Portland Neighborhood Search

Coursera IBM Data Science Capstone Project

Maribeth Todd

May 2020

Business Problem

- Locate the right neighborhood for purchasing a house in the Portland, OR metropolitan region
- Develop a systematic approach using public data that can be tailored to different preferences
- Analysis should be of interest to
  - anyone looking to move to a new neighborhood
  - real estate professionals helping clients choose a neighborhood

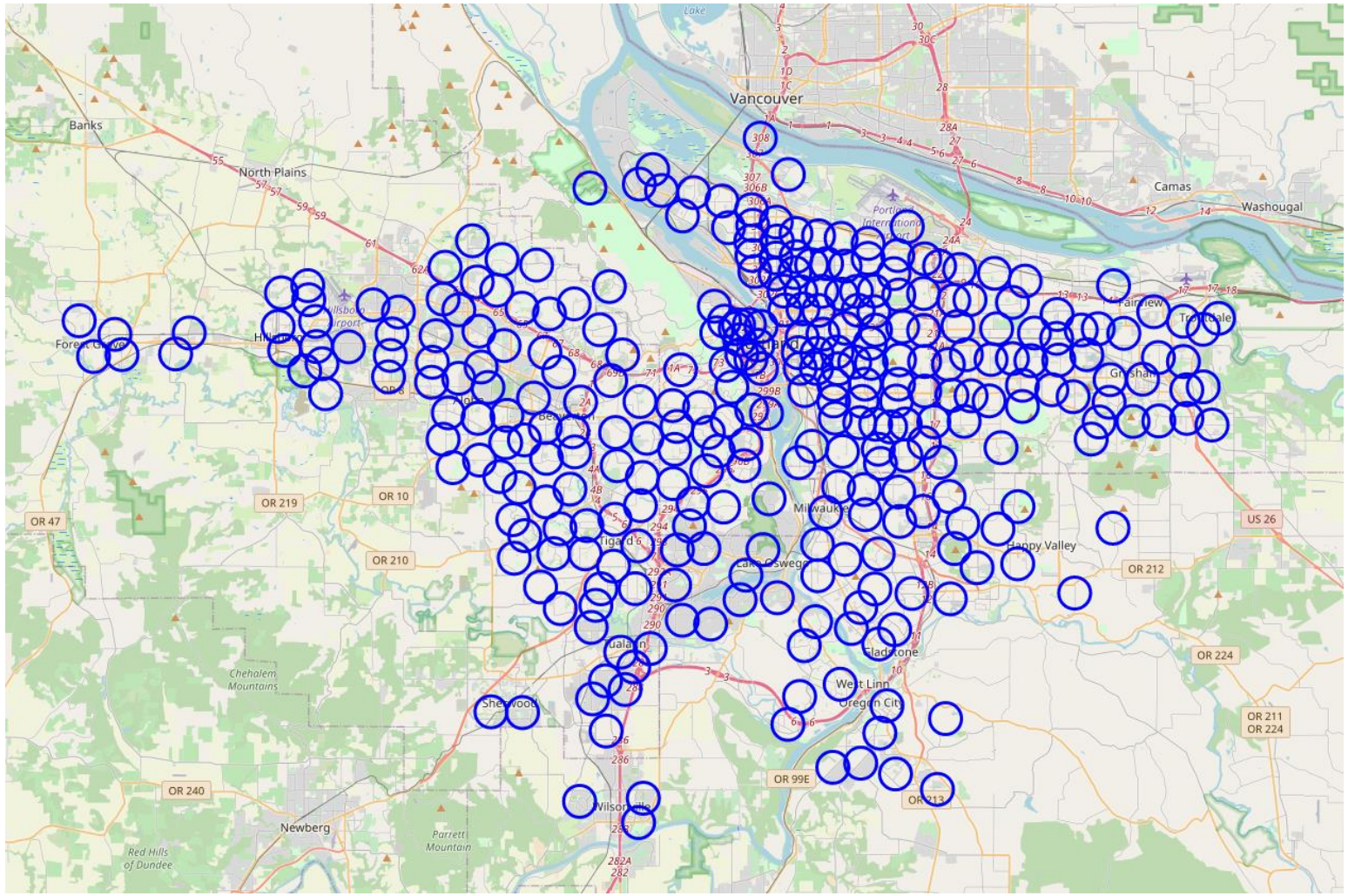
Data sources

- Mean Centers of Population for Census tracts
  - Define the neighborhoods for analysis
  - Mean population centers better represent location of population within a tract than geographic centers
- Foursquare API
  - Businesses and venues available in each neighborhood
- Census American Community Survey API
  - Demographics, prices, and other neighborhood information
- Mapquest Geocoding API
  - Attach zip codes to neighborhoods for linking to real estate listings

Methodology

Neighborhoods are based on the center of population for each Census tract

Radius of 800 meters, about a half mile, is used for the Foursquare venue search



Collect and prepare the data for cluster analysis

- Get Census data on population density and median house values
- For each neighborhood, summarize the number of Foursquare venues within a half mile for each of the following categories:
  - Grocery stores, hardware stores, restaurants, libraries
  - Libraries may be a limiting factor since 2/3 of the region's neighborhoods do not have easy access to one

	Number of neighborhoods with venues			
Number of venues	library	hardware	grocery	restaurant
0	206	153	115	0
1	78	66	71	32
2	12	45	58	13
3	6	14	32	8
4	6	14	12	10
5	1	5	9	7
6	1	4	3	7
7	1	3	5	9
8	0	1	2	7
9	0	1	3	6
10 or more	0	5	1	212

