Finding a house
Making it home

Selecting homes based on personalized criteria

- Realty aggregators already provide a number of filters to help home buyers narrow their search – list price, year built, type of residence, and square footage to name a few.
- Many additional criteria go into selecting a home. This project uses four criteria to rank homes and further narrow choices to aid me in finding an ideal home.

1. Distance from family members
2. Distance from favorite store (Costco)
3. Variety of nearby venues
4. Purchasing power (square feet $\$^{-1}$)

Data acquisition

- Housing data scraped from Trulia.com. Prefiltered by:
  - List Price: $100,000-$320,000
  - Size: 1,750+ sqft
  - Bedrooms: 3+
  - Bathrooms: 2+
  - Location: Lafayette/West Lafayette, Lebanon, Crawfordsville, Kokomo, Zionsville, Westfield
- Trulia data collected:
  - List price
  - Square footage of home
  - Number of bedrooms
  - Number of bathrooms
  - Street Address
- GPS coordinates of homes obtained using Bing geocoder
- Venue data retrieved using Foursquare API

trulia.com/IN/Lafayette/

trulia | Lafayette, IN | Buy | Rent | Mortgage | Saved Homes | Saved Sear

Any Price | All Beds | All Home Types | More | Save Search

Lafayette, IN Homes For Sale & Real Estate

226 homes available on Trulia

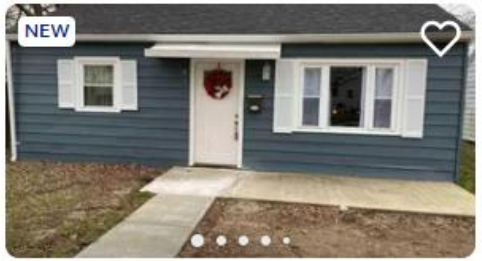
Sort: Just For You

$185,000
3bd 2ba 1,434 sqft
5004 Heritage Dr
Lafayette, IN

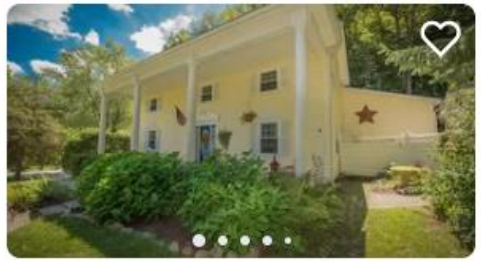
$359,900
3bd 4ba 3,340 sqft
5528 Blackberry Ln
Lafayette, IN

