Capstone Project

Finding a Philadelphia Neighborhood Most Similar to Almagro, Madrid

Introduction

The project objective is to find a neighborhood in Philadelphia, PA, USA, that is most similar to Almagro, Madrid, but the program is applicable for finding a neighborhood in any city that is similar to some other target neighborhood. The algorithm could be used by people who are planning to move to a new city and want to find a neighborhood that is most similar to a known neighborhood that they find desirable. The algorithm also incorporates real estate price data so that the user can compare similarity between neighborhoods and median housing prices to find a most optimal match for his/her particular situation.

Data Sources

The following data sources will be utilized in the algorithm:

- GPS data from Nominatim: www.openstreetmap.org
- The Foursquare Places API for exploring neighborhood venues https://developer.foursquare.com/
- A list of Philadelphia neighborhoods and housing price statistics published by Philadelphia magazine at https://www.phillymag.com/property/house-prices-philadelphia-suburbs/#philly

Methodology

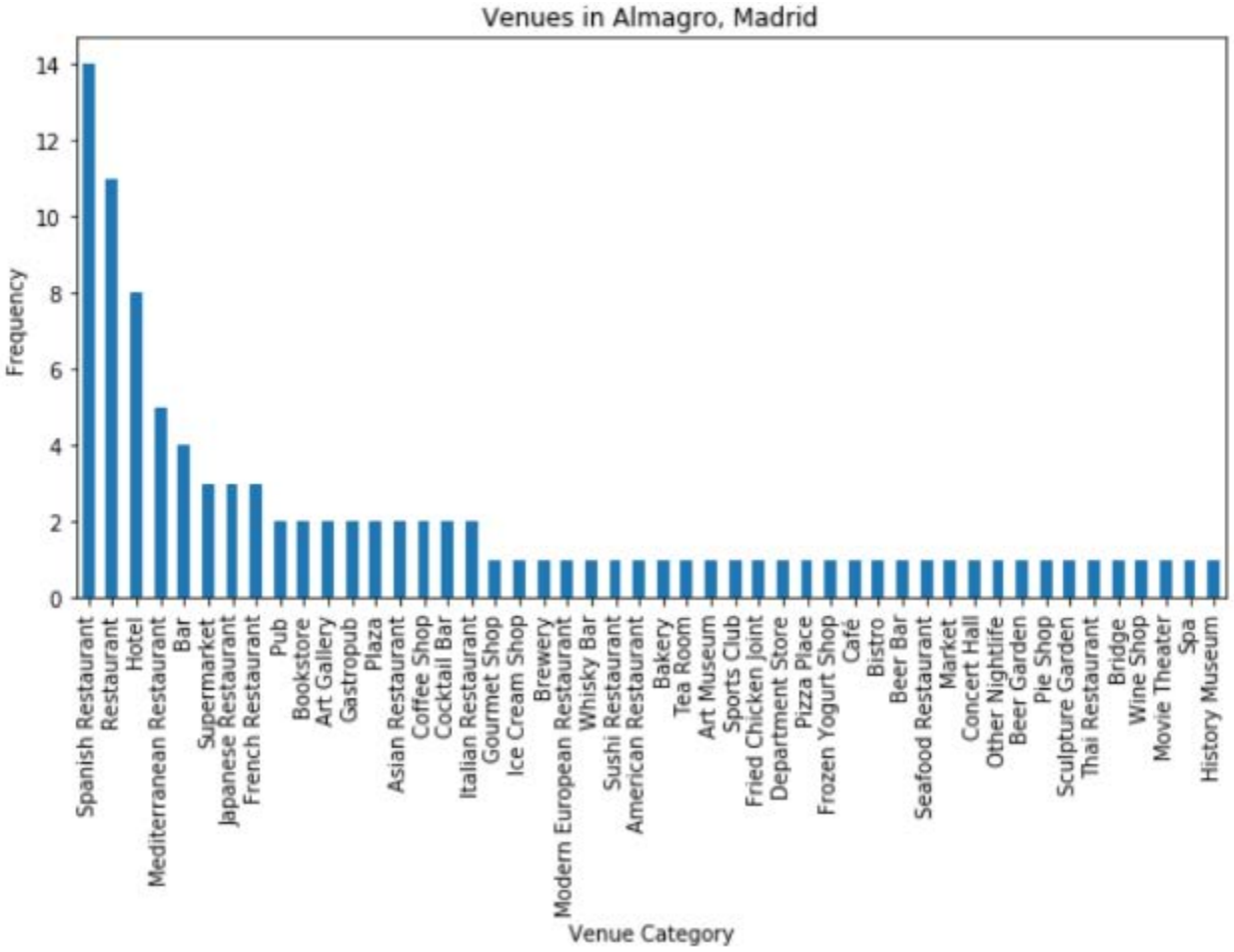
The target neighborhood, Almagro, Madrid will be characterized through an analysis of nearby venues utilizing the Foursquare Places API. Philadelphia neighborhoods will then be characterized in the same manner, and k-means clustering will be utilized to group Philadelphia neighborhoods by similarity in venue types. Almagro, Madrid will be included in the clustering analysis to identify the cluster of Philadelphia neighborhoods that is most similar to it.

