Moving out: Helping an international student

Daniel Monteiro de Miranda

May 28, 2019

Subject: The idea behind this project is to find out which are the best apartment options for an international student from Rio de Janeiro, based on rent price, location and ways of transportation, in this case subway. Venues around the neighborhood is a key factor too, because it's important to keep the surround of the student similar to his current neighborhood, to make easier to adapt to the new country.

1. Introduction

1.1. The situation that leads to a problem

The objective of this project is to suggest apartment options to a student who lives in Rio de Janeiro and is moving to Toronto in order to study at the Toronto University.
The decision is based on the location of the apartment and rental price. The student needs a apartment close to the university campus or a rent option that offers another way of transportation to the university (in this project, only subway will be considered).

1.2. The problem to be solved

Find the best option for the international student to rent an apartment, considering all the key factors in order to minimize the costs and yet, keep the surround of the student similar to his current neighborhood.
To do that, it's important to start analyzing the current neighborhood of the student, finding out which type of venues are more frequent around the area. To collect data from venues around neighborhoods, we'll be using the Foursquare API.
Next, we must create a list of the most frequent type of venues for each neighborhood in Toronto. After that, we'll be able to decide which neighborhoods are the most similar to the birth city of the student.

2. Data

2.1. Data required to solve the problem

The data required to find out a solution for the best rent option includes:

- Available options for rent (including: rental price, address, latitude, longitude, number of bedrooms)
- List of subway stations of the city (including: station name, address, latitude, longitude)
- Data about venues from the birth city of the international student
- Data about venues in each neighborhood in Toronto

2.2. Data sources

Information about rental data in Toronto can easily be found in Kaggle datasets (https://www.kaggle.com/rajacsp/toronto-apartment-price).
This dataset however, has not been updated since January. To complement the data, I scraped a rental website from Toronto and collected some more data. The website scraped was https://toronto.craigslist.org/search/ap.

The subway data could be found in a blogpost, available in CSV format. Here's the link to access the website: http://scruss.com/blog/2005/12/14/toronto-subway-station-gps-locations/.

Information about venues from Toronto were part of another project, and are already done. In case you're interested in checking it out, here's the link: https://github.com/danielmmir/Coursera_Capstone/blob/master/Scraping_Neigh_data_Toronto.ipynb.

2.3. Data manipulation

After collecting all the necessary data, I started preparing them to be used in my project, since most of the data collected were not ready to be used. Here's some adjustment I did to make sure the data were formatted the right way to be used in my main project:

- Rental dataset

The rental dataset collected from Kaggle had a problem with the price of the rent, since it was a 'object' Type, as we can see in the print taken from the jupyter Notebook.

```
Rental Price      object
```

In order to be able to collect information about the rental prices (average, 25th quartile, 75th quartile, minimum, maximum, etc.) it was necessary transform the type to 'float'.
After that, I had to clean all the rows with at least one empty entry.

- Subway data

Since the data were downloaded separately for each line of the subway, the only thing to do was concatenate the data frames.

3. Methodology

3.1. Steps to reach the solution

In order to reach a solution, it's important to start analyzing the birth city of the student, finding out which type of venues are more frequent around the neighborhood where the student currently lives. To collect data from venues around neighborhoods, we'll be using the Foursquare API.
Next, we must create a list of the most frequent type of venues for each neighborhood in Toronto. After that, we'll be able to decide which neighborhoods are the most similar to the birth city of the student.
After that, the strategy is based on mapping results from the type of venues for each neighborhood in Toronto and the rental data, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed: location near a subway, rental price and similar venues to the birth city of the student. This approach allows quick identification of location, price and feature, thus making the selection very easy.

The processing of these data and its mapping will allow to answer the key questions to make a decision:

- what is the cost of available rental places that meet the demands?
- what is the area of Toronto with best rental pricing that meets criteria established?
- What are the venues of the two best places to live? How the prices compare?
- How venues distribute among Toronto neighborhoods?
- Are there tradeoffs between size and price and location?
- Any other interesting statistical data findings of the real estate and overall data.

