CAPSTONE PROJECT

Battle of the neighborhoods

2021

2021-08-09

Applied Data Science Capstone

Authored by: Emile Strasheim

Table of contents

Introduction

This section contains the description of the background of the focus of this project, the business problem, as well as a description of who might be interested in the results of this project.

Background

As the world is becoming more connected and industries become integrated, more and more people relocate from one country to another in pursuit of new career opportunities. These relocations cause a disruption in the lives of these people and their families as they are required to familiarise themselves with a new country, new people, new culture, new language (perhaps) and new environment.

One of the biggest struggles of newly relocating families is finding a neighborhood and a place to live in that is compatible to their needs, and perhaps even similar to what they are used to.

Problem

A man and his wife, like many other couples and families, will be relocating to Amsterdam, Netherlands. They are relocating from Toronto, Canada and want to find neighborhoods that are similar to what they are used to. They are, unfortunately, not familiar with the Amsterdam's neighborhoods and need some help. They currently live in Willowdale, Toronto, Canada.

Solution approach

The approach that will be followed in pursuit of solving this problem will be to first cluster the Toronto neighborhoods according to the venues found within a 500m radius of a neighborhood's centroid. Then, using this built cluster model, classify the Amsterdam neighborhoods into the same clusters to find similar neighborhoods to the ones in Toronto. The data required for this will be discussed in the next section.

Interest

This is of interest to anyone relocating from one city to another, regardless of country, and require more information on their new city's neighborhoods, especially the similarity of these neighborhoods with their current city's neighborhoods.

Data

This section contains a description of the data to be used in this analysis, as well as an explanation of the data cleaning and preparation tasks that are required for the proposed analytical model.

Data required and sources

The data components required for the proposed analytical model are listed in the table below, as well as their respective sources.

Data component	Data source	Data fields
List of Toronto neighborhoods	Wikipedia page: https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=945633050.	- Neighborhood name - Neighborhood postal code
List of Amsterdam neighborhoods	Wikipedia page: https://en.wikipedia.org/wiki/Template:Neighborhoods_of_Amsterdam	- Neighborhood name - Neighborhood district
Geolocations of all neighborhoods	Bing Maps API	- Neighborhood latitude - Neighborhood longitude
Venue information per neighborhood	Foursquare API (Explore)	- Venue name - Venue latitude - Venue longitude - Venue category

Data cleaning and preparation

The following data cleaning tasks were done per data component:

Data component	Data cleaning tasks
List of Toronto neighborhoods	1. Create Soup from website, extract relevant data points 2. Remove Postal codes with unassigned neighborhoods
List of Amsterdam neighborhoods	1. Create Soup from website, extract relevant data points
Geolocations of all neighborhoods	1. Create a function that uses Bing Maps API to geocode an address 2. Loop through neighborhoods using the function to geocode addresses 3. Function retries geocoding when first attempts fail
Venue information per neighborhood	1. Create a function that uses Foursquare API to retrieve venue data per neighborhood, extract relevant data points per venue 2. Loop through neighborhoods and extract data through API using function

The following data preparation tasks were done to prepare the data for the analytical model:

Data preparation	Data preparation tasks
Determine frequency of occurrence of venues by category per neighborhood (used by clustering model)	1. Combine venue datasets into one dataframe (to ensure all categories are included for features in clustering model) 2. One hot encode each venue according to the venue category 3. Summarise the neighborhoods by calculating the mean of the frequency of occurrence of venues by venue category for each neighborhood
Identify the top 10 venue categories per neighborhood (used to view resulting clusters' neighborhoods)	1. Create a function that will sort venue categories per neighborhood by frequency of occurrence 2. Use function to sort venue categories and identify top 10 venue categories per neighborhood 3. Create a dataframe containing the 1^{st} most common venue category, up to 10^{th} most common venue category, per neighborhood.

Exploratory data analysis

This sub-section contains descriptions of the exploratory data analysis tasks that were done to gain a better understanding of the data that is required for this analytical model.

