IBM Data Science Capstone

Finding Similar Neighborhoods in Philadelphia

Problem

- How can I find a neighborhood in another city that is similar to a given neighborhood somewhere else?
  - If I have a successful business in Wilmington, DE and I want to expand to the much larger market of Philadelphia, PA, where should I look?
  - I like my current neighborhood, and I think it was a factor in the success of the business. Why not look for similar neighborhoods in Philadelphia?

Solution

- Use the Foursquare venue data to build a feature set around each neighborhood of Philadelphia.
- Build a ‘profile’ of the neighborhoods based on the frequency of the types of venues in the neighborhoods.
- Use the k-means algorithm to build a model that groups similar neighborhoods.
- Build a similar profile for the neighborhoods you want to emulate, and use the model to predict the cluster that is most similar.

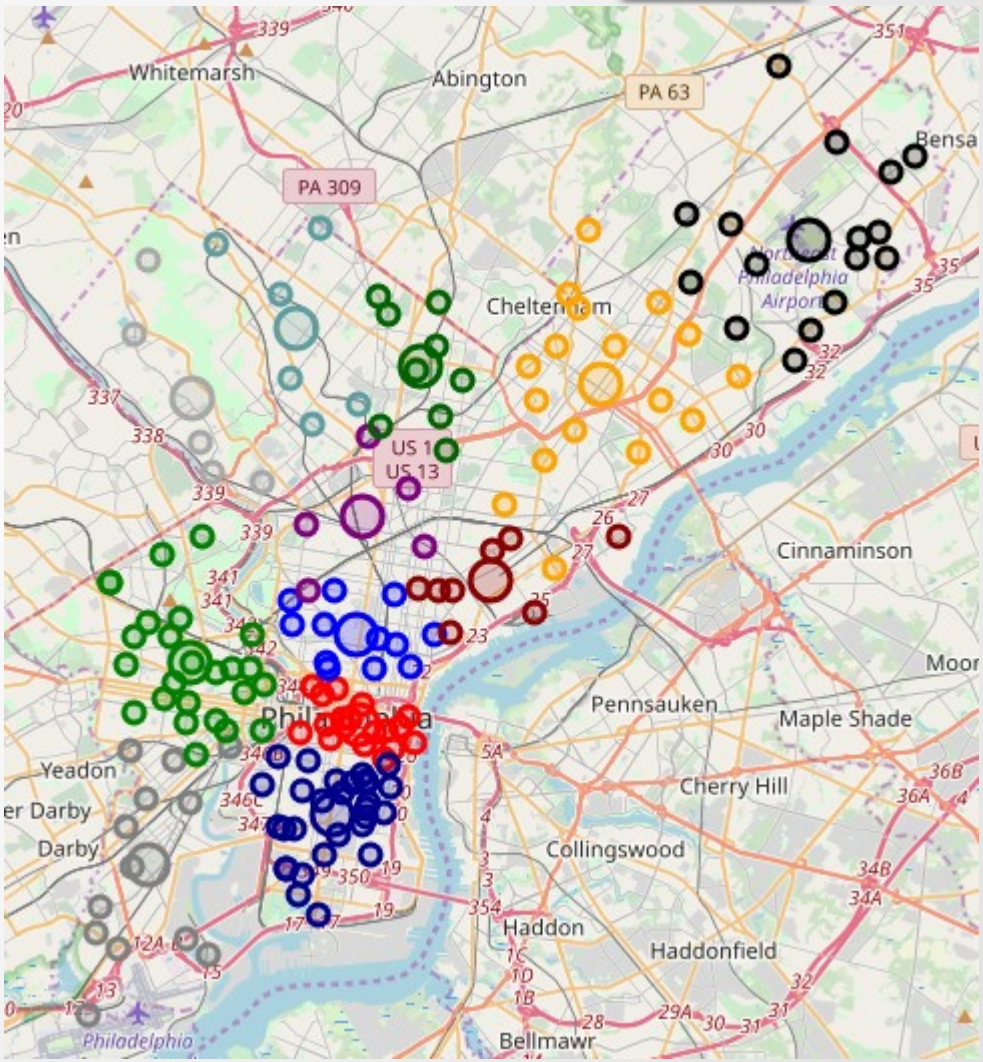
Feature Data

- Below is an excerpt of the feature data for the neighborhoods
- The venue type columns have values that indicate the relative frequency of the venue type in the neighborhood.