$309,900
4bd 3ba 2,382 sqft
3852 Basalt St
Lafayette, IN

$194,900
3bd 2ba 1,412 sqft
2814 Remington Dr
Lafayette, IN

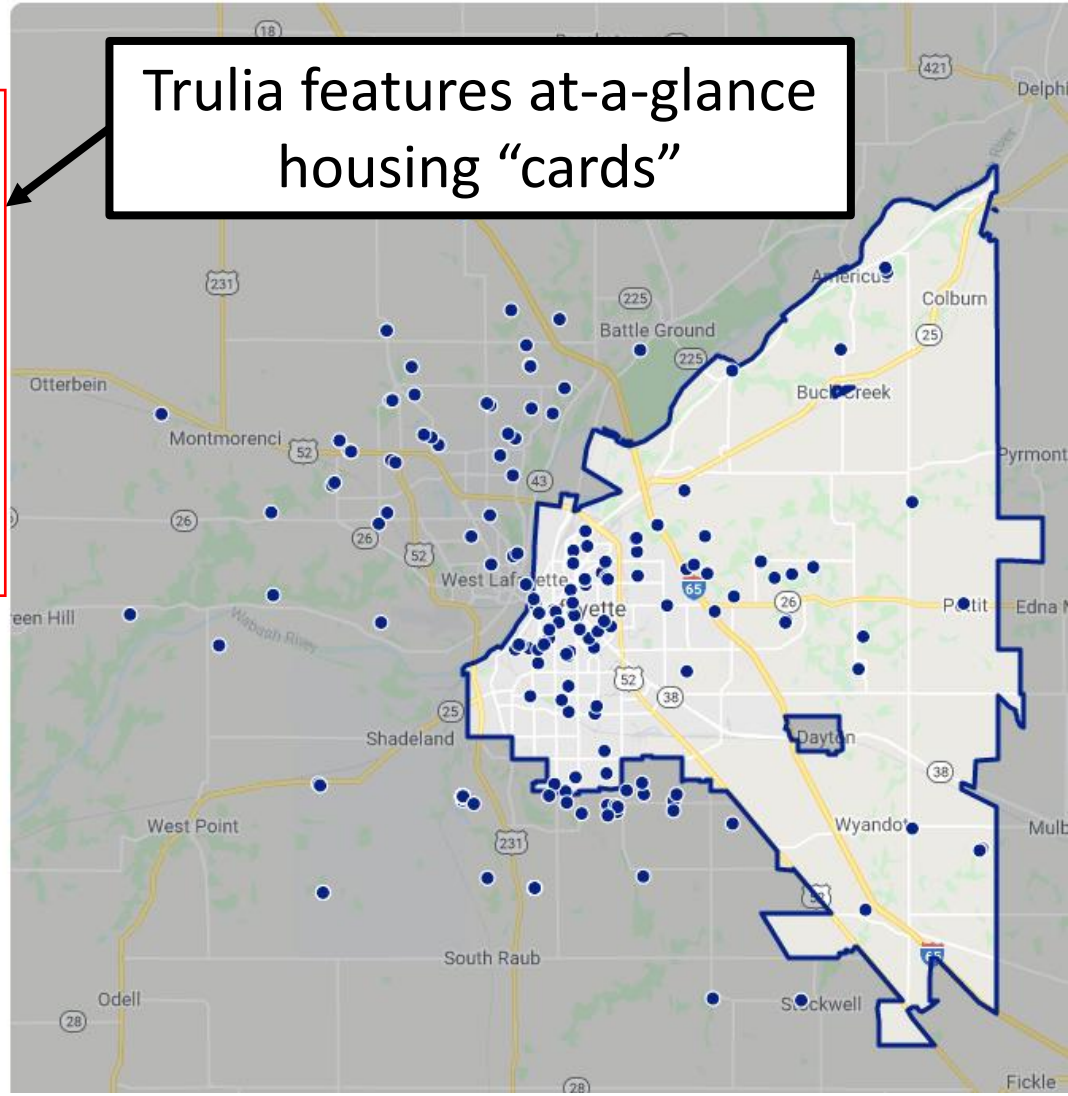


$137,500
4bd 1ba 1,064 sqft
1411 Oak Ct
Lafayette, IN



$350,000 ↓
4bd 4ba 2,903 sqft
4509 E 50 N
Lafayette, IN

Trulia features at-a-glance housing "cards"



trulia.com/IN/Lafayette/

trulia | Lafayette, IN | Buy | Rent | Mortgage | Saved Homes | Saved Sear

Any Price | All Beds | All Home Types | More | Save Search

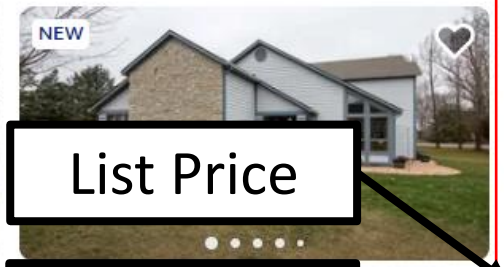
Lafayette, IN Homes For Sale & Real Estate

