A Simple Recommender System for Pseudo Neighborhoods

Kevin Rayco

May 18, 2020

Introduction

- Moving to a new place can sometimes be difficult.
- Let's assume that one way to ease the moving process is to make sure that the new place is similar in terms of the neighborhood.
- This project is a simple recommendation system that gives a ranking of pseudo neighborhoods within a target location based on the similarity of venues in the neighborhood with an origin location.

Data

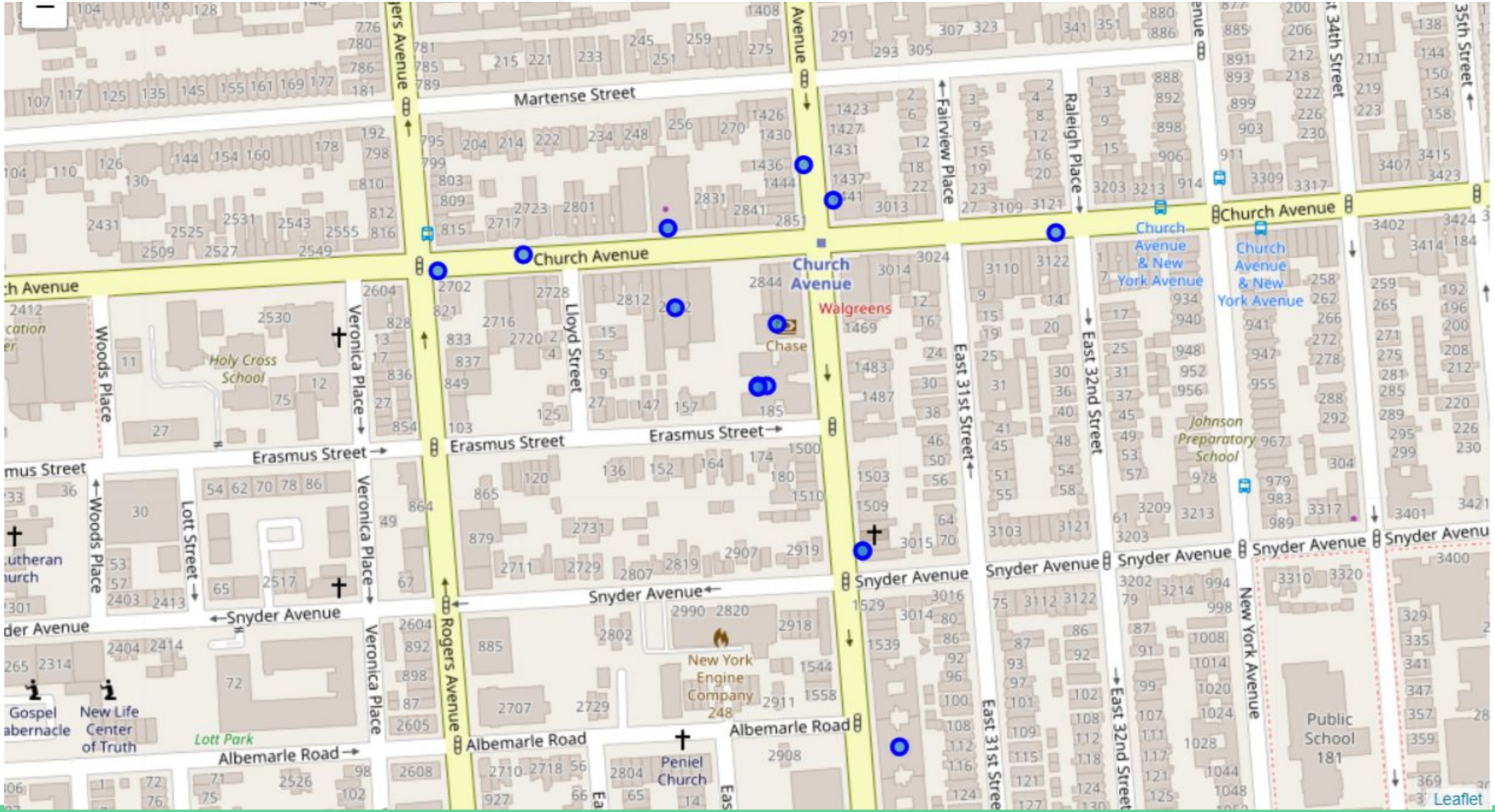
- Data needed:
  - **coordinates** of origin location and target location
  - venue information: **venue ID, name, coordinates, category name, category ID**
- Data sources:
  - **GeoPy** for the coordinates
  - **Foursquare API** for the venue information

