

Coursera Capstone Project: Applied Data Science

Athanasis Serafeim, 2020

Overview

Introduction

Data

- Neighborhoods

- Geocoding

- Foursquare API data

Methodology

- Geocoding API

- Folium

- One hot encoding

- Optimal number of Clusters

Results

Discussion/Conclusion

Introduction

The purpose of this report is to conduct an extensive study in the areas of interest of the city of Toronto. Toronto is one of the world's most popular travel destinations and due to each multicultural variability has a variety of options and experiences to offer. The aim is to help tourists in the city of Toronto choose the right destination for them, depending on the facilities each neighborhood of Toronto has to offer.

This can also help locals find out places they do not know about and also take decisions about moving to another neighborhood. Findings will help stakeholders make better decisions and address any concerns they have, based on information about the different kinds of cuisines, provision stores and facilities each place of the city has to offer.

Data

Neighborhoods:

Data from: "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M"

- 1.borough: Name of Neighborhood
- 2.town: Name of borough
- 3.post_code: Postal codes for Toronto.

Data

Neighborhoods:

Postal Code	Borough	Neighbourhood
0	M3A	North York Parkwoods
1	M4A	North York Victoria Village
2	M5A	Downtown Toronto Regent Park, Harbourfront
3	M6A	North York Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto Queen's Park, Ontario Provincial Government
...
98	M8X	Etobicoke The Kingsway, Montgomery Road, Old Mill North
99	M4Y	Downtown Toronto Church and Wellesley
100	M7Y	East Toronto Business reply mail Processing Centre, South C...
101	M8Y	Etobicoke Old Mill South, King's Mill Park, Sunnylea, Hu...
102	M8Z	Etobicoke Mimico NW, The Queensway West, South of Bloor,...

Data

Geocoder API:

Data from Geocoder API online including:

1. Latitude: Latitude for Neighborhood
2. Longitude: Longitude for Neighborhood

Data

Geocoder API:

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476
...
98	M9N	43.706876	-79.518188
99	M9P	43.696319	-79.532242
100	M9R	43.688905	-79.554724
101	M9V	43.739416	-79.588437
102	M9W	43.706748	-79.594054

Data

Foursquare API data:

- 1.Neighborhood: Name of the Neighborhood
- 2.Neighborhood Latitude: Latitude of the Neighborhood
- 3.Neighborhood Longitude: Longitude of the Neighborhood
- 4.Venue: Name of the Venue
- 5.Venue Category: Category of Venue

Data

Foursquare API data:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Tim Hortons	Coffee Shop
4	Victoria Village	43.725882	-79.315572	Portugril	Portuguese Restaurant
...
1306	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	South St. Burger	Burger Joint
1307	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Wingporium	Wings Joint
1308	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Dollarama	Discount Store
1309	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Artisano Bakery Café	Bakery
1310	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Healthy Planet	Supplement Shop

Methodology

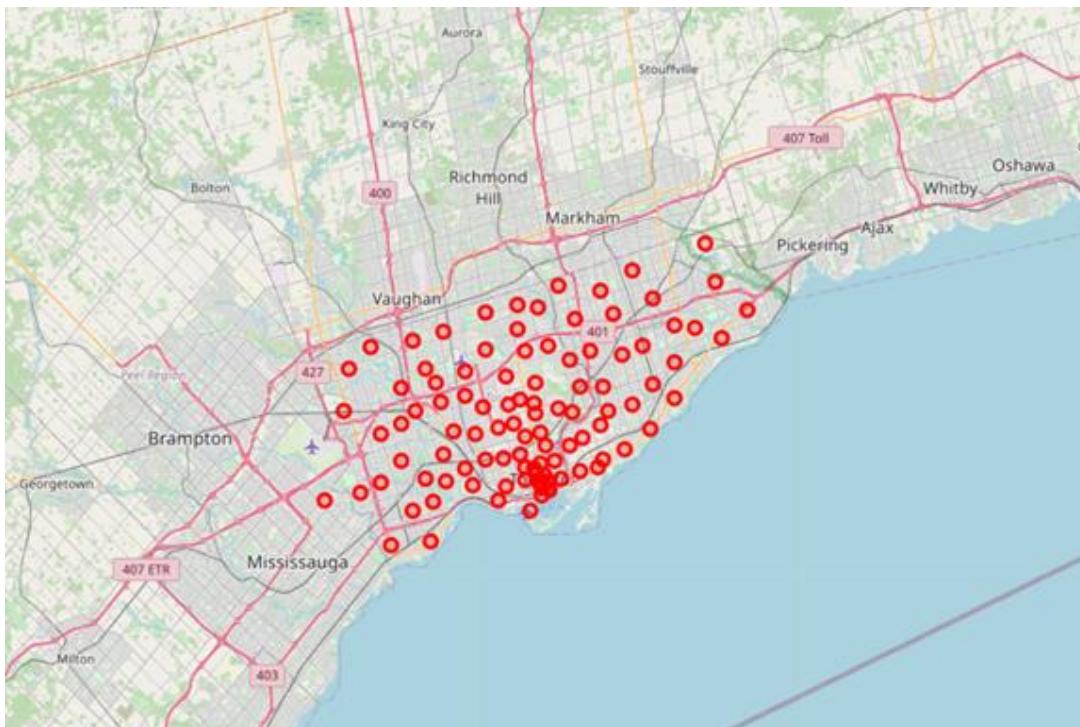
Geocoding API:

A thorough analysis of the principles of methods have been made in order to ensure the results are as accurate as possible. In the initial development phase with Open Geocoder API, the number of errors were of an appropriate amount, so this is the method we kept up with.