226 homes available on Trulia

Sort: Just For You

NEW

$185,000
3bd 2ba 1,434 sqft
5004 Heritage Dr
Lafayette, IN



NEW

5528 Blackberry Ln

NEW

$309,900
4bd 3ba 2,382 sqft
3852 Basalt St
Lafayette, IN

List Price

Bedrooms

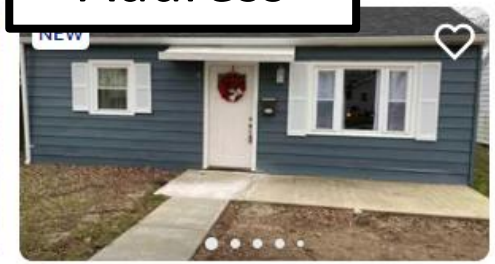
Address

Bathrooms

Size of home

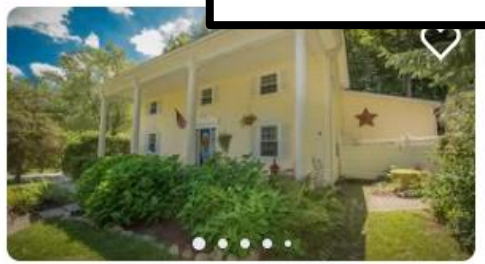
NEW

$194,900
3bd 2ba 1,412 sqft
2814 Remington Dr
Lafayette, IN

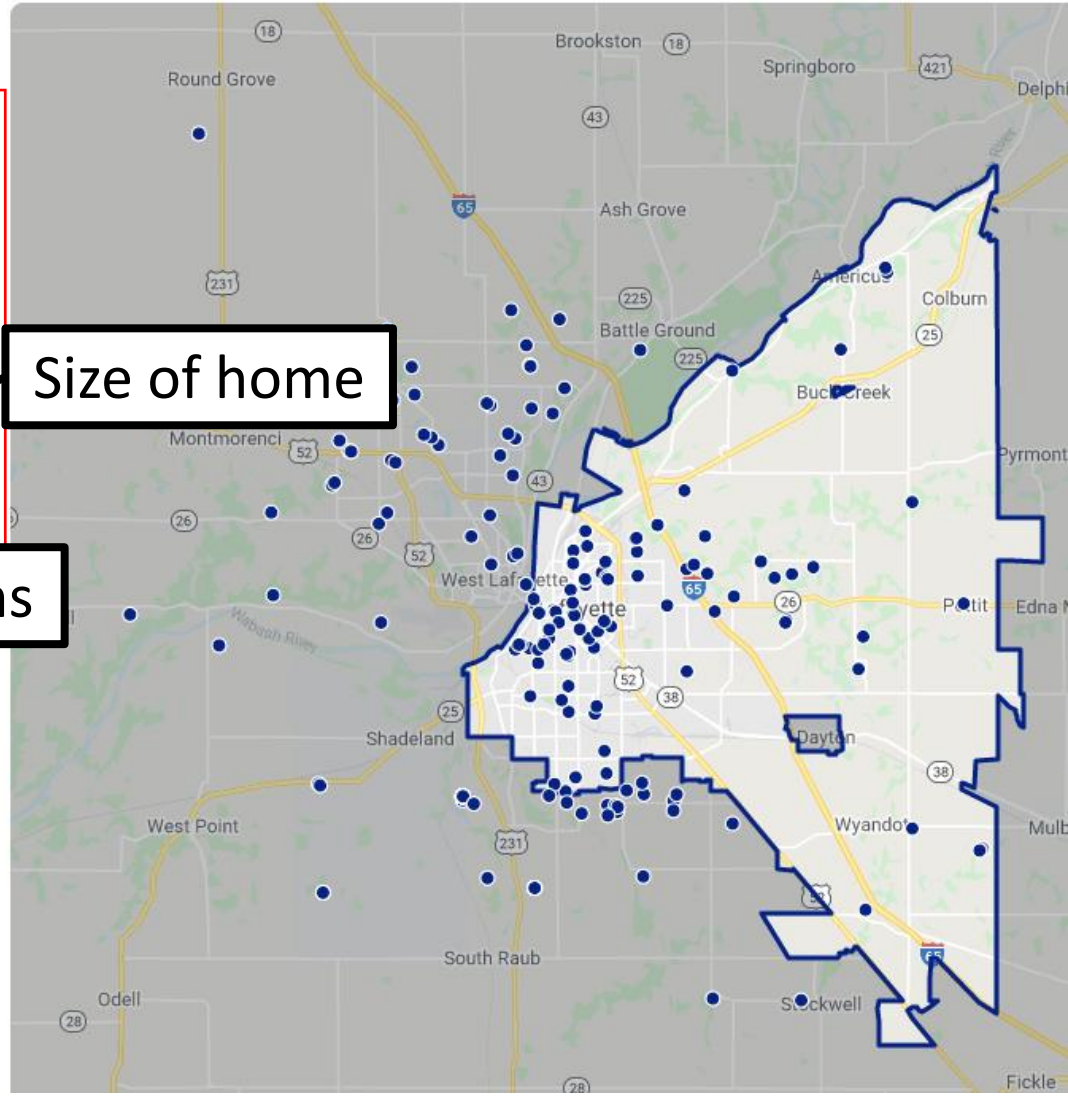


NEW

$137,500
4bd 1ba 1,064 sqft
1411 Oak Ct
Lafayette, IN



$350,000 ↓