Data

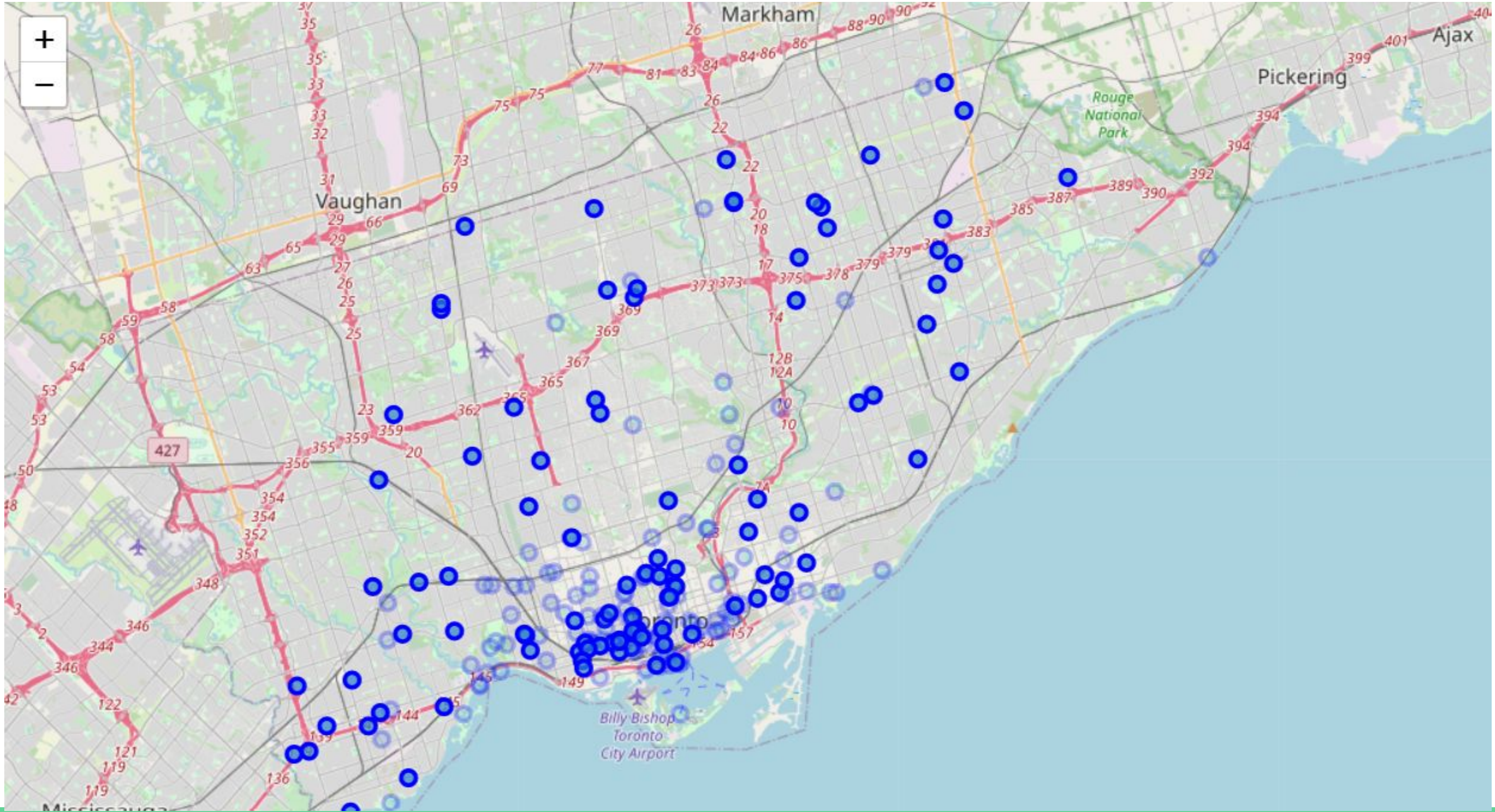
Sample Data:

- Origin Location: Brooklyn, New York
  - Total Venues: 15 closest venues
- Target Location: Toronto, Canada
  - Total Venues: 222 venues
  - Venues of Interest: 98 Venues

Data: Brooklyn, New York



Data: Toronto, Canada



Methodology

- Neighborhood Definition:
    - Define neighborhoods by venues near each other
    - Use k-means clustering to define neighborhoods
        - Set number of clusters to 42 (ratio of coordinates of standard deviations)
        - Features are latitude and longitude
- Neighborhood Profiling:
    - Define factors influencing similarity
    - Use presence of venues of interest, regardless of frequency
- Neighborhood Ranking:
    - Recommend neighborhood by probability
    - Use Gaussian Naive Bayes for multiclass probabilistic classification

Methodology

Neighborhood Definition:

```
Number of clusters: 42
```

	Neighborhood ID	index	Venue ID	Name	Latitude	Longitude	Category ID	Category
216	5	227	4bd444f5462cb71393ebdf07	Francesca Bakery	43.787716	-79.256852	4bf58dd8d48988d16a941735	Bakery
217	11	228	5675fb22498e98ce8baff885	Crown Pizza	43.824307	-79.247393	4bf58dd8d48988d1ca941735	Pizza Place
218	11	229	4aef94c7f964a52060d921e3	Walmart	43.833671	-79.256036	4bf58dd8d48988d10f951735	Pharmacy
219	8	230	4e0b137722713e13018e7117	Tim Hortons / Esso	43.801863	-79.199296	4bf58dd8d48988d148941735	Donut Shop
220	28	231	53669066498eb02e34d50dcd	Yoga Grove - Small Classes. Big Difference.	43.649227	-79.506812	4bf58dd8d48988d102941735	Yoga Studio

Methodology

Neighborhood Profiling:

Target Neighborhood Profiles:

	Neighborhood ID	Playground	Gym / Fitness Center	Bank	Yoga Studio	Furniture / Home Store	Donut Shop	Juice Bar	Caribbean Restaurant	Pharmacy	Bakery	Pizza Place
0	0	0	1	0	1	0	0	0	1	1	1	0
1	1	0	0	0	1	0	0	1	0	0	1	1
2	2	0	0	1	1	0	0	0	0	1	1	0
3	3	0	0	1	0	0	0	0	0	1	1	1
4	4	0	0	0	0	1	0	0	0	1	1	1

Analysis

- Neighborhood profile similarities are calculated by presence of venues of interest
- Used Gaussian Naive Bayes, which is used for multiclass probabilistic classification
- Predicted probability tends to result in extreme values of one and zeros
- Used log-probability to rank neighborhood profile
- Output Range:
  - Negative infinity (zero probability)
  - Up to 0 (equivalent to 100% probability).

Analysis

Log-probability and Mean Coordinates of Top Neighborhoods

	Log Probability	Neighborhood ID	Mean Latitude	Mean Longitude
0	0.000000e+00	0	43.669676	-79.388641
1	-2.075294e+09	1	43.645195	-79.418322
2	-2.075294e+09	2	43.648460	-79.399714
3	-2.075294e+09	3	43.776172	-79.256715
4	-4.150588e+09	4	43.658614	-79.408111

