



THE BATTLE OF NEIGHBORHOODS

IBM Applied Data Science Capstone

Neighborhood Hunting in Toronto, Canada

Autor: Henrique Mandt



Neighborhood Hunting in Toronto, Canada

- Evaluating houses is a deeply personal and complex process, impacted by diverse factors ranging from the physical characteristics and local amenities to politic-economic factors.
- A factor that can affect the decision of evaluating houses is the proximity to the things that matter most to them.
- In this project, Python's data analysis and geospatial analysis packages was used to analyze the whole spectrum of available listings in a market, evaluate and rank properties based on venue category frequency and arrive at a shortlist of neighborhoods in Toronto, similars to Yorkville.



Data acquisition

- Housing data for the city of New York was collected from the New York University Libraries [1]. The New York dataset has a total of 5 boroughs and 306 neighborhoods, as well as the latitude and longitude coordinates of each neighborhood.
- Housing data for the city of Toronto was scraped from Wikipedia [2] that contains a list of postal codes, boroughs and neighborhoods, with a total of 10 boroughs and 217 neighborhoods.

[1] https://geo.nyu.edu/catalog/nyu_2451_34572

[2] https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M



Methodology

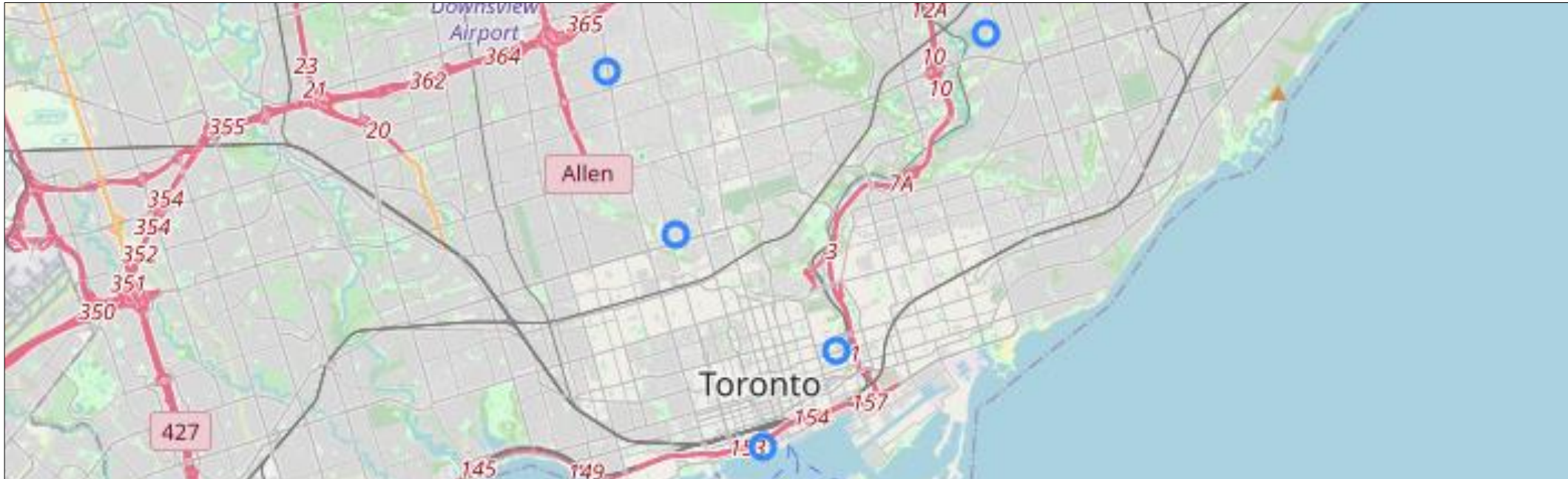
- Firstly, the boroughs and neighborhoods of New York was collected from the New York University Libraries [1]; and postal codes, boroughs and neighborhoods of Toronto was scraped from Wikipedia [2] using Pandas package. The Geocoder package was used to convert address into latitude and longitude.
- Next, Foursquare API was used passing the geographical coordinates to get the top 100 venues that are with a radius of 750 meters. Then, each neighborhood was grouped and the mean of the frequency of each venue category. Then, a shortlist of Toronto neighborhoods was defined by filtering the data with the top 10 most frequent venue category of Yorkville.

Results

- The results from ranking show that Harbourfront, Lawrence Manor, Parkwoods, Regent Park and Victoria Village are the most similar neighborhoods to Yorkville, based on the most frequent venue category.
- The image bellow shows the top 10 venue categories of Yorkville.

	Neighborhood	Coffee Shop	Ice Cream Shop	Pizza Place	Gym	Italian Restaurant	Deli / Bodega	Bagel Shop	Wine Shop	Bar	Mexican Restaurant
0	Yorkville	0.07	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03

MANDT



	Neighborhood	Coffee Shop	Ice Cream Shop	Pizza Place	Gym	Italian Restaurant	Deli / Bodega	Bagel Shop	Wine Shop	Bar	Mexican Restaurant
0	Harbourfront	0.063075	0.005779	0.030619	0.009808	0.013503	0.009807	0.001723	0.000688	0.007991	0.010083
1	Lawrence Manor	0.065583	0.015518	0.031787	0.004697	0.019946	0.002273	0.000000	0.000000	0.007108	0.009367
2	Parkwoods	0.070725	0.018485	0.030372	0.015229	0.008921	0.006578	0.000000	0.000000	0.007644	0.005332
3	Regent Park	0.072782	0.007766	0.064346	0.008766	0.016415	0.003538	0.002019	0.000000	0.007092	0.004394
4	Victoria Village	0.054502	0.008497	0.069860	0.000000	0.008497	0.011364	0.000000	0.000000	0.012529	0.000000



Conclusion

- In this case study, the input data set was scraped from different sources and was spatially enriched with information about access to different venues. It demonstrates how data science can be employed to one aspect of the real estate industry. Buying a home is a personal process, however a lot of decisions are heavily influenced by location. As shown in this study, Python libraries such as Pandas can be used for visualization and statistical analysis, and libraries such as the Foursquare API for Python for spatial analysis. The methods adopted in this study can be applied to any other real estate market to build other recommendation engines.