000
8	2.476471	43.870588	45.488235	36.964706	66.147059	10.594118	10.694118	2.170588	122273.823529	1125.352941	70.376471
7	3.207692	43.453846	42.407692	32.815385	66.730769	9.392308	9.200000	5.484615	110807.384615	1057.846154	73.684615
1	2.368182	50.077273	49.736364	40.331818	64.713636	9.395455	11.486364	8.481818	99845.409091	1134.636364	68.950000
9	2.521739	42.847826	37.539130	36.552174	54.839130	13.169565	16.995652	5.878261	92447.782609	959.173913	70.295652
11	2.259375	41.784375	39.140625	31.781250	44.009375	15.981250	16.131250	9.303125	78821.031250	906.812500	64.787500
2	2.582927	39.882927	36.092683	30.163415	51.343902	13.785366	20.060976	13.758537	71296.097561	900.926829	74.239024
12	2.316667	38.011111	34.838889	28.333333	37.872222	20.727778	20.872222	10.955556	61236.666667	902.833333	70.961111
4	2.244643	37.955357	36.114286	26.928571	36.005357	21.123214	19.801786	13.662500	53235.696429	998.875000	73.412500

Murraywood cluster

- Murraywood has the second highest number of Parks
  - the first is cluster #3
- Murraywood has the second lowest number of units with mortgage (64%)
  - the first is cluster #10
- Murraywood is fairly close to convenience stores (5th closest)

- Recommend Murraywood cluster
- Recommend clusters #3 and #10

Map of Recommended Neighborhoods

- Red -- Murraywood cluster
- Yellow/Orange -- clusters #3 and #10
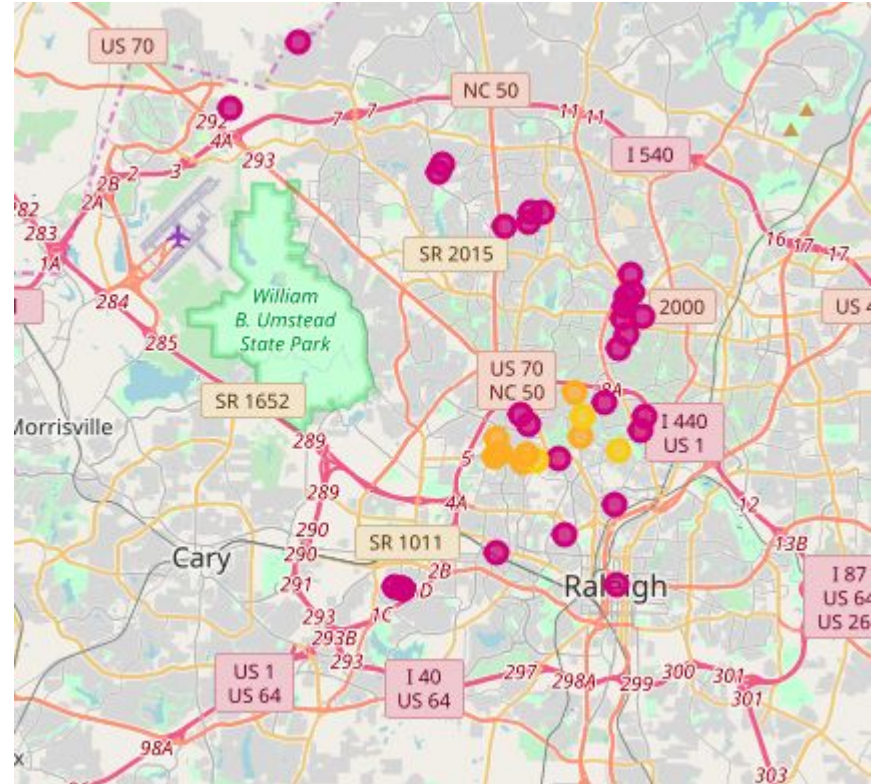
Conclusion

- Useful analysis, matches other research;
- Future work -- consider including neighborhoods in Durham, Cary, Chapel Hill, etc
- Future work -- include data on schools, police and nuisance reports, accidents data, etc.

Appendix

Python notebook on Github: https://bit.ly/2I0u1zb

Report: available upon request -- abezuglov@gmail.com