Neighborhoods data

There are 210 neighborhoods in Toronto, which have all been geocoded to obtain their latitudinal and longitudinal coordinates.

```
[10]: df_t.head(10)
```

[10]:

	PostalCode	Neighborhood	Address	Latitude	Longitude	Long_Lat
2	M3A	Parkwoods	Parkwoods, M3A, Toronto	43.755997	-79.329544	[43.75599670410156, -79.32954406738281]
3	M4A	Victoria Village	Victoria Village, M4A, Toronto	43.728336	-79.314789	[43.728336334228516, -79.31478881835938]
4	M5A	Harbourfront	Harbourfront, M5A, Toronto	43.655376	-79.365005	[43.65537643432617, -79.36500549316406]
5	M6A	Lawrence Heights	Lawrence Heights, M6A, Toronto	43.72192	-79.450676	[43.721920013427734, -79.45067596435547]
6	M6A	Lawrence Manor	Lawrence Manor, M6A, Toronto	43.725235	-79.439537	[43.72523498535156, -79.43953704833984]
7	M7A	Queen's Park	Queen's Park, M7A, Toronto	43.661913	-79.389937	[43.6619131706686, -79.3899373128105]
9	M9A	Islington Avenue	Islington Avenue, M9A, Toronto	43.667498	-79.533481	[43.6674979500411, -79.5334806996528]
10	M1B	Rouge	Rouge, M1B, Toronto	43.822937	-79.177452	[43.82293701171875, -79.17745208740234]
11	M1B	Malvern	Malvern, M1B, Toronto	43.8022	-79.223869	[43.80220031738281, -79.22386932373047]
13	M3B	Don Mills North	Don Mills North, M3B, Toronto	43.740825	-79.344493	[43.7408247192986, -79.3444934347813]

```
[43]: df_t.describe()
```

[43]:

	PostalCode	Neighborhood	Address	Latitude	Longitude	Long_Lat
count	210	210	210	210.000000	210.000000	210
unique	103	208	210	198.000000	197.000000	198
top	M9V	Runnymede	Harbourfront East, M5J, Toronto	43.643871	-79.381714	[43.64387130737305, -79.3958511352539]
freq	8	2	1	3.000000	3.000000	3

There are 77 neighborhoods in Amsterdam, which have all been geocoded to obtain their latitudinal and longitudinal coordinates.

```
[14]: df_a.head(10)
```

	District	Neighborhood	Address	Latitude	Longitude	Long_Lat
0	Centrum	Binnenstad (Oude Zijde - Nieuwe Zijde)	Binnenstad (Oude Zijde - Nieuwe Zijde), Amsterdam	52.45822	5.03278	[52.45822, 5.03278]
1	Centrum	Grachtengordel (Negen Straatjes)	Grachtengordel (Negen Straatjes), Amsterdam	52.40387	4.88928	[52.40387, 4.88928]
2	Centrum	Haarlemmerbuurt	Haarlemmerbuurt, Amsterdam	52.384697	4.886757	[52.38469696044922, 4.886756896972656]
3	Centrum	Jodenbuurt	Jodenbuurt, Amsterdam	52.369171	4.9025	[52.369171142578125, 4.902500152587891]
4	Centrum	Jordaan	Jordaan, Amsterdam	52.373295	4.879922	[52.373294830322266, 4.879921913146973]
5	Centrum	Kadijken	Kadijken, Amsterdam	52.368889	4.91556	[52.36888885498047, 4.915559768676758]
6	Centrum	Lastage	Lastage, Amsterdam	52.373085	4.903207	[52.373085021972656, 4.903206825256348]
7	Centrum	Oostelijke Eilanden (Czaar Peterbuurt)	Oostelijke Eilanden (Czaar Peterbuurt), Amsterdam	52.37003	4.92934	[52.37003, 4.92934]
8	Centrum	Oosterdokseiland	Oosterdokseiland, Amsterdam	52.37648	4.906045	[52.37648010253906, 4.906044960021973]
9	Centrum	Plantage	Plantage, Amsterdam	52.364498	4.910798	[52.364498138427734, 4.910798072814941]

```
[42]: df_a.describe()
```

	District	Neighborhood	Address	Latitude	Longitude	Long_Lat
count	77	77	77	77.00000	77.00000	77
unique	8	77	77	72.00000	72.00000	72
top	Centrum	Osdorp (De Aker - Middelveldsche Akerpolder)	Jodenbuurt, Amsterdam	52.36154	5.03846	[52.36154, 5.03846]
freq	14	1	1	5.00000	5.00000	5

Key observation: the number of Amsterdam neighborhoods is less than half of the number of Toronto neighborhoods. Luckily, Toronto's neighborhoods will be used to create and train the clustering model, so it is advantageous to have more data for this.

