The Accidental Tourist Reccomendation System

Gustavo Sbardelotto Moreira
gustamo@gmail.com
Coursera
June, 22nd 2020

Table of Contents

Introduction/Business Problem

A lot of business travelers don't like to travel at all. Instead, they would prefer to stay at the comfort of their homes. The Accidental Tourist is one of my favorite movies, where the main character is a writer of travels books that gives a lot of tips on how the business travelers can fly to the most different cities, but still have the feeling they're still at home.

Inspired by the movie, I've built a Jupyter Notebook where the user will input which neighborhood he lives. For the sake of simplicity, the system only allows travellers that live in New York and are flying to Toronto. "Travelling with Accidental Tourist is like going on a cocoon", one reader said.So the system will list thes neighborhoods in Toronto that are very similiar with the one he lives. It will be very nice for the user to leave his hotel and find similar venues, restaurants, parks, and só on.

Methodology

Research Methods and Data Sources

All neighborhoods from Toronto where fetched from **Wikipedia** and from **Coursera** files. The information on these datasets are: Postal Code, Neighborhood and Borough. Also, their geolocations were fetched (Latitude and Longitude) from a CSV file.

The same information for New York is fetched online from this source:
https://cocl.us/new_york_dataset

With this data compiled, we give an option for the user to pick which neighborhood he lives in New York:

Select your NY neighborhood on the list below, so later the system will find similar Neighborhoods in Toronto:

```
[33]: display(dropdown_Neighb)
```

Neighborh... Manhattan Beach

Then, this neighborhood is merged with all Toronto neighborhoods in a single Dataset (all Toronto + 1 New York). Here is a sample of the dataframe:

	Borough	Neighborhood	Latitude	Longitude	city	Postalcode
77	Brooklyn	Manhattan Beach	40.577914	-73.943537	NY	NY
0	North York	Parkwoods	43.753259	-79.329656	TO	M3A
1	North York	Victoria Village	43.725882	-79.315572	TO	M4A
2	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636	TO	M5A
3	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763	TO	M6A

The next step is to fech venues data on **Foursquare,** for those neighborhoods. The most important information is the Venue Category, because it will be the basis for our clustering. Here is a sample of the data fetched:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Manhattan Beach	40.577914	-73.943537	Manhattan Beach	40.577370	-73.945531	Beach
1	Manhattan Beach	40.577914	-73.943537	manhattan beach playground	40.577115	-73.946587	Playground
2	Manhattan Beach	40.577914	-73.943537	Carvel Express	40.577962	-73.943551	Ice Cream Shop
3	Manhattan Beach	40.577914	-73.943537	Chillax Manhattan Beach Cafe	40.578836	-73.938229	Café
4	Manhattan Beach	40.577914	-73.943537	MTA Bus - B1/B49 - Oriental Blvd & Hastings St	40.577933	-73.944004	Bus Stop

One important decision that this project made is that **only the venues that appears more than 25 times** were selected. This avoided the great diversity that would affect on the clustering.

```
Venue Category
Coffee Shop            175
Café                   100
Restaurant              66
Park                    54
Pizza Place             50
Sandwich Place          44
Italian Restaurant      44
Bakery                  43
Hotel                   40
Japanese Restaurant     39
Clothing Store          33
Gym                     33
Sushi Restaurant        30
Grocery Store           29
Bar                     29
Fast Food Restaurant    26
Pub                     26
Bank                    25
American Restaurant     25
Breakfast Spot          24
Seafood Restaurant      22
Thai Restaurant         21
Pharmacy                21
Ice Cream Shop          19
Diner                   18
Name: Neighborhood, dtype: int64
```

The data is subited to one hot encoding, to convert the categories into columns. Here is a sample:

	American Restaurant	Bakery	Bank	Bar	Breakfast Spot	Café	Clothing Store	Coffee Shop	Diner	Fast Food Restaurant	Grocery Store	Gym	Hotel	Ice Cream Shop	Italian Restaurant	Japanese Restaurant	Park	Pharmacy	Pizza Place	Pub	Restaurant	S
2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	