4. Data analysis

4.1. Exploring current student neighborhood

Using the foursquare API, I collected venues around the neighborhood of the current student home, which is Barra da Tijuca - Rio de Janeiro, BR. This was an important part because understanding the environment where the student currently lives helps the search for a similar place to live in the new country.

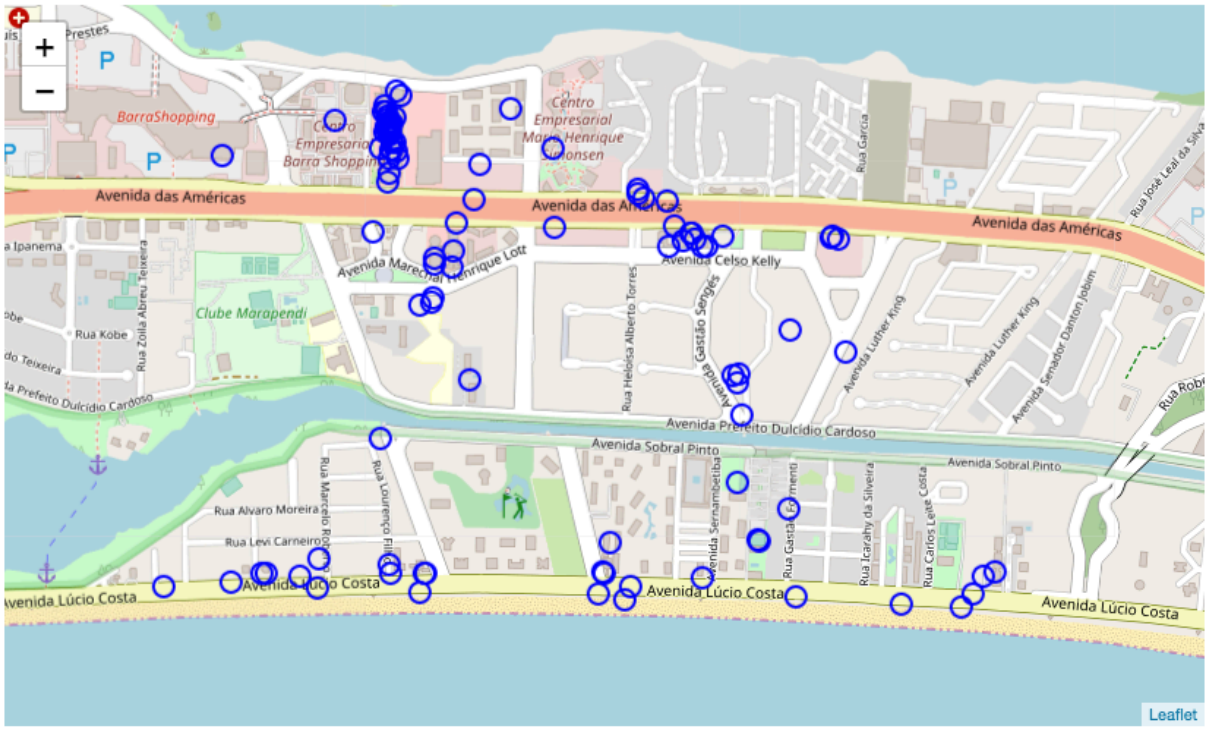
Here are the results from the search:

```
# Venues near current student residence place

SGnearby_venues.head(15)
```