Venues data

There are 4810 different venues across Toronto's neighborhoods, that are categorised into 328 different venue categories.

```
[18]: toronto_venues.head(10)
```

	Neighborhood	Neighb. Latitude	Neighb. Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	City
0	Parkwoods	43.755997	-79.329544	Brookbanks Park	43.751976	-79.332140	Park	Toronto
1	Parkwoods	43.755997	-79.329544	TTC Stop #09083	43.759655	-79.332223	Bus Stop	Toronto
2	Parkwoods	43.755997	-79.329544	DVP at York Mills	43.758899	-79.334099	Intersection	Toronto
3	Parkwoods	43.755997	-79.329544	Chick-N-Joy	43.759900	-79.326520	Fried Chicken Joint	Toronto
4	Victoria Village	43.728336	-79.314789	Tim Hortons	43.725517	-79.313103	Coffee Shop	Toronto
5	Victoria Village	43.728336	-79.314789	Portugril	43.725819	-79.312785	Portuguese Restaurant	Toronto
6	Victoria Village	43.728336	-79.314789	The Frig	43.727051	-79.317418	French Restaurant	Toronto
7	Victoria Village	43.728336	-79.314789	Eglinton Ave E & Sloane Ave/Bermondsey Rd	43.726086	-79.313620	Intersection	Toronto
8	Victoria Village	43.728336	-79.314789	Pizza Nova	43.725824	-79.312860	Pizza Place	Toronto
9	Victoria Village	43.728336	-79.314789	Wigmore Park	43.731023	-79.310771	Park	Toronto

```
[52]: toronto_venues.describe(include=object)
```

	Neighborhood	Venue	Venue Category	City
count	4810	4810	4810	4810
unique	205	2356	328	1
top	St. James Town	Starbucks	Coffee Shop	Toronto
freq	131	114	382	4810

There are 2163 different venues across Amsterdam's neighborhoods, that are categorised into 257 different venue categories.

```
[22]: amsterdam_venues.head(10)
```

[22]:

	Neighborhood	Neighb. Latitude	Neighb. Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	City
0	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	Posthoorn	52.460915	5.035037	French Restaurant	Amsterdam
1	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	De Waegh	52.459240	5.036258	French Restaurant	Amsterdam
2	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	Beuqz	52.459148	5.036137	Café	Amsterdam
3	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	De Koperen Vis	52.459852	5.036748	Diner	Amsterdam
4	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	Cafe 1614	52.459243	5.035988	Pub	Amsterdam
5	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	Coffee & Cacao Lunchroom, Chocolaterie, Patiss...	52.459110	5.036719	Café	Amsterdam
6	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	De Zwaan	52.459129	5.036301	Café	Amsterdam
7	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	Bierderij Waterland Organic Brewery & Tasting ...	52.458379	5.039865	Brewery	Amsterdam
8	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	SPAR express Monnickendam Oost	52.458756	5.031091	Convenience Store	Amsterdam
9	Binnenstad (Oude Zijde - Nieuwe Zijde)	52.45822	5.03278	Four Seasons	52.461610	5.036410	Chinese Restaurant	Amsterdam

```
[53]: amsterdam_venues.describe(include=object)
```

[53]:

	Neighborhood	Venue	Venue Category	City
count	2163	2163	2163	2163
unique	71	1529	257	1
top	Oosterdokseiland	Albert Heijn	Hotel	Amsterdam
freq	100	24	113	2163

Key observation: to be able to use the same clustering model to classify Amsterdam's neighborhoods, one has to use all the venue categories found across Toronto and Amsterdam's neighborhoods when training the clustering model.

Top venues per neighborhood

The venues can be sorted according to frequency of occurrence of the venue category, per neighborhood. The top occurring venue categories can then be viewed. These datasets are used to view clusters' underlying neighborhoods once clustered.