The next step is to get the 10 most common venues of each neighborhood in a DataFrame:

	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	Agincourt	Breakfast Spot	Thai Restaurant	Gym	Bakery	Bank	Bar	Café	Clothing Store	Coffee Shop	Diner
1	Alderwood / Long Branch	Pizza Place	Sandwich Place	Pub	Coffee Shop	Gym	Thai Restaurant	Fast Food Restaurant	Bakery	Bank	Bar
2	Bathurst Manor / Wilson Heights / Downsview North	Bank	Coffee Shop	Diner	Sandwich Place	Restaurant	Pizza Place	Pharmacy	Ice Cream Shop	Sushi Restaurant	Grocery Store
3	Bayview Village	Bank	Café	Japanese Restaurant	Thai Restaurant	Gym	Bakery	Bar	Breakfast Spot	Clothing Store	Coffee Shop
4	Bedford Park / Lawrence Manor East	Italian Restaurant	Sandwich Place	Restaurant	Coffee Shop	Thai Restaurant	Breakfast Spot	Café	Grocery Store	Sushi Restaurant	American Restaurant

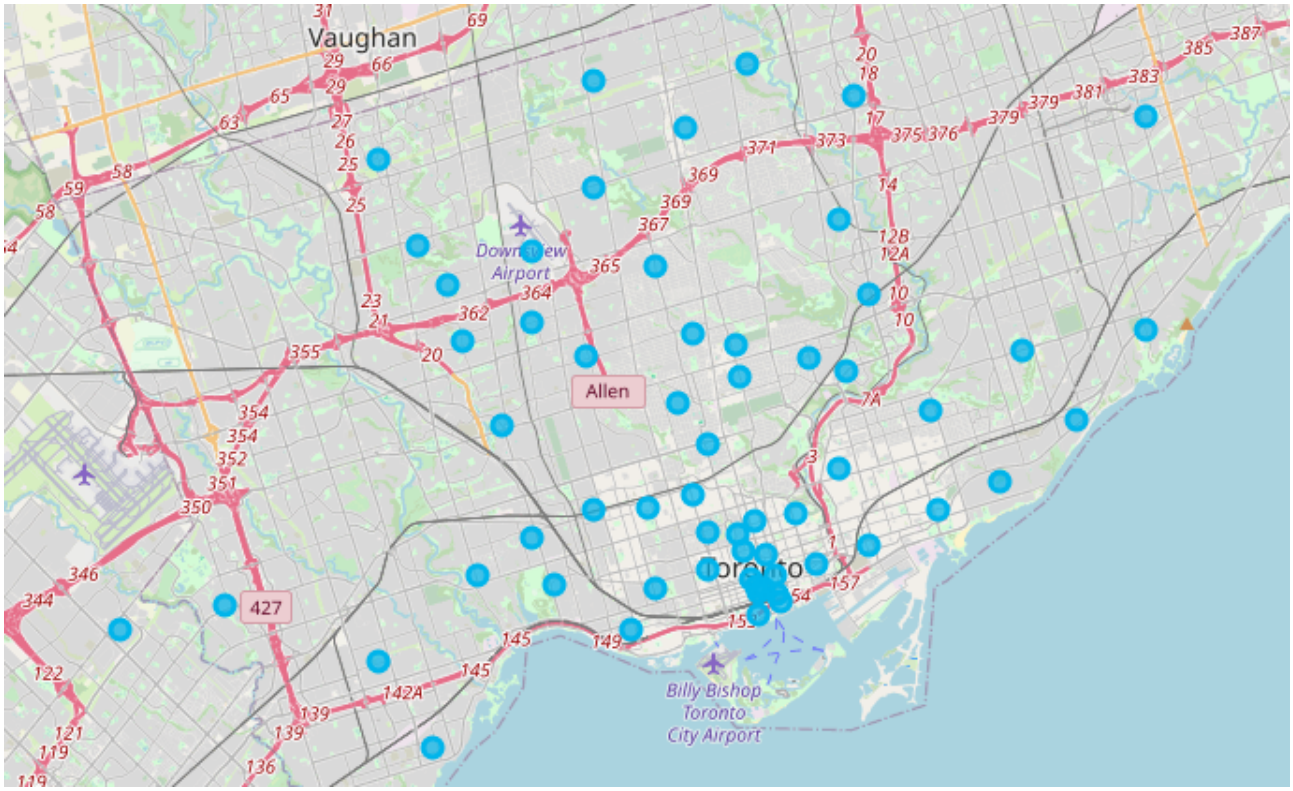
The clustering algorithm runs, and classify the cluster into 5 categories. Here is a sample of the clustering numbers:

	Borough	Neighborhood	Latitude	Longitude	city	Postalcode	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
77	Brooklyn	Manhattan Beach	40.577914	-73.943537	NY	NY	2	Café	Sandwich Place	Pizza Place	Ice Cream Shop	Thai Restaurant	Grocery Store	Bakery	Bank	Bar	Breakfast Spot
0	North York	Parkwoods	43.753259	-79.329656	TO	M3A	3	Park	Thai Restaurant	Gym	Bakery	Bank	Bar	Breakfast Spot	Café	Clothing Store	Coffee Shop
1	North York	Victoria Village	43.725882	-79.315572	TO	M4A	1	Pizza Place	Coffee Shop	Thai Restaurant	Gym	Bakery	Bank	Bar	Breakfast Spot	Café	Clothing Store
2	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636	TO	M5A	2	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Café	Bank	Restaurant	Ice Cream Shop	Thai Restaurant
3	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763	TO	M6A	2	Clothing Store	Coffee Shop	Thai Restaurant	Gym	Bakery	Bank	Bar	Breakfast Spot	Café	Diner

Then the system finds on which cluster of Toronto the “home neighborhood” from New York, and display the dataset for this cluster:

	Neighborhood	Latitude	Postalcode	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
77	Manhattan Beach	40.577914	NY	2	Café	Sandwich Place	Pizza Place	Ice Cream Shop	Thai Restaurant	Grocery Store	Bakery	Bank	Bar	Breakfast Spot
2	Regent Park / Harbourfront	43.654260	M5A	2	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Café	Bank	Restaurant	Ice Cream Shop	Thai Restaurant
3	Lawrence Manor / Lawrence Heights	43.718518	M6A	2	Clothing Store	Coffee Shop	Thai Restaurant	Gym	Bakery	Bank	Bar	Breakfast Spot	Café	Diner
4	Queen's Park / Ontario Provincial Government	43.662301	M7A	2	Coffee Shop	Sushi Restaurant	Diner	Sandwich Place	Bank	Bar	Café	Park	Italian Restaurant	Gym
7	Don Mills	43.745906	M3B	2	Restaurant	Japanese Restaurant	Coffee Shop	Gym	Sandwich Place	Café	Italian Restaurant	Clothing Store	Thai Restaurant	Fast Food Restaurant
9	Garden District, Ryerson	43.657162	M5B	2	Clothing Store	Coffee Shop	Café	Japanese Restaurant	Italian Restaurant	Pizza Place	Diner	Fast Food Restaurant	Thai Restaurant	Gym
10	Glencairn	43.709577	M6B	2	Pub	Japanese Restaurant	Thai Restaurant	Grocery Store	Bakery	Bank	Bar	Breakfast Spot	Café	Clothing Store

Finally, all the Toronto neighborhoods from this specific cluster are displayed on a map to make it easier for the user to identify their location. He may, for example, pick one neighborhood that is closer to the airport:


Vaughan
Allen
427
Billy Bishop
Toronto
City Airport