	name	categories	lat	lng
0	Windsor Marapendi	Hotel	-23.009734	-43.351354
1	Deli Golden Green	Deli / Bodega	-23.009934	-43.350338
2	Novotel Barra da Tijuca	Hotel	-23.009942	-43.351312
3	Praia da Barra da Tijuca	Beach	-23.010654	-43.344461
4	Posto 6	Bathing Area	-23.010300	-43.344277
5	Golden Sucos	Sandwich Place	-23.009952	-43.350278
6	Posto 7	Bathing Area	-23.010342	-43.353454
7	Hotel Laghetto Stilo Barra Rio	Hotel	-23.009566	-43.353403
8	Bodytech	Gym / Fitness Center	-23.010077	-43.342172
9	Barrabella Hotel Residência	Hotel	-23.009137	-43.344867
10	Supermercado Zona Sul	Supermarket	-23.009920	-43.345096
11	Gracie Barra Praia	Martial Arts Dojo	-23.006299	-43.351604
12	The Bife	Steakhouse	-23.002715	-43.350476
13	Quiosque do Ceará	Beach Bar	-23.010454	-43.350465
14	Academia de Dança Ramalho's	Dance Studio	-23.009080	-43.340516

In the image above, we can see a list of 15 venues around the current student home. The idea is to understand what type of venues are around the area, and then look for a similar place in Toronto in order to make easier for the student to adapt to a new country.



Above, there's a map showing venues around the current neighborhood of the student.

4.2. Analysis of Toronto neighborhoods venues

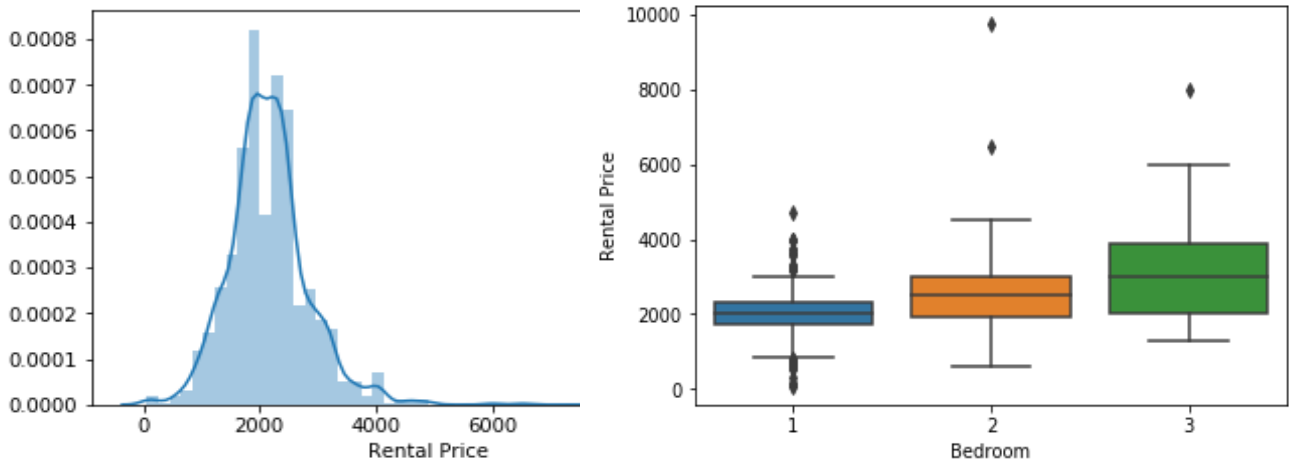
After understanding the surround of the student current neighborhood, it was time to move on to Toronto.

In the image bellow we can see part of the dataset which show us neighborhoods with clusters classification based on the type of venues around each one.