[110]:

	City	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	Amsterdam	Admiralenbuurt	Bar	Restaurant	Deli / Bodega	Snack Place	Coffee Shop	Supermarket	Plaza	Japanese Restaurant	Middle Eastern Restaurant	Massage Studio
1	Amsterdam	Apollobuurt	Hotel	Steakhouse	Supermarket	Baby Store	Donut Shop	Lounge	Coffee Shop	Tram Station	Health Food Store	Bistro
2	Amsterdam	Banne Buiksloot	Bus Stop	Park	Supermarket	Shopping Mall	Drugstore	Bakery	Business Service	Turkish Restaurant	Café	Restaurant
3	Amsterdam	Bijlmer	Bus Stop	Arcade	Dog Run	ATM	Performing Arts Venue	Pastry Shop	Park	Paper / Office Supplies Store	Palace	Pakistani Restaurant
4	Amsterdam	Binnenstad (Oude Zijde - Nieuwe Zijde)	Café	Bus Stop	Snack Place	French Restaurant	Playground	Tourist Information Center	Pharmacy	Chinese Restaurant	Athletics & Sports	Asian Restaurant
5	Amsterdam	Bos en Lommer (Kolenkitbuurt - Landlust)	Restaurant	Fast Food Restaurant	Bakery	Park	Seafood Restaurant	Gym / Fitness Center	Bagel Shop	Bar	Mediterranean Restaurant	Bookstore
6	Amsterdam	Buiksloot	Bus Stop	Park	Supermarket	Shopping Mall	Drugstore	Bakery	Business Service	Turkish Restaurant	Café	Restaurant
7	Amsterdam	Buikslotermeer	Supermarket	Electronics Store	Bakery	Shoe Store	Sandwich Place	Seafood Restaurant	Discount Store	Bar	Market	Brasserie
8	Amsterdam	Buitenveldert	Hotel	Bakery	Drugstore	Restaurant	Sandwich Place	Coffee Shop	Supermarket	Park	Grocery Store	Massage Studio
9	Amsterdam	Bullewijk	Coffee Shop	Hotel	Café	Restaurant	Cafeteria	Bus Stop	Buffet	Brewery	Performing Arts Venue	Hostel

[111]:

	City	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
71	Toronto	Adelaide	Coffee Shop	Bar	Café	Hotel	Taco Place	Restaurant	Concert Hall	Seafood Restaurant	Theater	Thai Restaurant
72	Toronto	Agincourt	Chinese Restaurant	Print Shop	Vietnamese Restaurant	Bakery	Skating Rink	Badminton Court	Pharmacy	Sandwich Place	Discount Store	Coffee Shop
73	Toronto	Agincourt North	Chinese Restaurant	Liquor Store	Fast Food Restaurant	Dim Sum Restaurant	Clothing Store	Frozen Yogurt Shop	Fried Chicken Joint	Wings Joint	Bank	Park
74	Toronto	Albion Gardens	Grocery Store	Coffee Shop	Sandwich Place	Pizza Place	ATM	Other Great Outdoors	Park	Paper / Office Supplies Store	Palace	Pakistani Restaurant
75	Toronto	Alderwood	Pizza Place	Playground	Pharmacy	Convenience Store	Coffee Shop	ATM	Park	Paper / Office Supplies Store	Palace	Pakistani Restaurant
76	Toronto	Bathurst Manor	Playground	Park	Convenience Store	Baseball Field	ATM	Paper / Office Supplies Store	Palace	Pakistani Restaurant	Outdoors & Recreation	Outdoor Supply Store
77	Toronto	Bathurst Quay	Coffee Shop	Café	Park	Grocery Store	Gym	Japanese Restaurant	Pizza Place	Bank	Caribbean Restaurant	Ramen Restaurant
78	Toronto	Bayview Village	Trail	Flower Shop	Construction & Landscaping	Organic Grocery	Pastry Shop	Park	Paper / Office Supplies Store	Palace	Pakistani Restaurant	Outdoors & Recreation
79	Toronto	Beaumond Heights	Grocery Store	Sandwich Place	Pizza Place	Caribbean Restaurant	Beer Store	Coffee Shop	Fast Food Restaurant	Pharmacy	Fried Chicken Joint	Auto Garage
80	Toronto	Bedford Park	Coffee Shop	Italian Restaurant	Sandwich Place	Comfort Food Restaurant	Hobby Shop	Cupcake Shop	Restaurant	Pizza Place	Park	Fast Food Restaurant

Final dataset (model ready):

An excerpt from the final dataset containing the frequency of occurrence of venue categories in a format that is ready for model can be seen below. This dataset consists of 276 rows (Toronto and Amsterdam neighborhoods combined) and 387 columns (387 unique venue categories across Toronto and Amsterdam neighborhoods)

[81]:

	City	Neighborhood	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
0	Amsterdam	Admiralenbuurt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	Amsterdam	Apollobuurt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Amsterdam	Banne Buiksloot	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Amsterdam	Bijlmer	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Amsterdam	Binnenstad (Oude Zijde - Nieuwe Zijde)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

[112]:
```
all_grouped.shape
```

[112]:
```
(276, 387)
```

Methodology & Results

This section contains an explanation of the analytical approach that was followed in the pursuit of solving the problem as stated in the Introduction section.