4bd 4ba 2,903 sqft
4509 E 50 N
Lafayette, IN

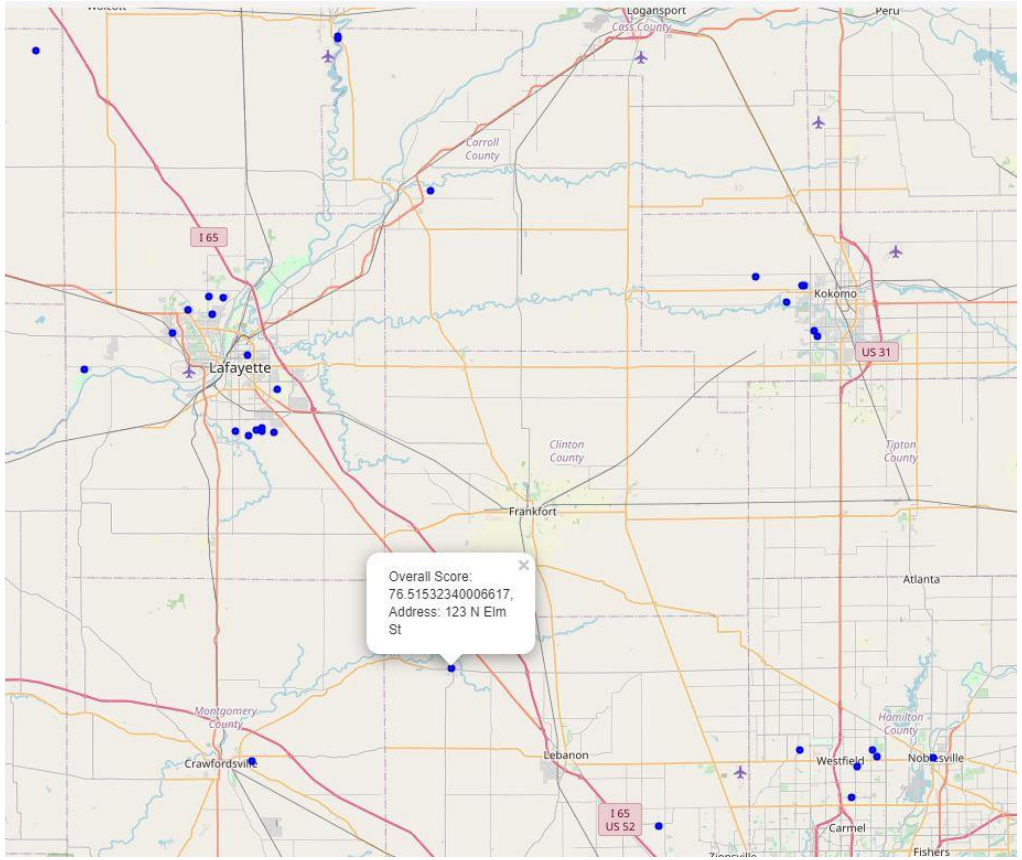


Data cleaning

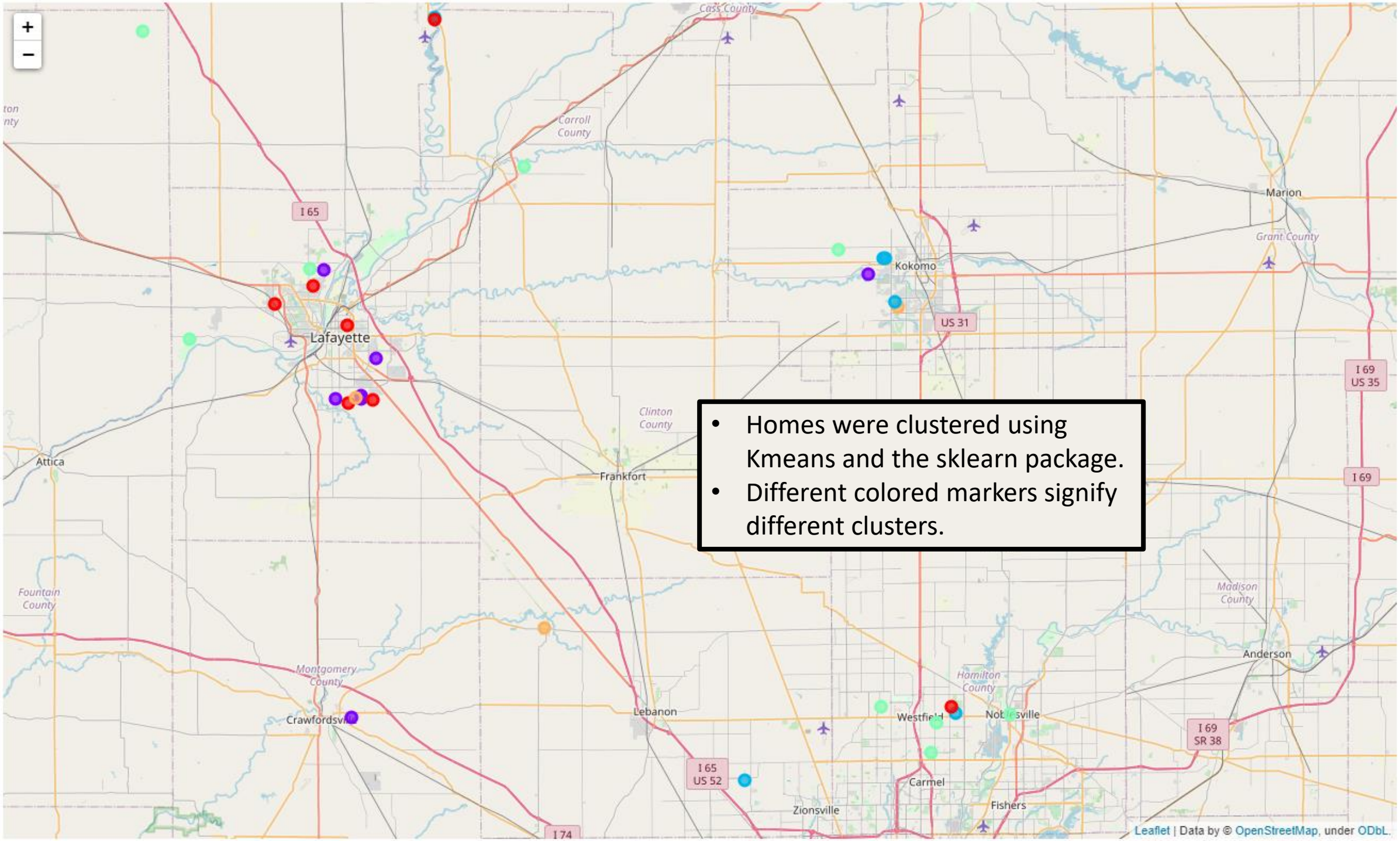
- Raw data set contained 219 homes
- Removing duplicates left 41 unique homes
- Restricting geographical range left 32 homes