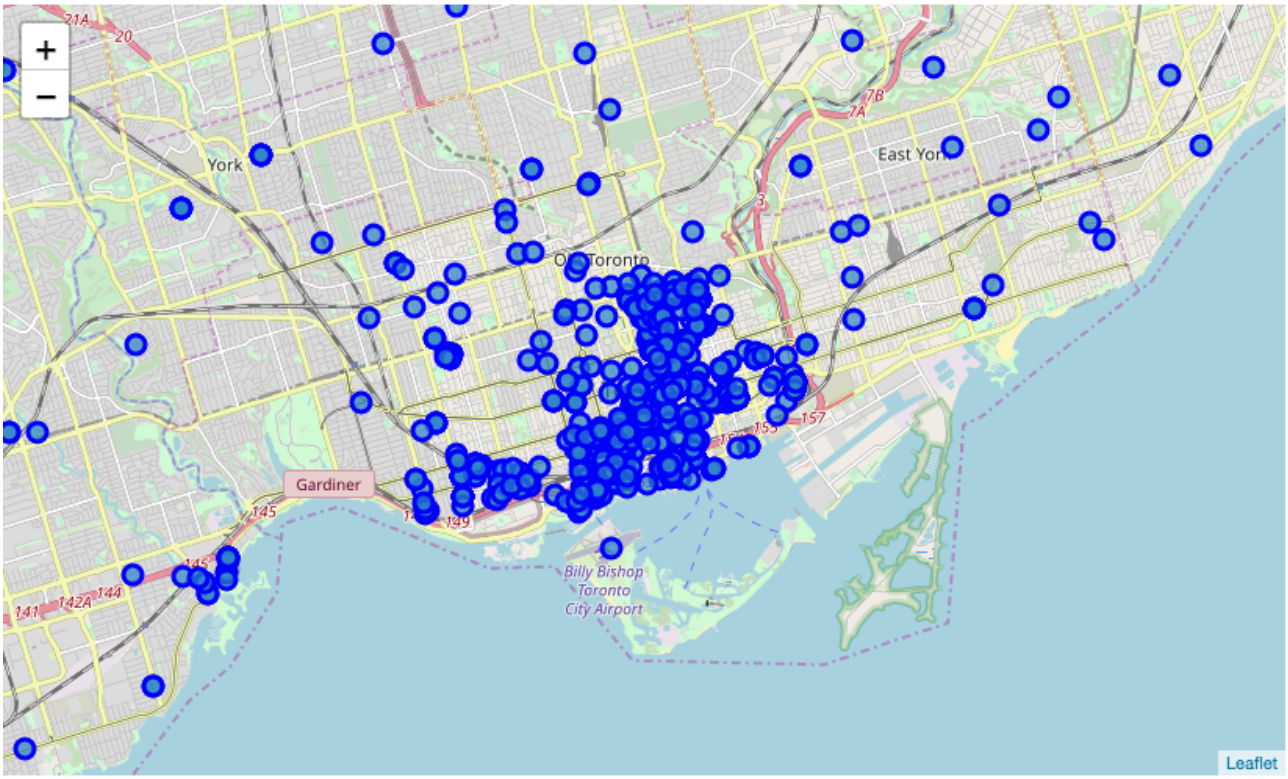
Postcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	0.0	Venue Category_Fast Food Restaurant	Venue Category_Print Shop
M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	0.0	Venue Category_Construction & Landscaping	Venue Category_Bar
M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	0.0	Venue Category_Intersection	Venue Category_Spa
M1G	Scarborough	Woburn	43.770992	-79.216917	0.0	Venue Category_Coffee Shop	Venue Category_Convenience Store
M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0.0	Venue Category_Hakka Restaurant	Venue Category_Bakery

Now that we have information about the venues around the neighborhoods, it's time to focus on the rental option in each area.