Methodology

Folium:

Folium build on data bonds combining python capabilities with the mapping strengths of the leaflet.js libraries. All cluster visualization are done by Folium which generates a Leaflet map with OpenStreetMap technology.



Methodology

One hot encoding:

Categorical variables are converted into a table form that can be provided to ML algorithms and achieve higher level predictions.

Methodology

Optimal number of Clusters:

- For K-means Clustering Algorithm.
- All unique items under Venue Category are one-hot encoded.
- The top 10 common venues and a new Data Frame is made.
- Silhouette Score is a measure of how similar an object is to its own cluster compared to other ones.
- The venue data is then trained using the K-means Clustering Algorithm to get the desired cluster based on the analysis.
- K- means was chosen as the variable's categories are large enough.

Results

The neighborhoods are divided into $n = 5$ clusters, where n is the number of clusters found using the optimal approach:



Results

	Borough	Cluster Labels	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	North York	0.0	Park	Food & Drink Shop	Yoga Studio	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
21	York	0.0	Park	Women's Store	Pool	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
35	East York	0.0	Park	Convenience Store	Intersection	Yoga Studio	Department Store	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop
49	North York	0.0	Park	Construction & Landscaping	Bakery	Yoga Studio	Department Store	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop
61	Central Toronto	0.0	Park	Swim School	Bus Line	Yoga Studio	Department Store	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
64	York	0.0	Park	Yoga Studio	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store
66	North York	0.0	Park	Convenience Store	Yoga Studio	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
85	Scarborough	0.0	Intersection	Playground	Park	Bakery	Yoga Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
91	Downtown Toronto	0.0	Park	Playground	Trail	Yoga Studio	Dance Studio	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center

Results

	Borough	Cluster Labels	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
1	North York	1.0	Pizza Place	Coffee Shop	Hockey Arena	Intersection	Portuguese Restaurant	French Restaurant	Dog Run	Donut Shop	Distribution Center	Drugstore
2	Downtown Toronto	1.0	Coffee Shop	Park	Bakery	Breakfast Spot	Theater	Distribution Center	Spa	Dessert Shop	Event Space	Farmers Market
3	North York	1.0	Clothing Store	Accessories Store	Boutique	Gift Shop	Furniture / Home Store	Event Space	Coffee Shop	Women's Store	Vietnamese Restaurant	Airport Terminal
4	Downtown Toronto	1.0	Coffee Shop	Yoga Studio	Diner	Sandwich Place	Portuguese Restaurant	Park	Mexican Restaurant	Italian Restaurant	Hobby Shop	Fried Chicken Joint
7	North York	1.0	Gym	Beer Store	Coffee Shop	Japanese Restaurant	Café	Dim Sum Restaurant	Supermarket	Restaurant	Caribbean Restaurant	Chinese Restaurant
...
98	Etobicoke	1.0	Pool	River	Yoga Studio	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
99	Downtown Toronto	1.0	Gym	Escape Room	Sushi Restaurant	Ice Cream Shop	Indian Restaurant	Ramen Restaurant	Beer Bar	Japanese Restaurant	Ethiopian Restaurant	Juice Bar
100	East Toronto	1.0	Pizza Place	Auto Workshop	Park	Light Rail Station	Burrito Place	Farmers Market	Fast Food Restaurant	Butcher	Restaurant	Recording Studio
101	Etobicoke	1.0	Construction & Landscaping	Baseball Field	Yoga Studio	Dessert Shop	Ethiopian Restaurant	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop
102	Etobicoke	1.0	Gym	Thrift / Vintage Store	Convenience Store	Discount Store	Burrito Place	Burger Joint	Sandwich Place	Kids Store	Fast Food Restaurant	Bakery

Results

Borough	Cluster Labels	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	
45	North York	2.0	Martial Arts School	Yoga Studio	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store
Borough	Cluster Labels	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	
11	Etobicoke	3.0	Print Shop	Yoga Studio	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store
Borough	Cluster Labels	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	
6	Scarborough	4.0	Fast Food Restaurant	Yoga Studio	Event Space	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center

Discussion/Conclusion

- Cluster 0 is an appropriate choice for anyone who loves a walk on a park and some sightseeing.
- Cluster 1 is aimed at people who would like to live permanently in the area and not just ordinary visitors.
- Clusters 2,3 and 4 are aimed at a very specific audience with specific requirements and needs.