Approach

Clustering is done on the Toronto neighborhoods to create clusters of similar neighborhoods based on the venues found in those neighborhoods. The model built from the clustering is then used to classify the Amsterdam neighborhoods into the same clusters, to find similar neighborhoods to the ones in Toronto.

K-means Clustering was found to be the most suitable for this problem because of its simplicity and strength, which is advantageous considering the number of features in the dataset. Also, there is no need for a hierarchy of clusters, so hierarchical clustering was not considered. The *k* parameter is specified as 10 (i.e. 10 clusters). 10 is chosen for *k* because of the meaningful split that results in the neighborhoods, compared to other, lower values for *k*.

K-means Clustering is applied using the KMeans package from the sklearn.cluster module. This package allows you to create a KMeans object which makes use of machine learning to build the clustering model, given certain parameters and training data. The training dataset consists of the frequency of occurrence of the venue categories (as features) per neighborhood. Clustering is done on the Toronto neighborhoods using the *fit()* function of the KMeans object.

A model is created and stored in the KMeans object, based on the clustering of the Toronto neighborhoods. This created model is then used to predict the cluster of each Amsterdam neighborhood, given the frequency of occurrence of the different venue categories in each neighborhood. Basically, the created clustering model stored in the KMeans object is used to classify the Amsterdam neighborhoods using the *predict()* function of the object.

The resulting cluster labels given to Toronto and Amsterdam neighborhoods are presented and analysed by plotting the neighborhoods on a geographical map and using colored markers (a different color for each cluster). This is done using the *folium* package. These maps can be used to browse through neighborhoods that are similar to the neighborhoods of Toronto.

Results, Discussion & Conclusion

This section contains the results from the built model, using the data prepared earlier in the process. The results are analysed and discussed in the last part of this section.

Results

The resulting cluster labels are combined with the dataset containing the neighborhood names, along with their top 10 occurring venue categories. Below figure shows an excerpt from this dataframe containing the Toronto neighborhoods.

[37]:

	PostalCode	Neighborhood	Address	Latitude	Longitude	Long_Lat	Cluster Labels	City	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
2	M3A	Parkwoods	Parkwoods, M3A, Toronto	43.755997	-79.329544	[43.75599670410156, -79.32954406738281]	1	Toronto	Park	Intersection	Bus Stop	Fried Chicken Joint	Accessories Store	Organic Grocery	Paper / Office Supplies Store	Palace	Pakistani Restaurant	Outdoor Supply Store
3	M4A	Victoria Village	Victoria Village, M4A, Toronto	43.728336	-79.314789	[43.728336334228516, -79.31478881835938]	3	Toronto	Park	Intersection	Portuguese Restaurant	French Restaurant	Coffee Shop	Pizza Place	Optical Shop	Paper / Office Supplies Store	Palace	Pakistani Restaurant
4	M5A	Harbourfront	Harbourfront, M5A, Toronto	43.655376	-79.365005	[43.65537643432617, -79.36500549316406]	4	Toronto	Coffee Shop	Italian Restaurant	Theater	Sandwich Place	Thrift / Vintage Store	Spa	Bakery	Gym / Fitness Center	Bar	Thai Restaurant
5	M6A	Lawrence Heights	Lawrence Heights, M6A, Toronto	43.72192	-79.450676	[43.721920013427734, -79.45067596435547]	4	Toronto	Clothing Store	Coffee Shop	Dessert Shop	Cosmetics Shop	Women's Store	Restaurant	Sushi Restaurant	Food Court	Toy / Game Store	Bakery
6	M6A	Lawrence Manor	Lawrence Manor, M6A, Toronto	43.725235	-79.439537	[43.72523498535156, -79.43953704833984]	6	Toronto	Park	Accessories Store	Optical Shop	Paper / Office Supplies Store	Palace	Pakistani Restaurant	Outdoor Supply Store	Outdoor Sculpture	Other Great Outdoors	Other Event

And below is an excerpt from the Amsterdam dataframe.

