

Capstone Project - The Battle of Neighborhoods

1. Introduction

1.1 Background

The trend of studying in Canada is getting more and more popular for International students. From the record, it shows that there are total 642,480 international students studied in Canada 2019[1]. There are 185% increase in international students in Canada from 2010 to 2019[1]. Nearly half of the students studied in Ontario and most of them studied in Toronto.

Toronto is the largest city of Canada with the area of 630.2 square kilometers, and population of 2,731,571. The rich culture diversity, colorful urban life, welcoming immigrants, increasing employment opportunity and high quality education make Toronto attract more and more international students.

1.2 Problems

As mentioned in Background Part, international student choose Toronto is not only for high quality education, also the culture and urban environment. Therefore, choosing a suitable place to live is first step. However, Increasing number of international students may increase the demand of house renting.

Toronto is the first city ranked by population, but only 33rd by area in Canada. For city like Toronto with high population density, renting an apartment with a suitable price and good environment is not easy for students who knows little about the city.

Therefore, I am going to investigate the data of Toronto Apartment Rental prices and venues info around the apartment, cluster the neighbors using k-means, predict the price of the apartment, and give the international students a best suggestion on renting an apartment.

1.3 Target audience

The target audiences of my project are:

1. Students who want to rent an apartment.
2. Rental agency

2. Data

2.1 Data Resources

The data resources I used in the project are listed as following:

1. Toronto Apartment Rental prices:
Kaggle: <https://www.kaggle.com/rajacsp/toronto-apartment-price>
The data consisted of number of bedroom, bathroom, address, latitude, longitude, and price of the apartment. Therefore, we can joint the data of apartment and the data of neighborhood, predict the rental price of the apartment, clustering the neighborhood and recommend a suitable neighborhood for student.
2. Foursquare location data:
According to the geographical coordinates of University and apartments, the venues info around can be found. According to the venues around we can determine whether it is a good place for students to live.
3. Table of postal codes of Canada,
Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
Obtain the postal code and neighborhood of Toronto.
4. Geocoder package or the csv file
Get the geographical coordinates according to postal code of Toronto.

2.2 Data Cleaning

Datasets with clear and useful information is the key of data science process. The first step of the project is download and clean the data.

1. Read and Clean the data of City Toronto:

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031

Table 1 Data of Toronto

2. Read and Clean the data of Apartment rent in City Toronto:

	Bedroom	Bathroom	Den	Address	Latitude	Longitude	Price	PostalCode
0	2	2.0	0	3985 Grand Park Drive, 3985 Grand Park Dr, Mis...	43.581639	-79.648193	2450.0	L5B
1	1	1.0	1	361 Front St W, Toronto, ON M5V 3R5, Canada	43.643051	-79.391643	2150.0	M5V
2	1	1.0	0	89 McGill Street, Toronto, ON, M5B 0B1	43.660605	-79.378635	1950.0	M5B
3	2	2.0	0	10 York Street, Toronto, ON, M5J 0E1	43.641087	-79.381405	2900.0	M5J
4	1	1.0	0	80 St Patrick St, Toronto, ON M5T 2X6, Canada	43.652487	-79.389622	1800.0	M5T

Table 2 Data of Apartment in Toronto

- Merge two table, get Apartment renting Price in City Toronto with neighborhood and borough.

	PostalCode	Borough	Neighbourhood	Latitude_x	Longitude_x	Bedroom	Bathroom	Den	Latitude_y	Longitude_y	Price
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1.235294	1.250000	0.264706	43.654019	-79.363601	1886.345238
1	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	1.459016	1.327869	0.032787	44.068058	-80.263063	1620.235955
2	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	1.166667	1.166667	0.166667	43.651160	-79.375618	1953.571429
3	M4E	East Toronto	The Beaches	43.676357	-79.293031	1.000000	1.000000	0.000000	43.679191	-79.287763	1450.000000
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	1.447368	1.434211	0.157895	43.645255	-79.376301	1854.545455

Table 3 Data of Apartment in Toronto with neighbourhood

3. Methodology

Apartment Rental Price in Toronto

For students, Price is one of the most important concerns when renting an apartment. Therefore, in this part we will analyze the data, predict the price based on status of apartment, and give a list of recommendation of apartment based on price.

The relationship of price with apartment status

The dataset we used in this part is Apartment Price in Toronto(Table 2).The correlation between variables was calculated, and from the table The price of apartment is highly correlated with the number of bedrooms, the number of bathrooms and Den.