4.3. Analysis of rental data



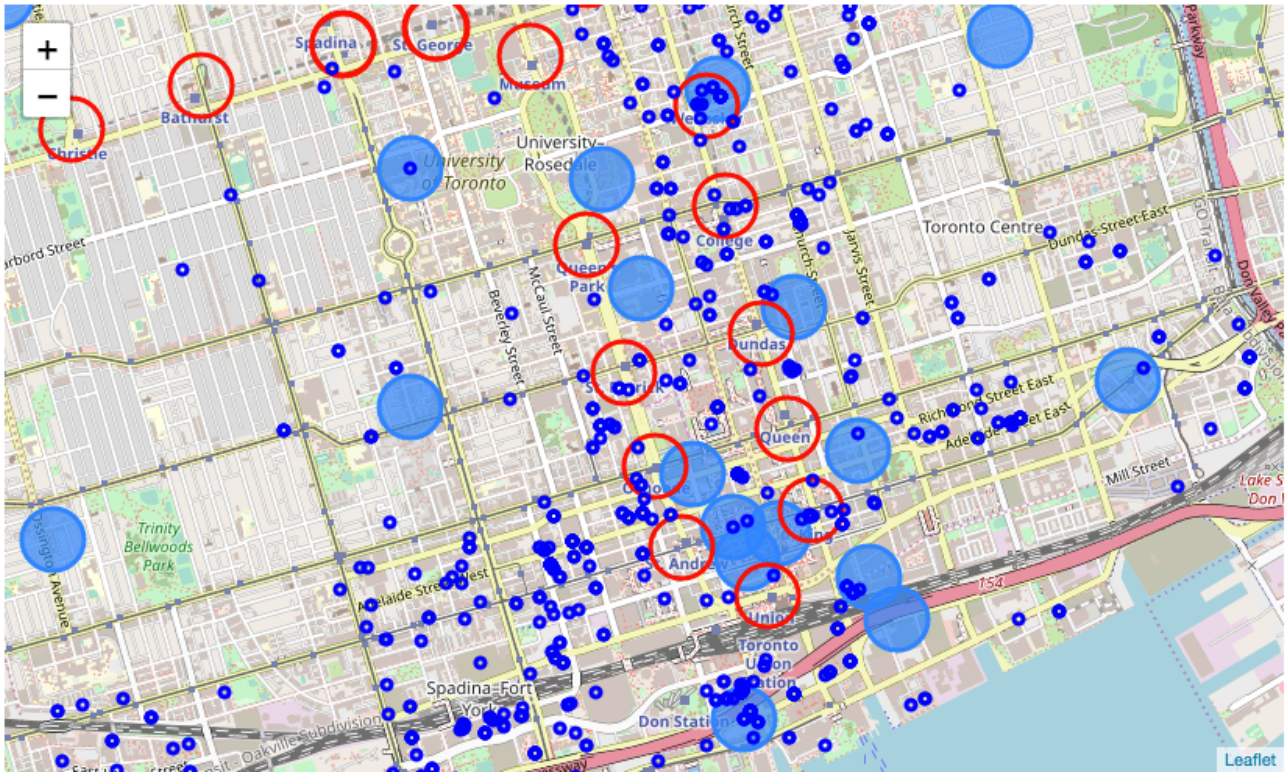
Above, a quick look at the rental prices around Toronto. The average rental price is around CAD 2.000,00, and we can see that, the price of the rent increases according to the number of bedrooms. More bedrooms, a more expensive rent, in general.