[40]:

	District	Neighborhood	Address	Latitude	Longitude	Long_Lat	Cluster Labels	City	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	Centrum	Binnenstad (Oude Zijde - Nieuwe Zijde)	Binnenstad (Oude Zijde - Nieuwe Zijde), Amsterdam	52.45822	5.03278	[52.45822, 5.03278]	4	Amsterdam	Bus Stop	Café	French Restaurant	Diner	Pizza Place	Coffee Shop	Snack Place	Athletics & Sports	Asian Restaurant	Harbor / Marina
1	Centrum	Grachtengordel (Negen Straatjes)	Grachtengordel (Negen Straatjes), Amsterdam	52.40387	4.88928	[52.40387, 4.88928]	0	Amsterdam	Restaurant	Hotel	Boat or Ferry	Steakhouse	Bagel Shop	Arcade	Bakery	General Entertainment	Harbor / Marina	Sandwich Place
2	Centrum	Haarlemmerbuurt	Haarlemmerbuurt, Amsterdam	52.384697	4.886757	[52.38469696044922, 4.886756896972656]	4	Amsterdam	Bar	Deli / Bodega	Plaza	Coffee Shop	Café	Restaurant	Sandwich Place	French Restaurant	Italian Restaurant	Music Venue
3	Centrum	Jodenbuurt	Jodenbuurt, Amsterdam	52.369171	4.9025	[52.369171142578125, 4.902500152587891]	4	Amsterdam	Hotel	Café	Coffee Shop	Bar	Cocktail Bar	History Museum	French Restaurant	Greek Restaurant	Grocery Store	Beer Bar
4	Centrum	Jordaan	Jordaan, Amsterdam	52.373295	4.879922	[52.373294830322266, 4.879921913146973]	4	Amsterdam	Bar	Coffee Shop	Café	Italian Restaurant	Hotel	Pizza Place	Ice Cream Shop	Cocktail Bar	Art Gallery	Chocolate Shop

The count for Toronto neighborhoods per cluster can be seen below.

[38]:

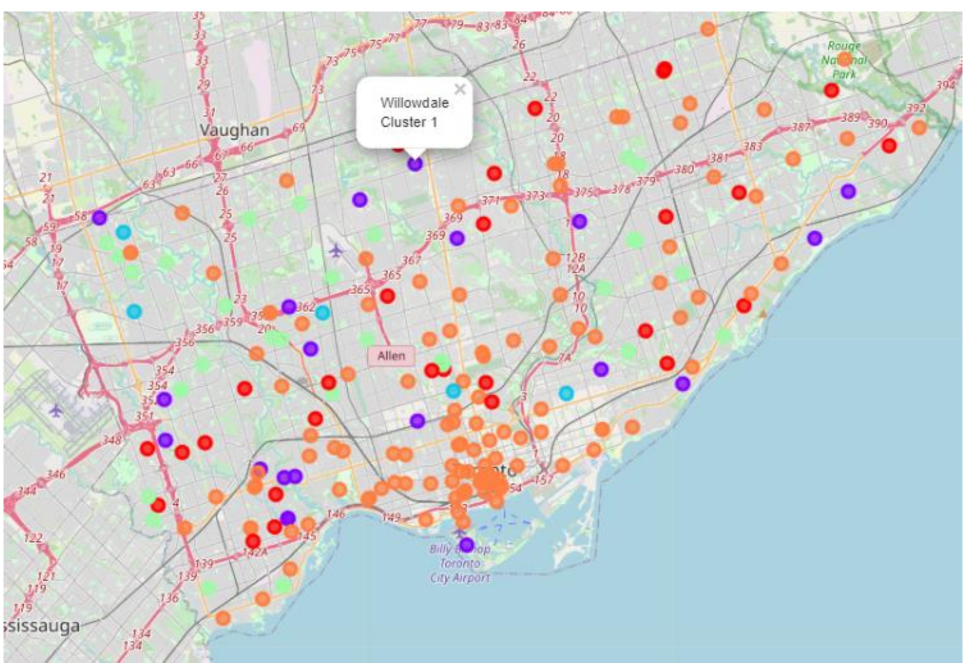
Cluster Labels	Neighborhood
0	8
1	20
2	5
3	31
4	123
5	2
6	9
7	5
8	3
9	1

And below is the count for Amsterdam neighborhoods per the same clusters.