In order to have a clear understanding of relationship between an individual variable and the price, scatterplots are shown as below:

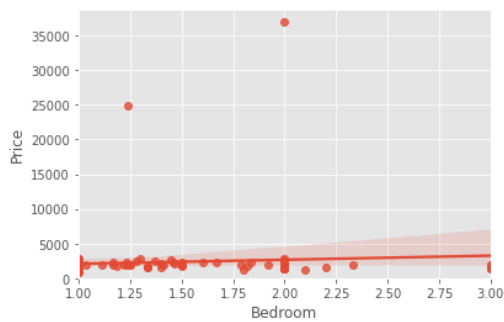


Figure 1 Price VS number of bedrooms

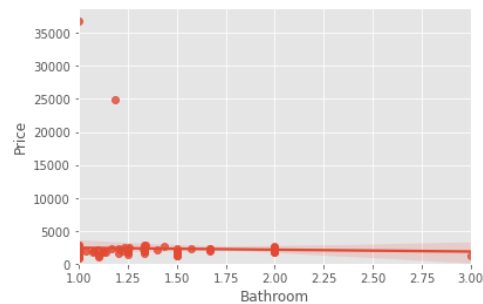


Figure 2 Price VS number of bathroom

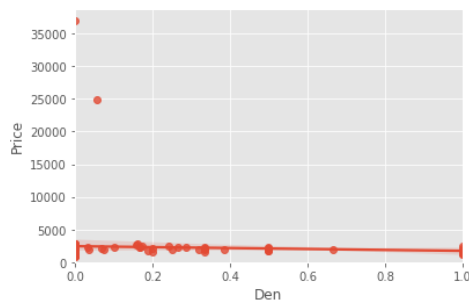


Figure 3 Price VS Dan

Multiple Linear Regression

From the figures above, we used number of bedrooms, number of bathrooms and Den in our model to predict price.

The equation is given by :

$$Y=2353+1347*X1-1470*X2-479*X3$$

According to distribution of the fitted values that result from the model and compare it to the distribution of the actual values, we can visualize the model for multiple Linear Regression.

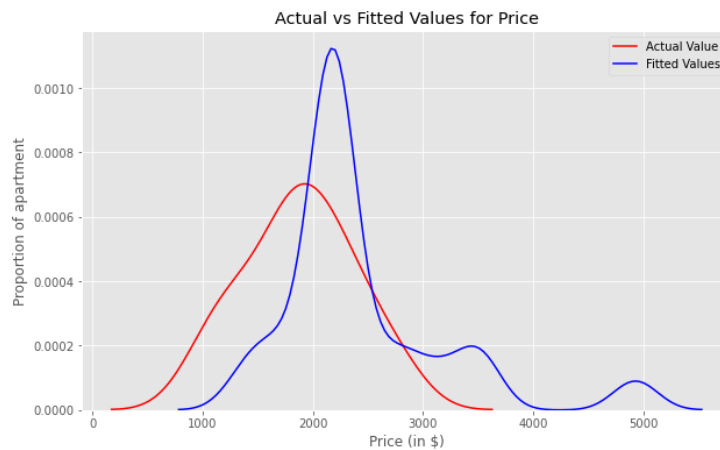


figure 4 Multiple Linear regression

From the Figure, The fitted value is not very closed to the actual values.

Linear Regression

we used number of bedrooms in our model to predict price. The equation is given by :

$$Y=1238+840*X1$$

From the plot, It is shows that the fitted values are close to the actual values compare with Multiple Linear Regression.



figure 5 Linear regression

Relationship of Number of bedroom and Price

From above analyze, the number of bedroom is highly related to the price of renting an apartment. Therefore, We have a further analyze on relationship between number of bedrooms and rental.

The propability of Apartments with different number of bedrooms

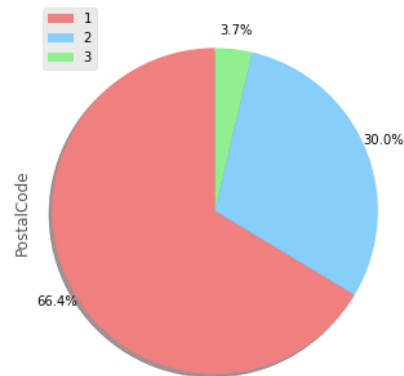


figure 6 Pie chart of probability of apartment with different number of bedroom

From the pie chart, over half of the apartment is one bedroom apartment, 30% of the apartment has two bedrooms. Only 3.7% of the apartment is three bedroom apartment. Therefore, apartment with one or two bedrooms is more easy to find.