To further identify the Philadelphia neighborhoods that are most similar to Almagro, exact Euclidian distance from Almagro will be calculated in the multidimensional venue space. The neighborhoods will then be plotted in a scatter plot against median housing

prices to provide an assessment of affordability along with similarity to Almagro. The ideal neighborhood will be similar to Almagro, but also affordable for the user of the algorithm.

Analysis, Results, and Discussion

The Foursquare Places API was utilized to characterize the target neighborhood, Almagro, Madrid by its categories and quantities of venues. As can be seen in the following bar chart, the Almagro neighborhood has a high concentration and wide variety of restaurants. Spanish and Mediterranean restaurants are most common. There are also a good number of bars/pubs and hotels, and some art galleries.

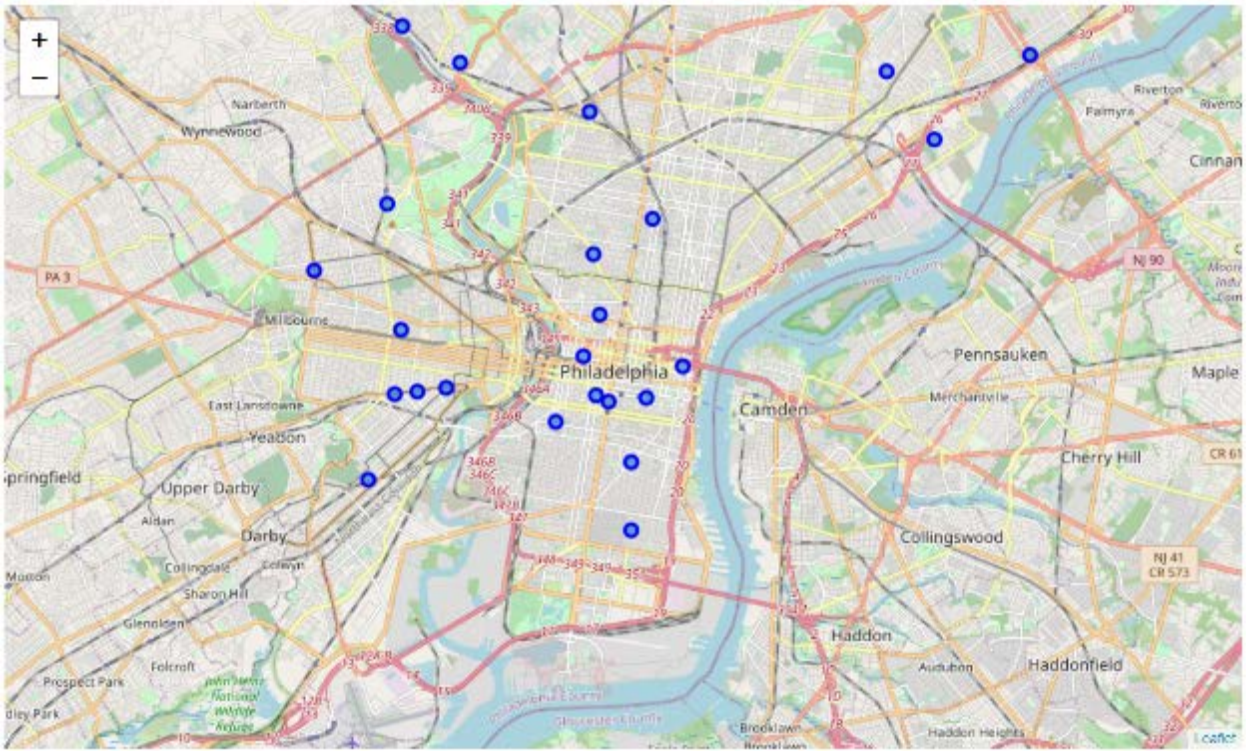


A list of Philadelphia neighborhoods and housing price data was scraped from an article published by Philadelphia magazine at https://www.phillymag.com/property/house-prices-philadelphia-suburbs/#philly. After reformatting the column data, the

neighborhood names were fed into the Nominatim database to collect GPS coordinates. The resulting dataframe is shown below.

	Neighborhood	Zip_Code	2017 Median Price	1-Yr. Change	3-Yr. Change	10-Yr. Change	Latitude	Longitude
0	Center City West: Broad to 16th, Callowhill to	19102	$580,500	-18%	37%	78%	39.946212	-75.165018
1	Rittenhouse	19103	$530,000	9%	12%	20%	39.955870	-75.171830
2	University City/West Philadelphia	19104	$218,375	-3%	28%	76%	39.948252	-75.209377
3	Old City/Society Hill	19106	$404,625	8%	4%	-1%	39.953578	-75.144727
4	Center City East: Broad to 6th, Callowhill to ...	19107	$335,000	15%	20%	-7%	39.947072	-75.154648
5	Fox Chase	19111	$170,000	10%	17%	0%	40.080375	-75.080391
6	Torresdale South	19114	$189,950	9%	19%	-6%	40.060628	-75.012622
7	Bustleton	19115	$235,000	11%	21%	0%	40.085084	-75.049773
8	Somerton	19116	$233,000	5%	8%	1%	40.116368	-75.018372
9	Chestnut Hill	19118	$451,000	-5%	-6%	11%	40.067015	-75.187339
10	Mount Airy	19119	$257,200	7%	14%	9%	40.052291	-75.188578
11	Olney	19120	$68,000	14%	42%	-5%	44.965675	1.827779