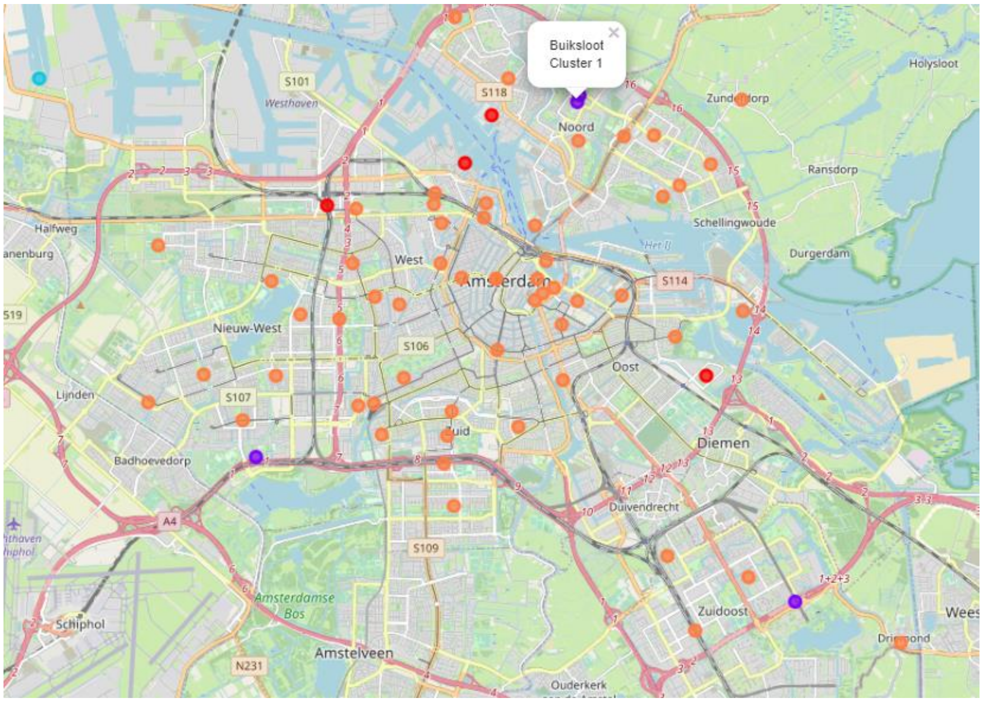
[41]:

Cluster Labels	Neighborhood
0	5
1	4
2	1
4	61

The clustered neighborhoods were plotted on a geographical map using a different color marker for a different cluster. Shown below is the resulting map for the Toronto neighborhoods. This map is interactive – when you click on a marker, a popup will appear showing the name of the neighborhood and its cluster.



Shown below is the resulting map for the Amsterdam neighborhoods.



When analysing the clusters, a count of the top 10 occurring venue categories gives one a good idea of the profile of the cluster. In the table below is shown the 20 venue categories that are found the most in the top 10 venue categories of the number 9 labeled cluster for Toronto.

```
[145]: Coffee Shop                     90
       Café                            63
       Restaurant                      49
       Park                            46
       Sandwich Place                  40
       Grocery Store                   34
       Italian Restaurant              31
       Pizza Place                     30
       Bakery                          28
       Fast Food Restaurant            26
       Pharmacy                        25
       Bar                             25
       Gym                             23
       Hotel                           20
       Bank                            20
       Japanese Restaurant             19
       Paper / Office Supplies Store   18
       Palace                          17
       ATM                             16
       Sushi Restaurant                14
       dtype: int64
```

And here are the counts for the 20 venue categories for Amsterdam.

```
[146]: Restaurant                    32
       Bar                           29
       Coffee Shop                   28
       Café                          28
       Park                          25
       Hotel                         24
       Bus Stop                      18
       Paper / Office Supplies Store 16
       Bakery                        16
       Supermarket                   15
       ATM                           14
       Palace                        14
       Snack Place                   13
       Italian Restaurant            12
       Pakistani Restaurant          11
       Plaza                          9
       Pizza Place                    9
       French Restaurant              9
       Cocktail Bar                   9
       Gym                            9
       dtype: int64
```

Discussion

When the number of neighborhoods per cluster is counted, it is clear that some clusters are more prevalent than others, indicating that there are more of these similar type neighborhoods in the city. The cluster that is predominant is labeled '4'. From the venue categories counts found in this cluster's top 10 venue categories, it is evident that these cluster 4 neighborhoods are dominated by restaurants, bars, coffee shops, cafés, parks, hotels, and so on.