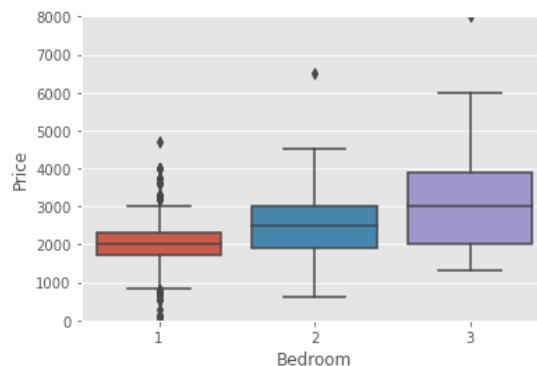


figure 7 box chart of price VS apartment with different number of bedroom

From the box plot, the price of 1 bedroom, 2 bedroom and 3bedroom is shown. 75% of the one bedroom apartment is under 2300, 50% of the one bedroom apartment is under 2000. 75% of the two bedroom apartment is under 3000(1500 per bedroom), 50% of the two bedrooms apartment is under 2500(1000 per bedroom).

Price Vs Different Area In Toronto

The bar chart shows the average price (per room) of apartment in different area of Toronto.

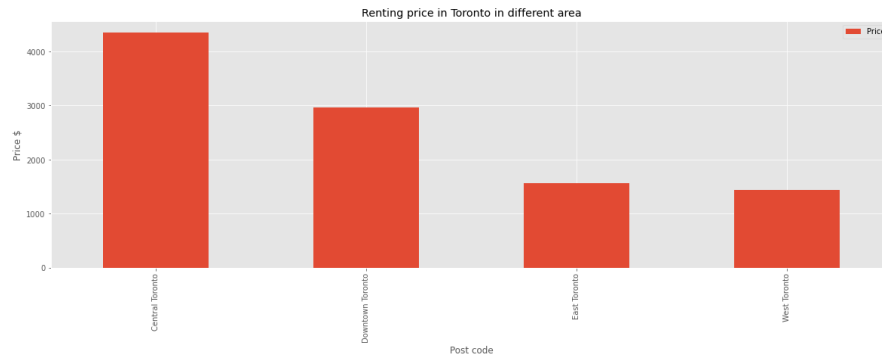


Figure 7 Price Based on different Borough

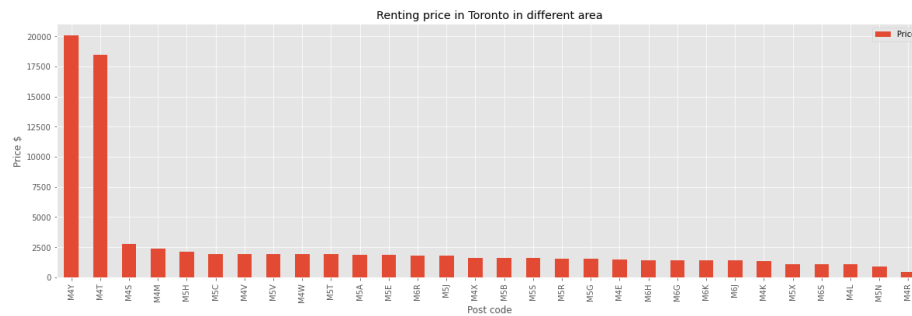


Figure 8 Price Based on different post code

From the chart, The central has the highest apartment rent. However, if we plot the bar chart of price with postcode, we can easily find except M4Y and M4T, the rental price of other areas have little differences.

The neighborhood and venues of apartment in Toronto

For students who want to rent an apartment, money is not the only concerns, they also care a lot about the environment and venues nearby.