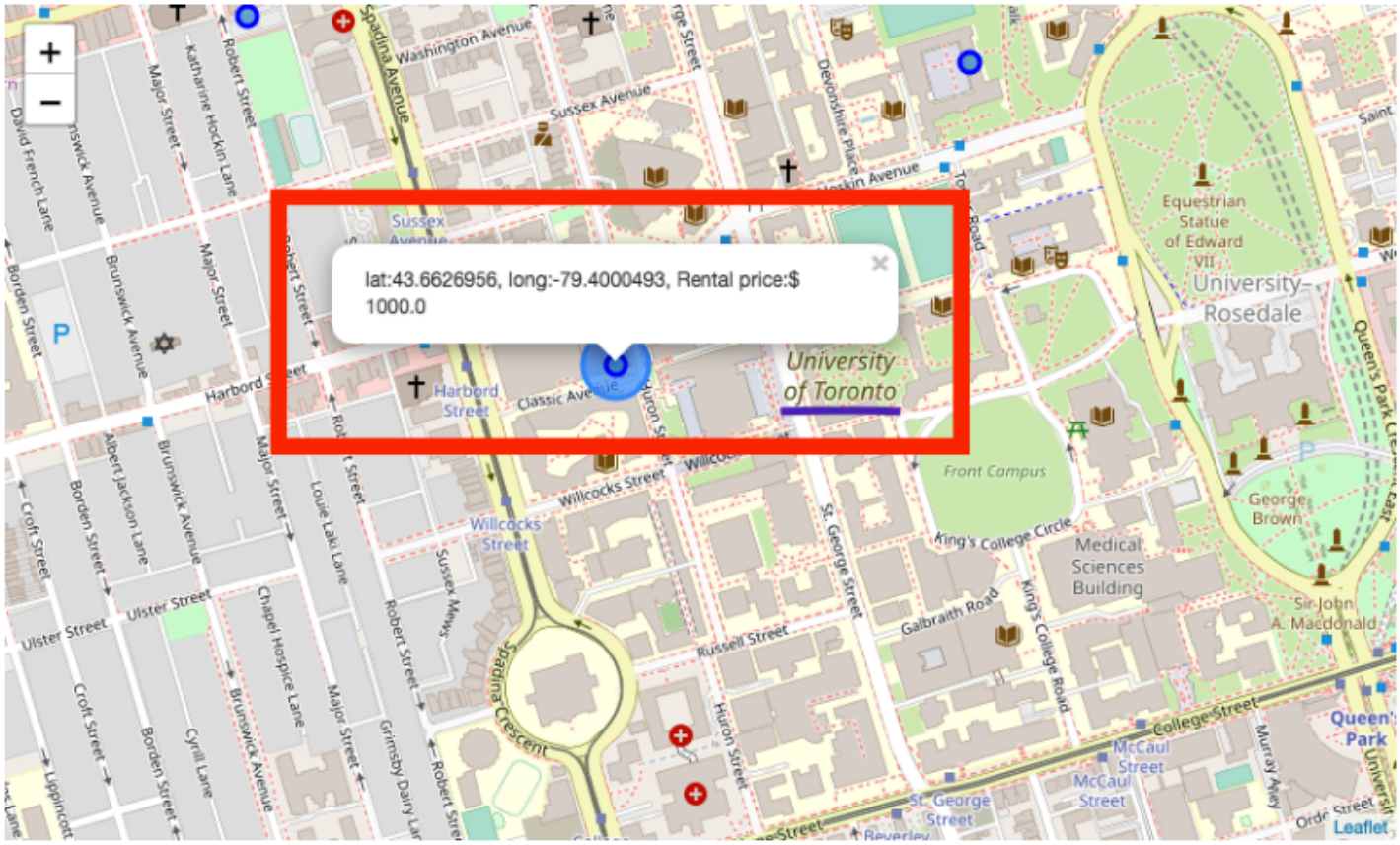
Above, all the apartments available for rent in Toronto. Most of the options are around downtown Toronto.

4.4. Final analysis

For the final analysis, I created a map view with all the factors that were key to find the best option for the student. Bellow, we have the view of the city, including clusters of venues (light blue circles filled with blue), apartment rental options (dark blue circles) and subway stations locations (red circles). Based on this map, 2 apartments were selected as options for the student to move in.



4.4.1. Option 1: Minimum cost




```
rents[rents['Lat']==43.6626956]
```

	Bedroom	Bathroom	Address	Lat	Long	Rental Price
1003	1	1.0	M5S, Canada	43.662696	-79.400049	1000.0

This option counts with 1 bedroom and 1 bathroom. It costs CAD 1000,00 a month, which is less than the half of the average price of rents in Toronto (CAD 2.175).
The apartment is extremely close to the university, but far from a subway station.
The venues around the apartment are similar to the ones around the student's current home.

4.4.2. Option 2: More expensive, but close to the subway


```
rents[(rents['Lat']==43.6655881) & (rents['Rental Price']==1850.0)]
```

	Bedroom	Bathroom	Address	Lat	Long	Rental Price
1026	1	1.0	28 Wellesley St E, Toronto, ON M4Y, Canada	43.665588	-79.383028	1850.0

This option counts with 1 bedroom and 1 bathroom as well. It costs CAD 1850,00 a month, which is less than the average rental price in Toronto (CAD 2.175), but almost two times the price of the option 1.
The apartment is slightly far from the university, but counts with a subway station right next to the apartment.
The venues around the apartment are also similar to the ones around the student's current home.

Based on the conclusions of this project, the student has to decide which of the options suits him better.