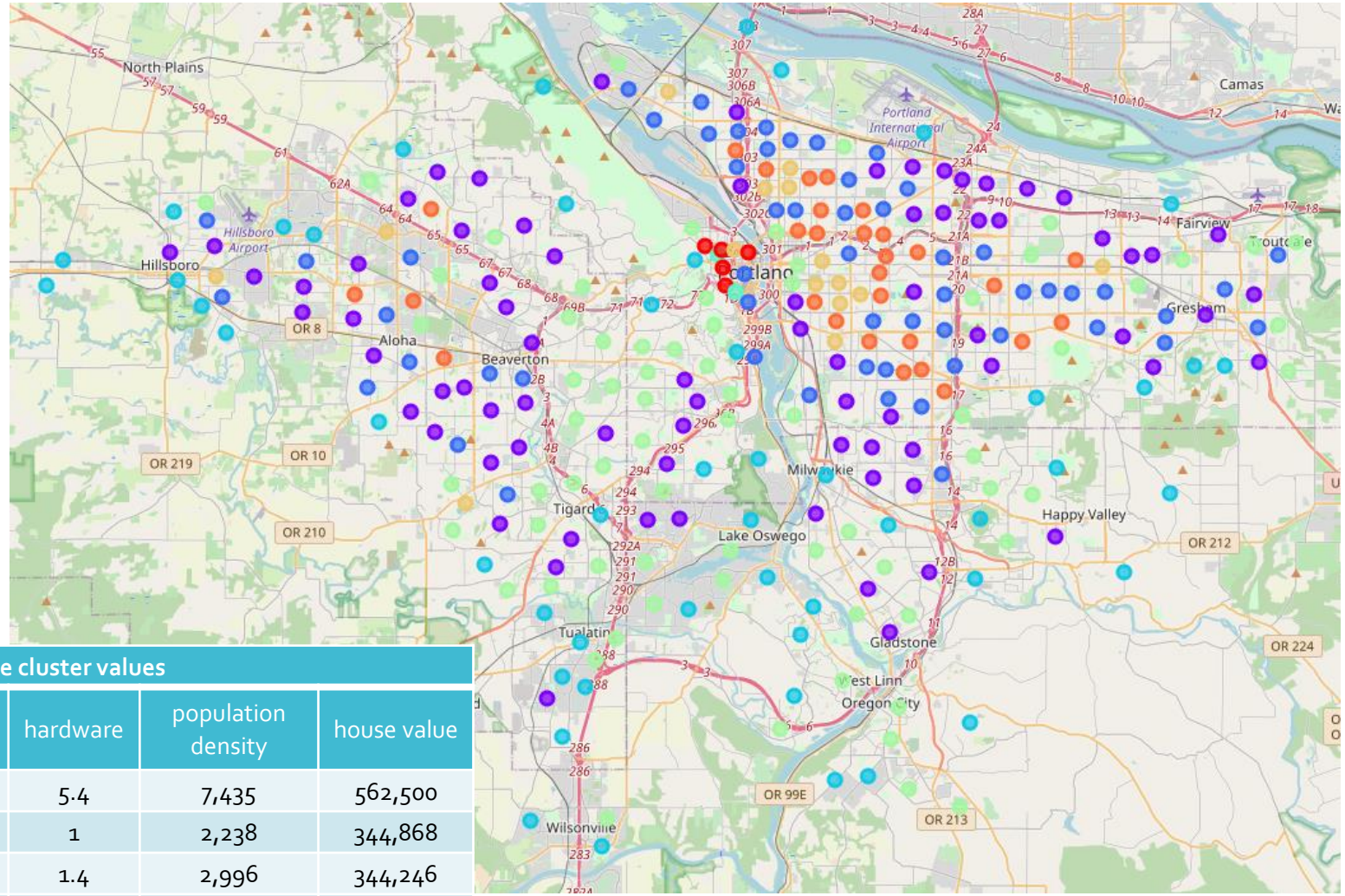
Cluster analysis inputs and parameters

- Use the four venue types and population density, house values will be used for filtering results after the clustering step
- K-means clustering using scaled data
- Evaluation metrics indicate model with 8 or 9 clusters may perform best, so set k=8
  - Silhouette and Calinksi-Harabaz scores: higher = better
  - Davies-Bouldin score: lower = better

k-means cluster metrics	6 clusters	7 clusters	8 clusters	9 clusters	10 clusters
Average silhouette score	0.534	0.536	0.561	0.557	0.548
Davies-Bouldin score	0.522	0.511	0.490	0.398	0.412
Calinski-Harabasz score	868	1054	1213	1363	1622