By joining data of apartment rent and data of city Toronto, we can obtain the table below:

	Neighbourhood	Latitude_x	Longitude_x	Bedroom	Bathroom	Den	Latitude_y	Longitude_y	Price
0	Berczy Park	43.644771	-79.373306	1.447368	1.434211	0.157895	43.645255	-79.376301	1854.545455
1	Brockton, Parkdale Village, Exhibition Place	43.636847	-79.428191	1.465116	1.093023	0.069767	43.639418	-79.425205	1414.269841
2	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.394420	1.281407	1.233668	0.241206	43.643802	-79.394146	1943.462745
3	Central Bay Street	43.657952	-79.387383	1.222222	1.222222	0.037037	43.656788	-79.385487	1518.181818
4	Christie	43.669542	-79.422564	2.000000	1.333333	0.000000	43.668401	-79.420232	1420.833333
5	Church and Wellesley	43.665860	-79.383160	1.239437	1.183099	0.056338	43.842522	-79.762470	20051.988636

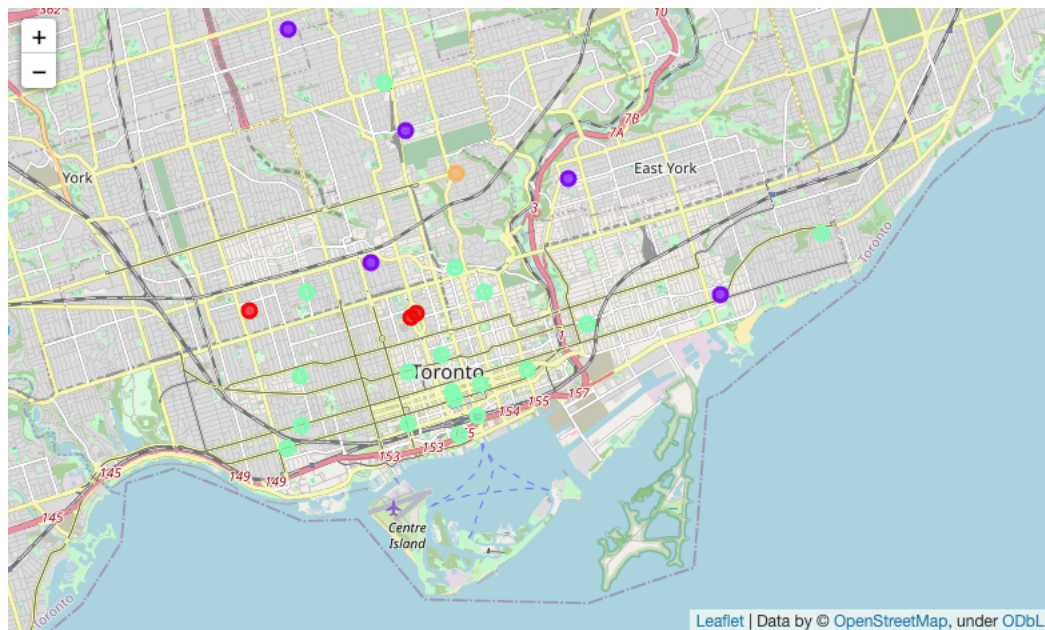
In this part, Foursquare is used to fetch the venues around the apartment. Using K-means cluster the city into 5 clusters, we can easily classified the city based on the characters of the environment around. The table of apartment information and venue around is shown as below:

Price	ClusterLabels	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
1854.545455	3.0	Coffee Shop	Restaurant	Hotel	Italian Restaurant	Sporting Goods Shop	Japanese Restaurant	Beer Bar	Café
1414.269841	3.0	Coffee Shop	Café	Restaurant	Furniture / Home Store	Bar	Supermarket	Bakery	Gift Shop
1943.462745	3.0	Coffee Shop	Restaurant	Hotel	Café	Pizza Place	Gym	Bar	Italian Restaurant
1518.181818	3.0	Coffee Shop	Italian Restaurant	Sandwich Place	Hotel	Restaurant	Sushi Restaurant	Middle Eastern Restaurant	Bubble Tea Shop
1420.833333	3.0	Grocery Store	Café	Playground	Park	Athletics & Sports	Baby Store	Coffee Shop	Candy Store

4. Result

Where to live?

The figure below reveals the classification of apartments based on the venues around them. From the figure, cluster 0 and 3, colored in red and green respectively, are the best two choices for students.



The students who prefer to live in quiet places can choose cluster 0. The venues in cluster 0 are parks, gyms and several restaurants which is suitable for students who enjoy quite and healthy life.