12	Strawberry Mansion/Brewerytown	19121	$170,000	45%	143%	138%	39.977210	-75.168987
13	Temple/Olde Kensington	19122	$305,000	36%	103%	118%	39.984489	-75.153013
14	Northern Liberties/East Spring Garden	19123	$418,500	-5%	28%	28%	39.968016	-75.223406
15	Juniata/Frankford	19124	$85,000	12%	48%	-9%	40.015194	-75.089053
16	Fishtown/Kensington	19125	$309,150	5%	44%	107%	44.121661	9.836675
17	Oak Lane	19126	$140,450	9%	27%	17%	44.111037	9.838797
18	Manayunk	19127	$235,000	5%	7%	-6%	40.024769	-75.221294
19	Roxborough/Andorra	19128	$239,450	6%	12%	3%	40.017049	-75.205852
20	East Falls	19129	$224,900	5%	8%	3%	40.006745	-75.170010
21	Fairmount	19130	$350,000	5%	16%	14%	39.964489	-75.167457
22	Wynnefield/Carroll Park	19131	$99,900	22%	32%	11%	39.987532	-75.225379
23	North Philadelphia West	19132	$27,500	26%	38%	-21%	44.086751	9.817743
24	North Philadelphia East	19133	$41,000	44%	39%	-9%	44.096731	9.775781
25	Richmond	19134	$84,450	39%	107%	43%	44.123910	9.783316
26	Tacony	19135	$107,000	13%	34%	-11%	40.018730	-75.049864
27	Holmesburg	19136	$139,000	9%	24%	1%	40.041957	-75.026135
28	Bridesburg	19137	$146,450	21%	36%	5%	40.031018	-75.075835
29	East Germantown	19138	$109,450	29%	37%	23%	44.111484	9.848488
30	West Market	19139	$60,000	36%	71%	9%	39.961289	-75.221523
31	Hunting Park	19140	$38,050	22%	46%	-24%	45.445883	1.578142
32	Logan	19141	$99,450	12%	28%	21%	40.585544	-3.121644
33	Elmwood	19142	$48,000	9%	45%	-31%	39.930088	-75.230725
34	Kingsessing	19143	$78,500	11%	32%	12%	39.948549	-75.217172
35	Germantown	19144	$119,400	20%	26%	15%	40.036057	-75.167824
36	South Philly: S. of Tasker, W. of Broad	19145	$209,000	18%	74%	84%	39.947720	-75.168289
37	Graduate Hospital/Point Breeze	19146	$365,000	9%	18%	78%	39.942056	-75.179274
38	Bella Vista/Queen Village/Passyunk Square North	19147	$439,500	19%	24%	37%	39.933811	-75.158858
39	South Philly: S. of Tasker, E. of Broad	19148	$216,000	10%	36%	51%	39.915435	-75.159808
40	Oxford Circle/Mayfair	19149	$146,000	14%	33%	10%	58.446465	17.927804
41	Cedarbrook	19150	$157,000	9%	15%	7%	45.229482	1.861242
42	Overbrook	19151	$132,500	9%	25%	10%	39.973653	-75.245311
43	Rhawnhurst	19152	$205,000	14%	28%	5%	40.051500	-75.058291
44	Eastwick	19153	$133,500	21%	27%	-1%	40.667254	-2.969596
45	Torresdale North	19154	$189,900	8%	18%	1%	48.926220	29.920578