	Address	Location	Price	Bed	Bath	Size
215	Thorntown, IN	123 N Elm St	$197,000	3bd	2ba	1,800 sqft
216	Ladoga, IN	5526 E 1200 S	$120,000	4bd	2ba	1,808 sqft
217	Crawfordsville, IN	307 Diamond Ln	$225,000	3bd	2ba	1,921 sqft
218	Thorntown, IN	123 N Elm St	$197,000	3bd	2ba	1,800 sqft
219	Ladoga, IN	5526 E 1200 S	$120,000	4bd	2ba	1,808 sqft

Last 5 entries of raw data set.



Visualization of final data set.



- Homes were clustered using Kmeans and the sklearn package.
- Different colored markers signify different clusters.

- Three of the five clusters are shown to the right.

- Variables were scaled using standard scaling with sklearn preprocessing before applying kmeans clustering.

- Visual analysis of the clusters shows that they are segregated by list price, with no other variables showing consistent trends between clusters.

```
df_model.loc[df_model['Cluster Labels'] == 0, df_model.columns[[1] + list(range(3, df_model.shape[1]))]]
```

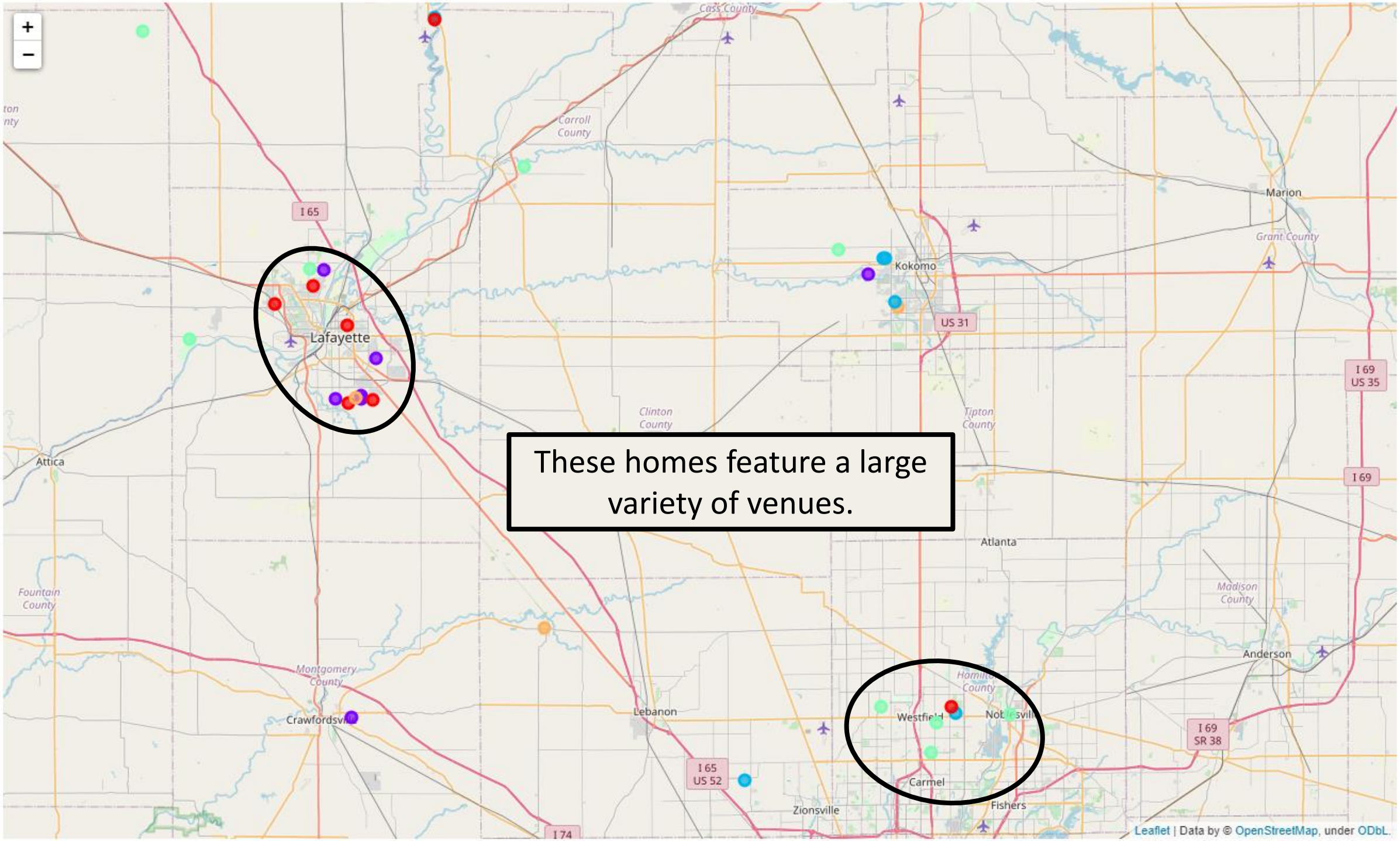
	index	Price	Bed	Bath	Size	City	latitude	longitude	overall_score	dist_to_costco	weighted_dist_score	unique_count	sqft_per_dollar
0	0	319000.0	4.0	4.0	3437.0	Lafayette	40.428314	-86.865967	63.280559	0.567368	-0.773607	0.834169	0.752358
5	5	319000.0	4.0	4.0	2432.0	West Lafayette	40.448970	-86.959076	66.937896	0.816603	-0.219054	0.951503	-0.572312
6	6	314900.0	4.0	3.0	2276.0	Lafayette	40.351856	-86.864540	88.151258	0.362079	-1.303002	0.423501	-0.738873
10	10	319900.0	3.0	3.0	2121.0	Lafayette	40.355357	-86.833444	89.023004	0.303741	-1.324218	-0.045833	-0.990100
11	11	315000.0	4.0	3.0	2250.0	West Lafayette	40.466883	-86.909771	70.945463	0.762410	-0.312718	0.892836	-0.774543
16	20	309900.0	4.0	3.0	3094.0	Monticello	40.727333	-86.754021	7.770966	1.320659	1.640396	-1.043170	0.420008
21	29	309999.0	3.0	2.0	2615.0	Noblesville	40.052323	-86.087628	86.174728	-1.605204	0.158783	1.186170	-0.231024

```
df_model.loc[df_model['Cluster Labels'] == 1, df_model.columns[[1] + list(range(3, df_model.shape[1]))]]
```