ClusterLabels	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	Park	Bar	Brazilian Restaurant	Gym	Café	Coffee Shop	Distribution Center	Falafel Restaurant	Event Space	Ethiopian Restaurant
0	Park	Office	Sushi Restaurant	Chinese Restaurant	Café	Coffee Shop	Theater	College Theater	Museum	Art Museum
0	Park	Coffee Shop	College Theater	Sandwich Place	Art Museum	Café	Office	Theater	Dog Run	Dive Bar

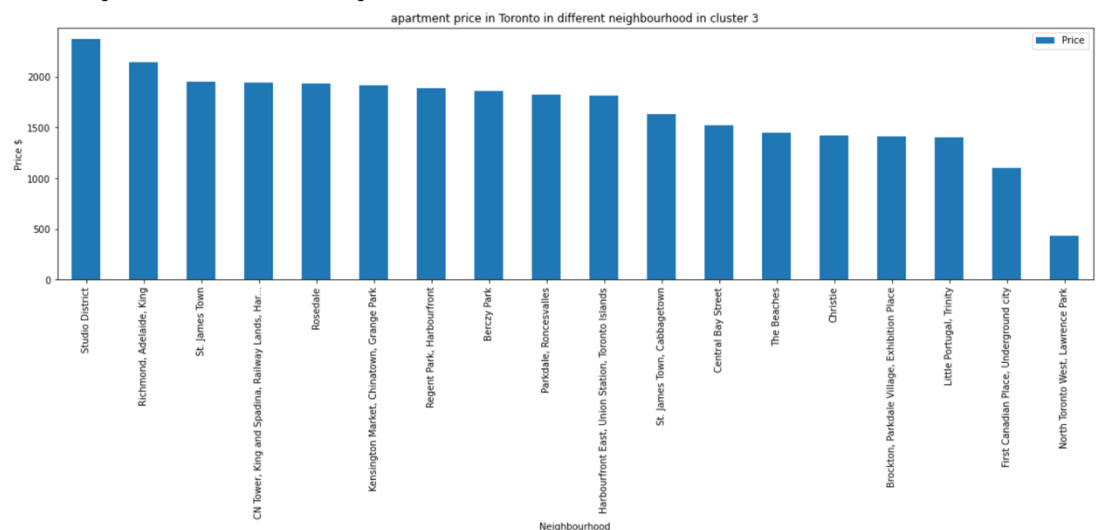
The table below shows the apartment status of cluster 0. From the table, The apartment near University of Toronto is the cheapest, which is really the best choice for students who studied in University of Toronto.

Neighbourhood	Bedroom	Bathroom	Den	Latitude_y	Longitude_y	Price	ClusterLabels
Dufferin, Dovercourt Village	1.187500	1.125000	0.187500	43.664819	-79.434812	1429.894737	0
Summerhill West, Rathnelly, South Hill, Forest...	1.000000	1.000000	0.000000	43.664388	-79.391946	1952.400000	0
University of Toronto, Harbord	1.230769	1.076923	0.384615	43.663759	-79.393333	1581.187500	0

while the students who want to enjoy the urban life can pick cluster 3. The venues around are coffee shops, restaurants, bar and shops, lists as below:

Coffee Shop 17
 Restaurant 17
 Hotel 17
 Italian Restaurant 17
 Sporting Goods Shop 17
 Japanese Restaurant 17
 Beer Bar 17
 Café 17
 Cocktail Bar 17
 Pub 17

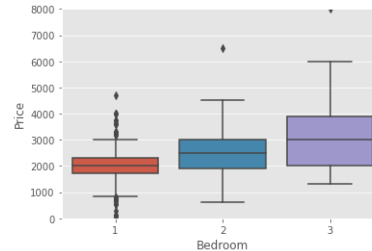
The apartment rental price is listed as below:



What kind of apartment should rent?

From the box chart we can obtain the table below:

Num of bedroom	50%	75%
One	\$2000	\$2300,
Two	\$2500	\$3000
Three	\$3000	\$4000



Therefore, students who have limited budget can share a two/three bedrooms apartment with classmates or friends. For students who have enough budget can choose one bedroom apartment.

5. Conclusion

From the analyze, we can offer a valuable suggestion to students who want to rent an apartment in Toronto based on their habit and budget. In the project.

we used the K-mean train the model and do the cluster. There are total five cluster, and two of them was recommend to student to live. In addition, liner regression is used to predict the price of renting the apartment.

Reference

[1] CBIE uses Immigration, Refugees and Citizenship Canada (IRCC) data.