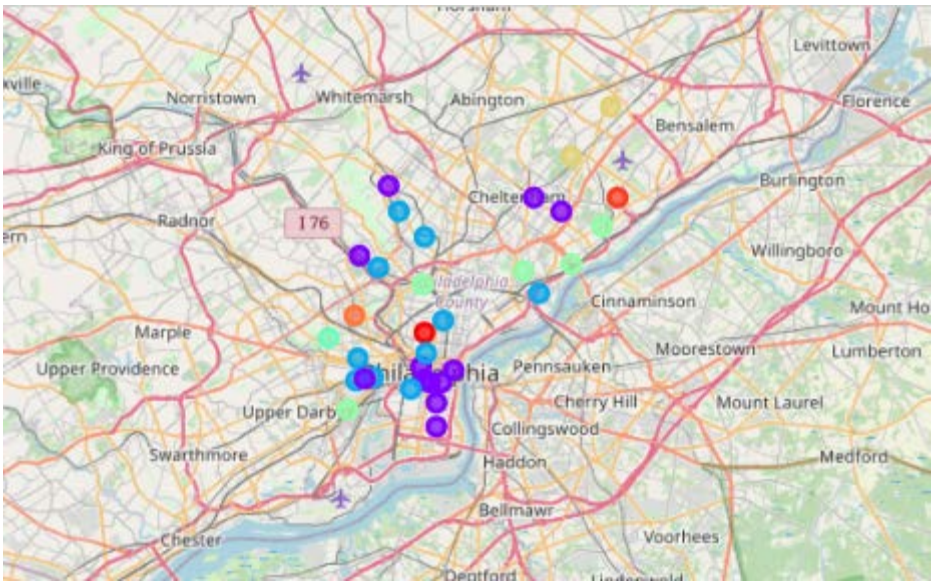
The GPS data were used to create a map of the Philadelphia neighborhoods to visualize their locations with respect to each other