Analysis

Target Neighborhood Profiles sorted by Rank

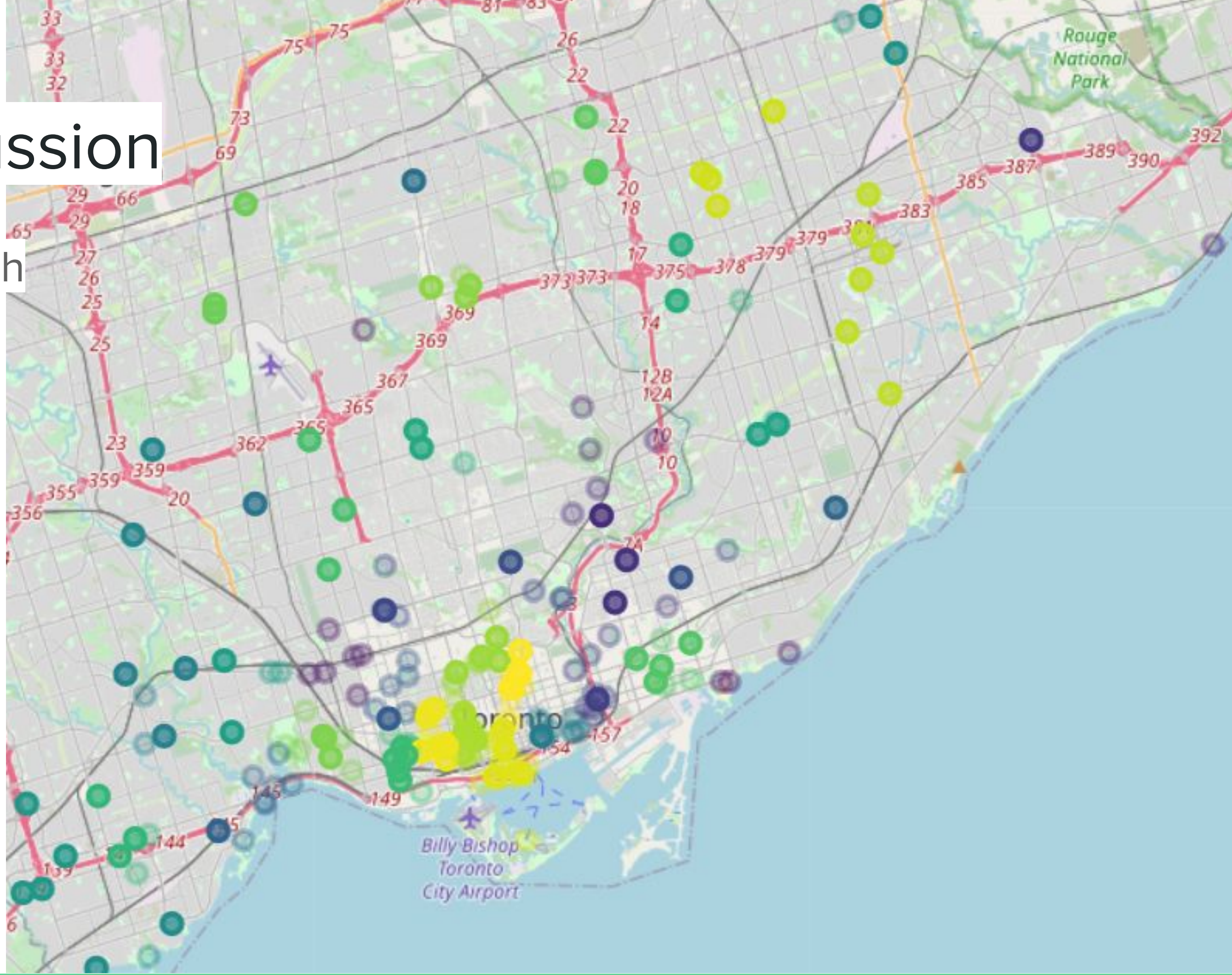
	Neighborhood ID	Playground	Gym / Fitness Center	Bank	Yoga Studio	Furniture / Home Store	Donut Shop	Juice Bar	Caribbean Restaurant	Pharmacy	Bakery	Pizza Place
0	0	0	1	1	1	0	0	0	1	1	1	0
1	1	1	0	0	1	0	0	0	0	1	1	1
2	2	0	0	0	1	0	0	1	0	0	1	1
3	3	1	0	0	1	1	0	0	0	0	1	0
4	4	0	0	0	0	1	0	0	0	1	1	1

Results & Discussion

- Let's view the map to see the neighborhood ranking
- Colored by rank of log probability
- We use viridis colormap to color each neighborhood.
  - Higher ranking neighborhoods tend to be more yellow
  - Lower ranking neighborhoods tend to be more blue
  - Venues that are not of interest are still shown but in lower opacity.

Results & Discussion

We can observe the high ranking neighborhoods from this as desired.



Conclusion and Recommendations

- Interested users and stakeholders can already utilize this system as is and gain desirable recommendations.
- Improvements:
  - Neighborhood clustering by finding global maximum or consistent clustering
  - Visualizing neighborhood boundaries
  - Redefinition of a neighborhood profile
  - Adding more features aside from venues