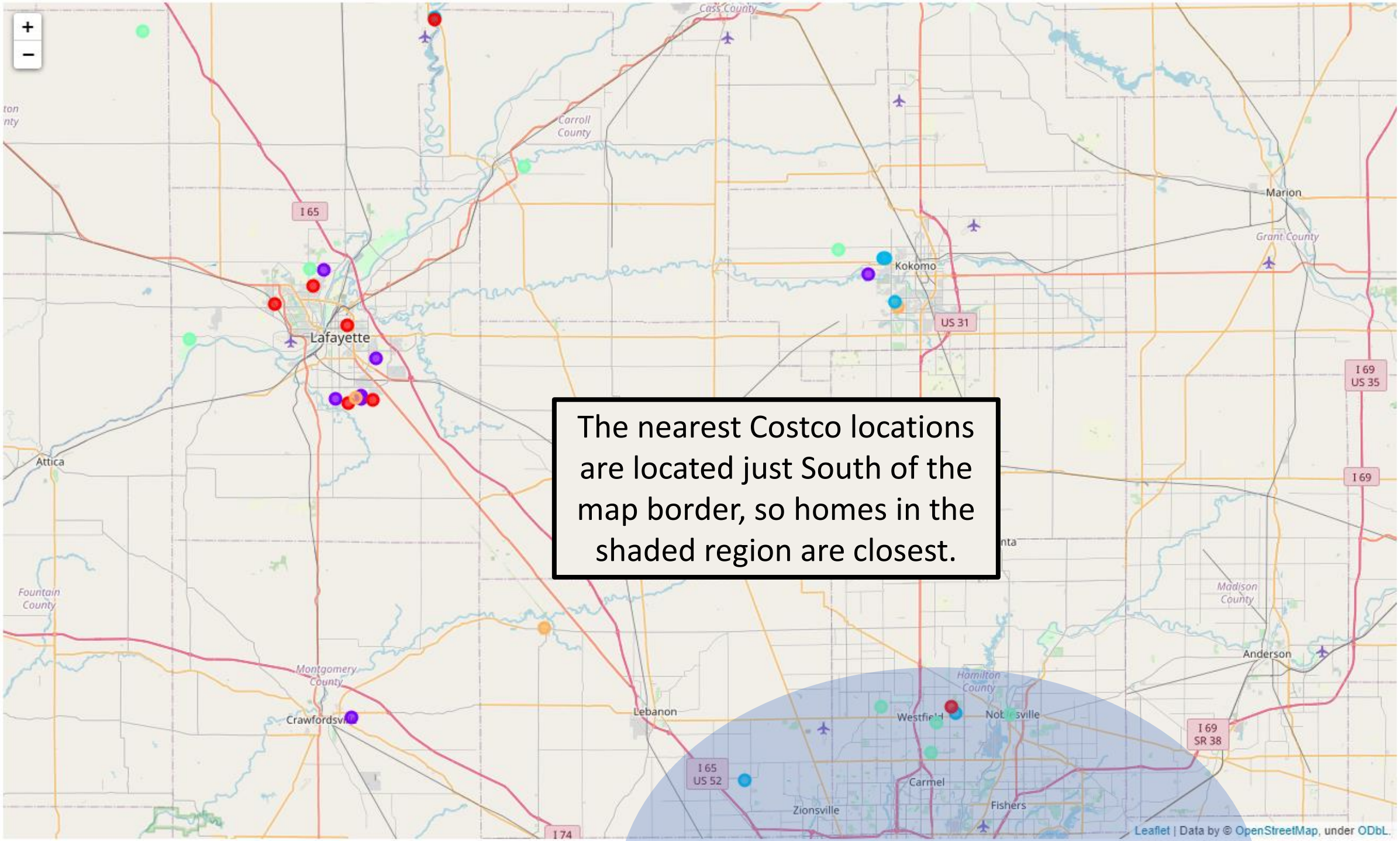
	index	Price	Bed	Bath	Size	City	latitude	longitude	overall_score	dist_to_costco	weighted_dist_score	unique_count	sqft_per_dollar
1	1	225000.0	3.0	3.0	2014.0	Lafayette	40.356407	-86.881485	78.647934	0.410850	-1.201686	0.540835	-0.014231
2	2	207000.0	3.0	3.0	1996.0	Lafayette	40.395855	-86.829083	71.702208	0.403172	-1.101664	0.188834	0.276480
3	3	214900.0	4.0	3.0	3592.0	Lafayette	40.355457	-86.847656	45.399266	0.334529	-1.333414	0.540835	3.250127
7	7	210000.0	3.0	3.0	3036.0	Lafayette	40.359222	-86.848447	55.714355	0.345970	-1.327620	0.540835	2.300876
9	9	215000.0	3.0	3.0	1920.0	West Lafayette	40.482346	-86.896234	60.850722	0.778551	-0.238191	0.834169	-0.023010
26	34	230000.0	3.0	2.0	2066.0	Kokomo	40.477871	-86.195257	44.216133	0.006560	0.849444	-0.397834	-0.000988
31	39	225000.0	3.0	2.0	1921.0	Crawfordsville	40.042246	-86.860882	66.560325	-0.216001	-0.318775	-0.573835	-0.188025

```
df_model.loc[df_model['Cluster Labels'] == 2, df_model.columns[[1] + list(range(3, df_model.shape[1]))]]
```