The GPS data were then run through the Foursquare places API to generate a list of venues in each neighborhood. The resulting json file was converted to a dataframe for subsequent one-hot encoding and analysis by k-means clustering. Prior to clustering, the venue data for Almagro, Madrid was added to the Philadelphia dataframe to identify the cluster into which it would fall. This would identify the cluster of most similar neighborhoods to Almagro.

K-means clustering showed that Almagro fell into cluster number 1, along with Center City West, Rittenhouse Square, Old City/Society Hill, and nine other Philadelphia neighborhoods, which are visualized by the purple markers in the following map. These neighborhoods had a high number and wide variety of restaurants, as well as hotels.

Two other clusters were identified in the Philadelphia neighborhoods. More casual venues like pizza resaurants and sandwich places were more popular in Cluster 4, and Cluster 8 included neighborhoods with fewer sit-down restaurants and more fast food venues.

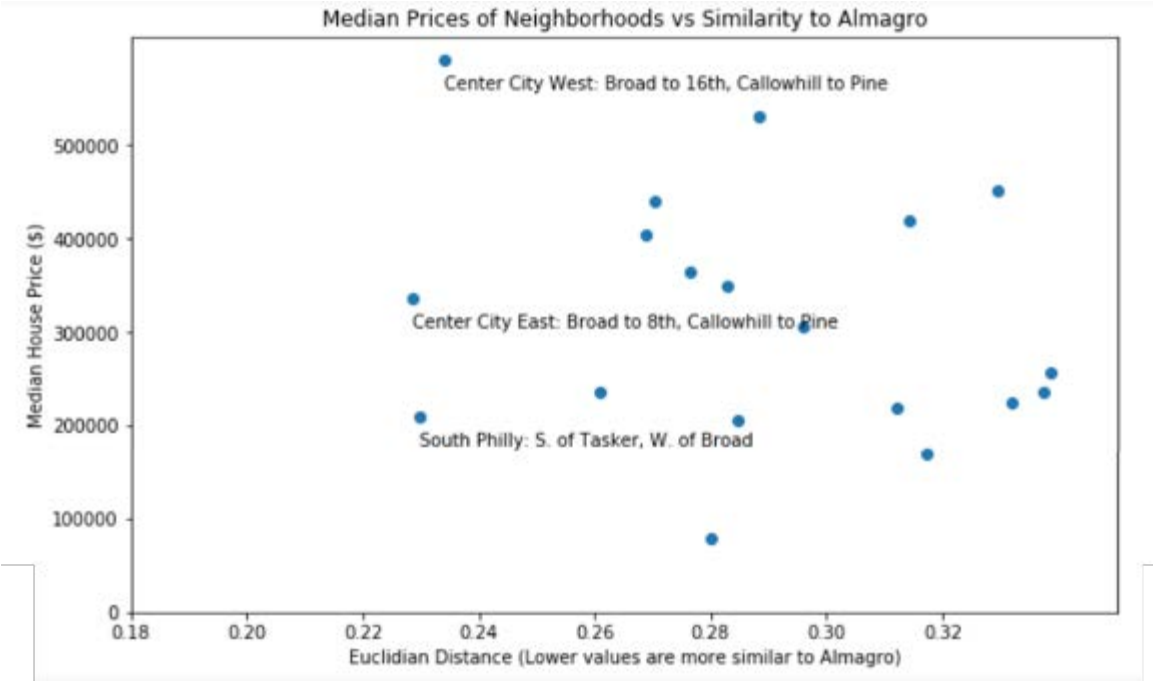


	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Almagro	Spanish Restaurant	Restaurant	Hotel	Mediterranean Restaurant	Bar	Supermarket	French Restaurant	Japanese Restaurant	Gastropub	Bookstore
1	Bella Vista/Queen Village/Passyunk Square North	Sandwich Place	Mexican Restaurant	Vietnamese Restaurant	Bakery	Coffee Shop	Gym	Bookstore	Restaurant	French Restaurant	Food Truck
2	Bridesburg	Pharmacy	Pizza Place	Bus Station	Food & Drink Shop	Gas Station	Bar	Food Truck	Restaurant	Sports Bar	American Restaurant
3	Bustleton	Indian Restaurant	Cosmetics Shop	Grocery Store	Pizza Place	Pharmacy	Salon / Barbershop	Diner	Chinese Restaurant	Skating Rink	Mediterranean Restaurant
4	Center City East: Broad to 8th, Callowhill to ...	Historic Site	Coffee Shop	American Restaurant	Sandwich Place	Bar	Japanese Restaurant	Restaurant	Italian Restaurant	Jewelry Store	Breakfast Spot
5	Center City West: Broad to 16th, Callowhill to...	Coffee Shop	Pizza Place	Vegetarian / Vegan Restaurant	Mexican Restaurant	Bakery	Mediterranean Restaurant	Yoga Studio	Italian Restaurant	Spa	Hotel
6	Chestnut Hill	Chinese Restaurant	Bakery	Hotel	Japanese Restaurant	Smoke Shop	Supermarket	Gas Station	Church	Paper / Office Supplies Store	Trail
7	East Falls	Clothing Store	Dessert Shop	Thrift / Vintage Store	Donut Shop	Fast Food Restaurant	Discount Store	Pharmacy	Restaurant	Bus Station	Ice Cream Shop