ID	Neighborhood	...	Theater	Courthouse	College Library	Parking	Church	Doctor's Office	Art Gallery	...
1	Avenue of the Arts		0.01123	0.01123	0.01123	0.01123	0.01123	0.01123	0.00000	
2	Callowhill		0.00000	0.02222	0.00000	0.07777	0.01111	0.00000	0.07777	
3	Chinatown		0.00000	0.00000	0.01063	0.04255	0.01063	0.00000	0.01063	
4	Elfreth's Alley		0.00000	0.00000	0.00000	0.01136	0.00000	0.01136	0.18181	
5	French Quarter		0.00000	0.00000	0.00000	0.02173	0.00000	0.01087	0.02173	
6	Logan Square		0.01098	0.00000	0.00000	0.01098	0.00000	0.00000	0.06593	
7	Naval Square		0.02083	0.02083	0.00000	0.02083	0.01041	0.00000	0.03125	
8	Jewelers' Row		0.00000	0.00000	0.00000	0.00000	0.00000	0.01087	0.00000	
9	Market East		0.00000	0.00000	0.00000	0.01162	0.00000	0.00000	0.03488	
10	Old City		0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.05555	
...										

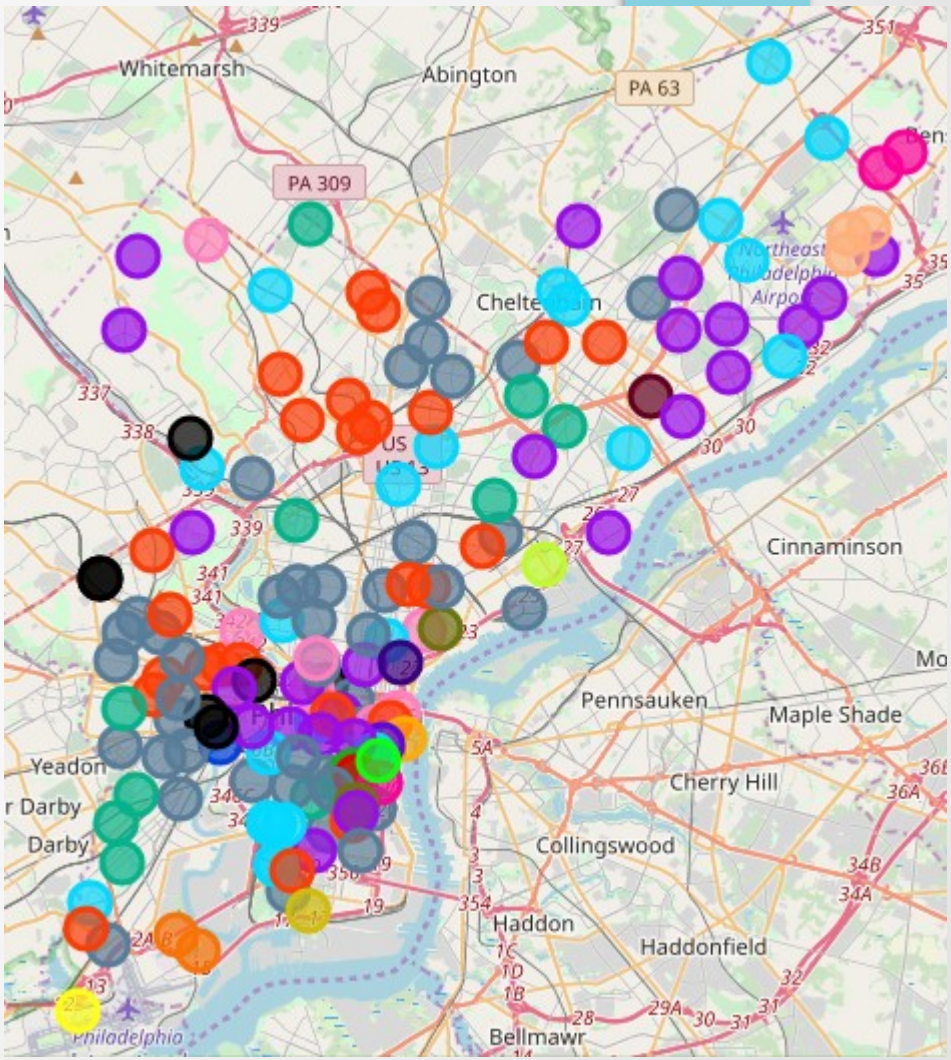
Unclustered Neighborhoods

- Neighborhoods are color coded by planning zone.
- They are grouped geographically.
- This is how one normally thinks of the city and its neighborhoods.



Clustered by Similarity

- The result of the k-means model (k=22), neighborhoods are color coded by cluster.
- Neighborhoods of the same color are in the same cluster, and are similar.



Locate the Desired Neighborhoods

- Philippine BBQ
  - 1700 Kirkwood Hwy, Wilmington, DE
  - Cluster #7 Violet
- De La Coeur (French Cafe)
  - 1836 Lovering Ave, Wilmington, DE
  - Cluster #17 Red orange
- Wang's Market
  - 276 E Main St, Newark, DE
  - Cluster #6 Neon Blue