Cluster results

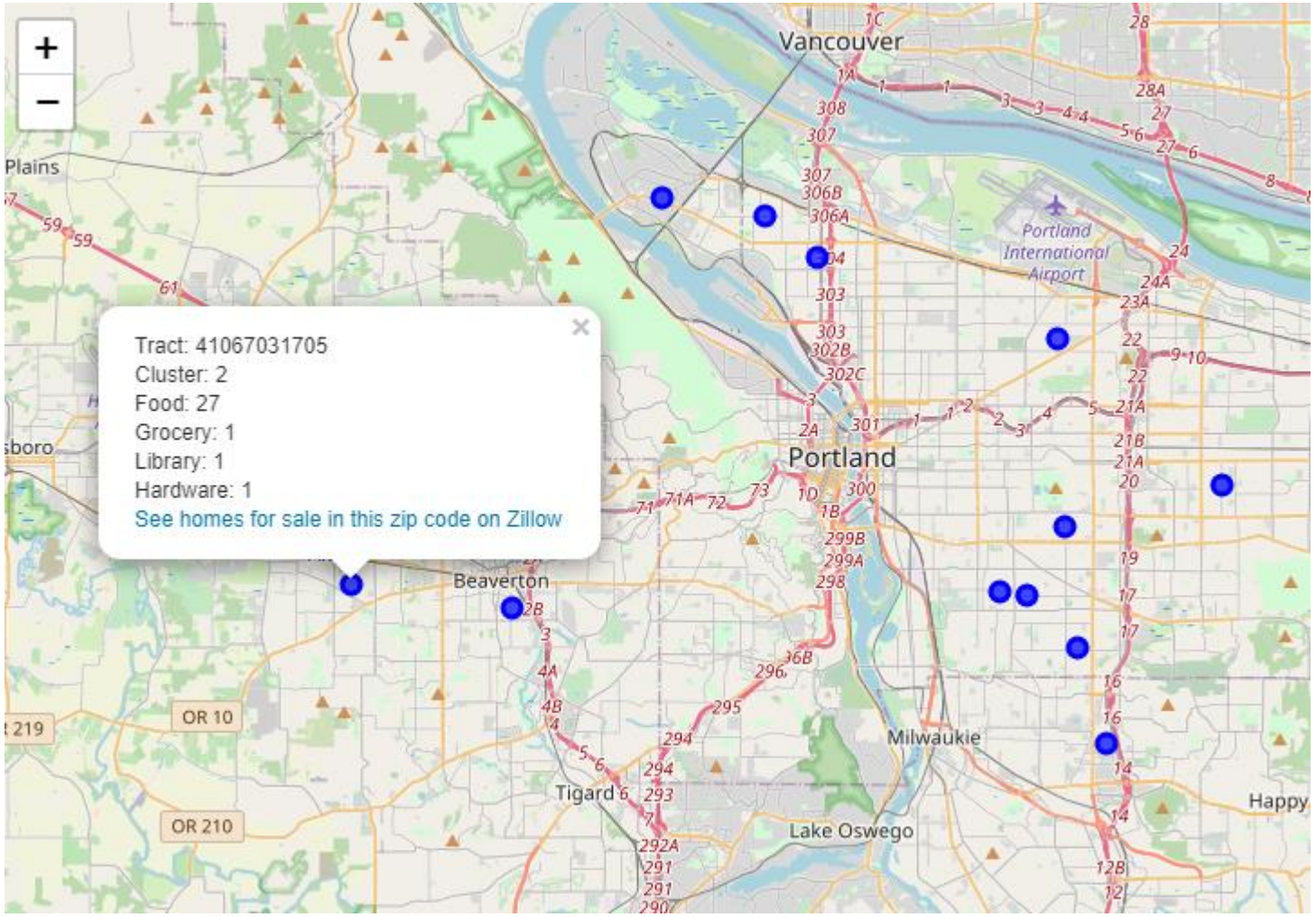
- Most neighborhoods in clusters 1, 2, 3, 4, 7
- Clusters 0 and 4 represent a few neighborhoods in the central city with very high population and service densities
- Cluster 2 (dark blue points) looks promising for further analysis with low house prices, moderate density, and most neighborhoods in the cluster with access to at least one of each of the venue types



		Average cluster values					
cluster	neighborhood count	food	grocery	library	hardware	population density	house value
0	5	30	6.4	3.4	5.4	7,435	562,500
1	78	15.9	1.1	0.4	1	2,238	344,868
2	56	21.8	2.2	0.6	1.4	2,996	344,246
3	54	10.4	0.6	0.3	0.7	628	397,498
4	2	30	6	2.5	4	11,852	353,250
5	72	13.9	1	0.2	1.4	1,508	414,321
6	15	29.1	4.5	1.2	2.2	4,763	397,587
7	29	23.6	2.2	0.9	1.7	3,755	412,283

Further analysis

- Filter the cluster 2 neighborhoods to focus on house values less than $400,000 and access to at least one library within a half mile
- This results in 12 neighborhoods for further exploration
- The map in the Jupyter Notebook provides links to homes currently for sale in each neighborhood
- Choosing the right neighborhood depends on other factors such as the types of houses for sale and commute to work or school
- If none of these neighborhoods provide a good fit, the analysis can be modified by looking at another cluster, or using different clustering and filtering criteria



Conclusion

- This analysis is based on easy walking access to grocery stores, hardware stores, libraries, and restaurants within a half mile but the approach can be easily customized for different preferences.
- Foursquare data indicate that restaurants and other food venues like coffee shops are readily available in most neighborhoods in the Portland region, and grocery stores are fairly well distributed as well. Hardware stores are less common, and easy access to a library is a much more limiting factor in choosing a neighborhood.
- The ultimate decision of which neighborhood to purchase a house in must include additional factors but this is a good way to start narrowing the search.
- This type of systematic neighborhood evaluation could be useful for other individuals looking for a new house, or real estate professionals assisting clients with their search for a place to live.