CAPSTONE PROJECT

Battle of the Neighborhoods

Emile Strasheim
August 2021

Agenda

1. Introduction
2. Data
3. Methodology
4. Results
5. Discussion
6. Conclusion

Introduction

- **Background**: many families relocating to new cities and/or countries
- **Problem**: which neighborhood to live in?
- **Target audience**: anyone relocating from one city to another
- **Case study**: relocating from Willowdale, Toronto, Canada to Amsterdam, Netherlands

- **Approach**: create clustering model using Toronto neighborhoods, then classify Amsterdam neighborhoods.

Data

Data required and sources:

Data component	Data source	Data fields
List of Toronto neighborhoods	Wikipedia page: List of postal codes of Canada: M (Toronto)	- Neighborhood name - Neighborhood postal code
List of Amsterdam neighborhoods	Wikipedia page: 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

Data cleaning:

- Create soups from website, extract data points
- Remove postal codes with unassigned neighborhoods
- Retry geocoding on failed addresses

Data

Exploratory data analysis:

- Toronto neighborhoods: 210
- Amsterdam neighborhoods: 77

```
[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

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

[42]:

	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**: Amsterdam much less. Luckily, Toronto used for training of algorithm.

Data

Exploratory data analysis:

- Toronto: 4810 venues, 328 venue categories
- Amsterdam: 2163 venues, 257 venue categories

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

[52]:

	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

```
[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**: need to combine datasets to use all venue categories for training of algorithm.

Data

Data preparation:

Data preparation	Data cleaning 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 1st most common venue category, up to 10th most common venue category, per neighborhood.

Data

Exploratory data analysis:

- Top venue categories by frequency of occurrence:

[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

[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

Data

Final input dataset:

[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

1. K-means clustering on Toronto Neighborhoods
    a. *k* parameter = 10 (number of clusters)
    b. sklearn.cluster.KMeans.fit()

2. Classify Amsterdam neighborhoods with cluster model
    a. sklearn.cluster.KMeans.predict()

3. Analyse clusters geographical map using *Folium* package

Results

Toronto neighborhoods:

[38]:

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

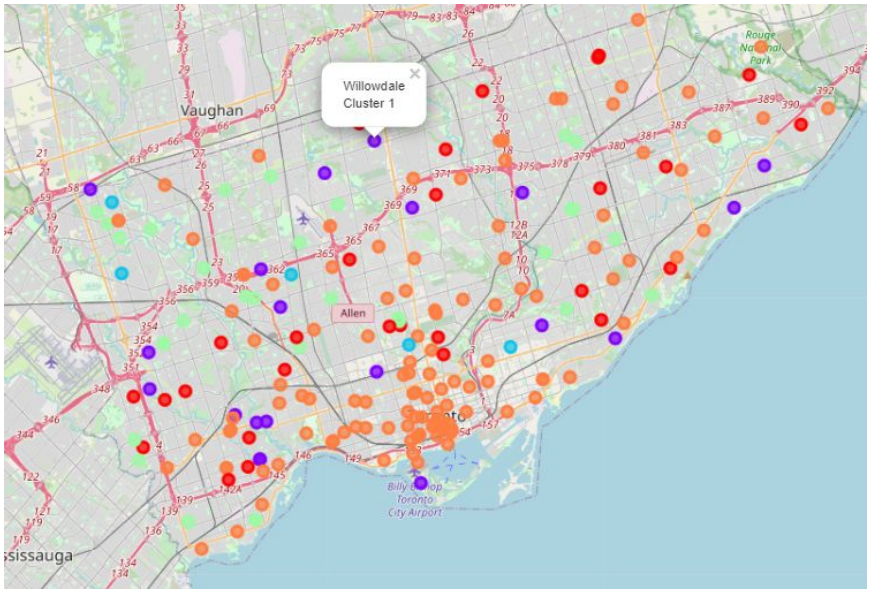
Amsterdam neighborhoods

[41]:

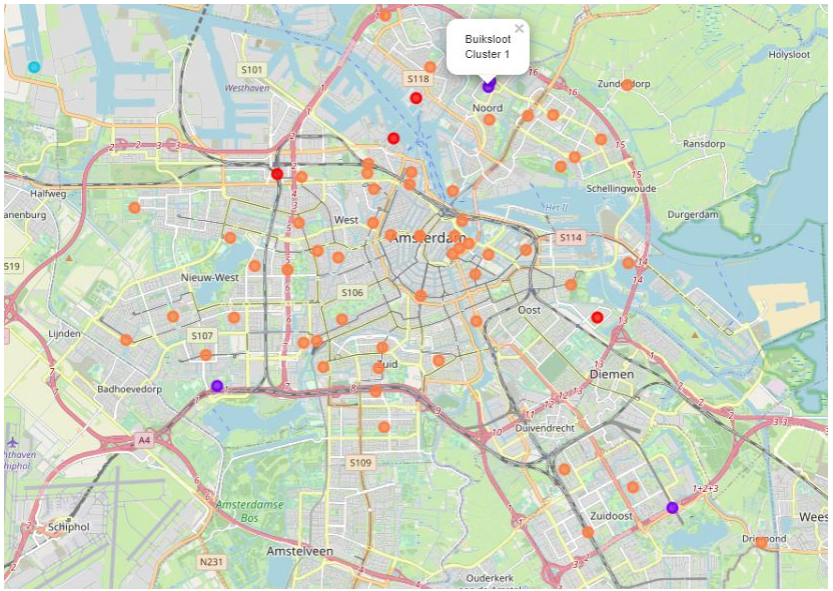
Cluster Labels	Neighborhood
0	5
1	4
2	1
4	61

Results

Toronto neighborhoods:



Amsterdam neighborhoods



Discussion

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) Palace (20) Coffee Shop (15) Pakistani Restaurant (15)	N/A

Cluster label	Toronto: 5 most prevalent top 10 venue categories (occurences)	Amsterdam: 5 most prevalent venue categories
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

Discussion

Cluster label	Toronto: 5 most prevalent top 10 venue categories (occurences)	Amsterdam: 5 most prevalent venue categories
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

Conclusion

- Some clusters more prevalent
- Predominant cluster: cluster 4
- Case study: Willowdale - cluster 1. Amsterdam cluster 1:

[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

Conclusion

Future improvements:

1. Use larger/smaller radius for venues
2. Test/compare performance of DBSCAN algorithm
3. Include other neighborhood data
4. Create clusters using both cities' data