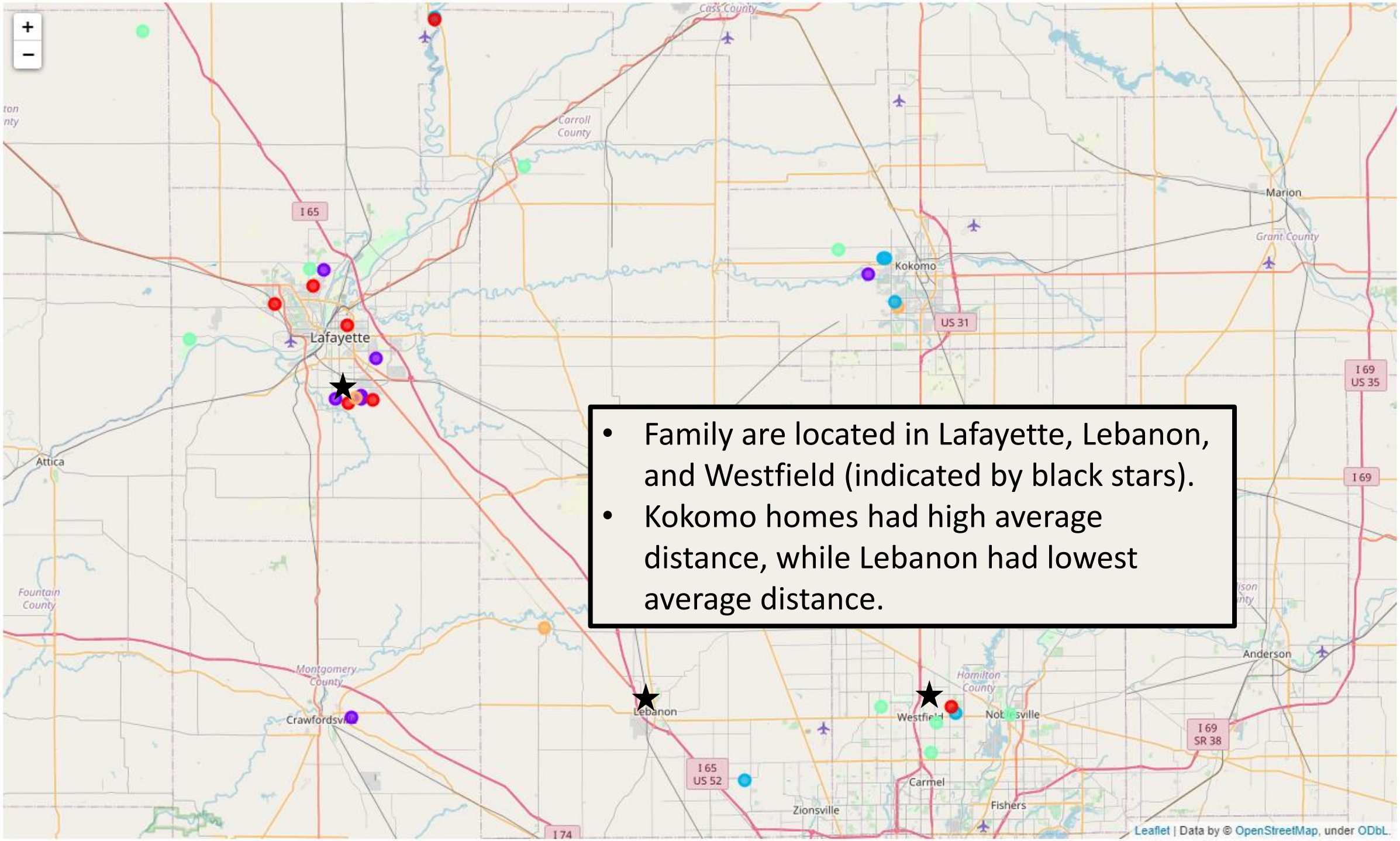
	index	Price	Bed	Bath	Size	City	latitude	longitude	overall_score	dist_to_costco	weighted_dist_score	unique_count	sqft_per_dollar
15	18	239900.0	3.0	2.0	1788.0	Monticello	40.729347	-86.753679	19.223152	1.327104	1.658298	-1.043170	-0.644094
20	28	250000.0	3.0	2.0	2214.0	Noblesville	40.046911	-86.082202	83.495584	-1.627783	0.196536	1.127503	-0.054222
23	31	265000.0	4.0	3.0	2514.0	Whitestown	39.980413	-86.353508	84.731763	-1.697207	-0.558419	-0.280501	0.211005
27	35	244900.0	3.0	2.0	1836.0	Kokomo	40.493983	-86.172819	47.746689	0.071931	1.037133	-0.339168	-0.625664
28	36	254900.0	4.0	3.0	3068.0	Kokomo	40.450988	-86.159844	30.148855	-0.088700	0.859941	-0.515168	1.282903
30	38	265900.0	3.0	2.0	1890.0	Kokomo	40.494111	-86.175175	49.715722	0.071927	1.027635	-0.339168	-0.789227



These homes feature a large variety of venues.



The nearest Costco locations are located just South of the map border, so homes in the shaded region are closest.



- Family are located in Lafayette, Lebanon, and Westfield (indicated by black stars).
- Kokomo homes had high average distance, while Lebanon had lowest average distance.

Key Results

- Homes in the Westfield/Carmel area had the highest average rank due to proximity to Costco, large venue diversity, and moderate average purchasing power and proximity to family.
- Lafayette/West Lafayette homes also scored well due to high venue diversity and purchasing power, with moderate proximity to family and high distance to Costco.
- Houses in rural areas and Kokomo scored poorly due to low venue diversity, high average distance from family and Costco, and moderate to low purchasing power.

Conclusion

- Algorithm narrowed choices according to criteria. However, selection could be improved.
  - Purchasing power should include lot size, as rural houses are unfairly punished by large lot size, which increases list price, but is not accounted for in current calculation of purchasing power (square footage of home / list price)
  - Other criteria important to potential buyers could be added (average school rating) to improve ranking algorithm.
  - Criteria suit target audience (me) but lacks generalizability in current form.