8	East Germantown	Hotel	Athletics & Sports	Fast Food Restaurant	Café	Smoke Shop	Electronics Store	Grocery Store	Multiplex	Pizza Place	Department Store
9	Eastwick	Hotel	Zoo Exhibit	Donut Shop	Flea Market	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Fabric Shop	Event Space	Ethiopian Restaurant

Euclidian distance from Almagro was then calculated in the multidimensional venue space to identify the neighborhoods that were most similar to Almagro. Finally, the housing price statistics were incorporated in the analysis to identify neighborhoods that might be more affordable for the user, yet still relatively similar to the target neighborhood.

When applied for finding a neighborhood in the city of Philadelphia, Pennsylvania, USA that is most similar to Almagro in Madrid, Spain, three neighborhoods are identified as promising candidates: (1) Center City East: Broad to 8th, Callowhill to Pine, (2) South Philly: S. of Tasker, W. of Broad, and (3) Center City West: Broad to 16th, Callowhill to

Pine. Of these neighborhoods, South Philly is clearly the most affordable, with a median house price of $209,000. These results are visualized in the following scatter plot of median housing prices versus similarity to Almagro.



Conclusions

An effective algorithm has been developed for identifying neighborhoods in a selected city that are most similar to some desirable neighborhood in another city. The algorithm utilizes k-means clustering to find groups of similar and dissimilar neighborhoods, and then identifies the most similar neighborhoods by calculating Euclidian distance from the target neighborhood in the multidimensional space. Finally, the housing price statistics are incorporated in the analysis to identify neighborhoods that might be more affordable for the user, yet still relatively similar to the target neighborhood.

When applied for finding a neighborhood in the city of Philadelphia, Pennsylvania, USA that is most similar to Almagro in Madrid, Spain, three neighborhoods are identified as promising candidates:

- Center City East: Broad to 8th, Callowhill to Pine
- South Philly: S. of Tasker, W. of Broad
- Center City West: Broad to 16th, Callowhill to Pine.

Of these neighborhoods, South Philly is clearly the most affordable, with a median house price of $209,000.