Below follows a brief profile of each cluster, listing the 5 venue categories that are most prevalent in each of these cluster's neighborhoods' top 10 occurring venue categories.

Cluster label	Toronto: 5 most prevalent top 10 venue categories (occurences)	Amsterdam: 5 most prevalent venue categories
0	Restaurant (8) Palace (6) Park (6) Paper / Office Supplies Store (4) Pakistani Restaurant (4)	Restaurant (5) Hotel (3) Sandwich Place (2) Coffee Shop (2) Convenience Store (2)
1	Park (20) Palace (19) Pakistani Restaurant (19) Paper / Office Supplies Store (18) Outdoor Supply Store (16)	Park (4) Café (3) Bus Stop (3) Palace (2) Bakery (2)
2	Palace (5) Paper / Office Supplies Store (5) Coffee Shop (5) Park (5) Pastry Shop (4)	Pakistani Restaurant (1) Park (1) Accessories Store (1) Organic Grocery (1) Palace (1)
3	Pizza Place (25) Paper / Office Supplies Store (21)	N/A

	Palace (20) Coffee Shop (15) Pakistani Restaurant (15)	
4	Coffee Shop (90) Café (59) Restaurant (45) Park (40) Bakery (34)	Restaurant (30) Coffee Shop (29) Café (28) Bar (26) Hotel (22)
5	Pakistani Restaurant (2) Park (2) Pharmacy (2) Accessories Store (2) Palace (2)	N/A
6	Other Great Outdoors (9) Outdoor Sculpture (9) Palace (9) Pakistani Restaurant 9) Outdoor Supply Store (9)	N/A
7	Trail (5) Pakistani Restaurant (5) Outdoor Sculpture (5) Palace (5) Outdoor Supply Store (5)	N/A
8	Pakistani Restaurant (3) Park (3) Other Great Outdoors (3) Accessories Store (3) Palace (3)	N/A
9	Seafood Restaurant (1) Pakistani Restaurant (1) Other Great Outdoors (1) Accessories Store (1) Palace (1)	N/A

In the end, the Toronto clusters weren't as transferrable to the Amsterdam neighborhoods, considering only four of the nine clusters were actually 'found' in the Amsterdam neighborhoods. This could be due to data limitations or accuracy.

Conclusion

The problem was solved using the proposed analytical approach. It is possible to classify neighborhoods of one city, using a clustering model based on another city's neighborhoods. Using this model, one can identify neighborhoods that are similar to the one's found in the training data, which could assist people relocating from one city to another in finding a suitable neighborhood to look for accommodation.

In the case of the couple moving from Willowdale, Toronto: Willowdale is clustered into the cluster labeled ‘1’. The Amsterdam neighborhoods that are classified as similar to Willowdale’s cluster is shown below.

[51]:

	Neighborhood	Long_Lat	Cluster Labels	City	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
20	Sloten	[52.339561462402344, 4.816626071929932]	1	Amsterdam	Café	Park	Hotel	Bus Stop	Diner	Other Event	Pastry Shop	Paper / Office Supplies Store	Palace	Pakistani Restaurant
23	Banne Buiksloot	[52.40760803222656, 4.916283130645752]	1	Amsterdam	Park	Bus Stop	Supermarket	Restaurant	Bakery	Café	Turkish Restaurant	Shopping Mall	Drugstore	Office
24	Buiksloot	[52.406494140625, 4.9156270027160645]	1	Amsterdam	Park	Bus Stop	Supermarket	Restaurant	Bakery	Café	Turkish Restaurant	Shopping Mall	Drugstore	Office
72	Gaasperdam	[52.31224060058594, 4.982944011688232]	1	Amsterdam	Food & Drink Shop	Park	Tunnel	Bus Station	Organic Grocery	Paper / Office Supplies Store	Palace	Pakistani Restaurant	Outdoor Supply Store	Outdoor Sculpture

The profile of this cluster can be viewed in the Discussion section.

There is, however, room for improvement in this model. The following is a list of proposed possible improvements that can be explored for this model:

1. Use a larger (or smaller) value for the radius around a neighborhood’s centroid from which to retrieve venue information.
2. Test and compare the DBSCAN clustering algorithm’s performance and accuracy.
3. Include other neighborhood data that could improve the profiling of neighborhoods, not just venues.
4. Do clustering on all neighborhoods (both cities).