2. Introduction Section:

Description of the Problem and Background

Scenario:

As an international student from University of Alberta (UA), I deeply felt how hard it was to find a proper one-bedroom rental before moving into a new city. There exist adequate sources of one-bedroom rentals around the university, but because of the lack of information, most of us have to select university dormitories', which the rental fee is much higher.

To help any other incoming international students from UA have more selections for their residences, my project will select Edmonton as the targeted neighborhood. I will try to figure out where are the most common living area in Edmonton and select the most proper region for students.

Because most first-year international students are more likely to live as close to university as possible, but not want to live in university dormitories, I will select the potential available regions from common living area, then compare the rental price to get the conclusion.

Business Problem:

To classify each regions in Edmonton, and identify what regions belong to the core living area.

Select the most cost-effective region from the core living area by analyze the rental price levels.

Interested Audience

The incoming international students for the University of Alberta.

The incoming stuffs work for the University of Alberta.

People who just moving from any other cities, and want to live in Edmonton for a long time.

2. Data Section:

Description of the data and its sources that will be used to solve the problem

Description of the Data:

How the data will be used to solve the problem

The data will be used as follows:

A table with the data about the borough, neighborhood and postal codes: This table would be scraped from Wikipedia and well organized in order to get the specific location data for Edmonton. In this project, I would use X-path package to scrape data.

A table with the data about each borough's coordinates data: This table would use the above table as the input to find out corresponding coordinates data. The table could provide the required information to map each region point. I would use geopy package to catch their coordinates data.

A table with the data about the existing restaurant's types, location, evaluation etc.: This table would contain the information existing restaurant's types, location, evaluation to help me analyze the current situation and figure whether existing low-evaluated restaurants could be replaced by new ones. I would use Foursquare API to get the data.

Two tables contain the rental price for Downtown/Downtown Fringe region and University/Strathcona Place region: Because most of rental websites have set the anti-craping software, it is really hard to directly scrape the rental price data from the website. All the information is manually collect from Rental.ca.

3. Methodology section:

I use the X-path to package to scrape the postcodes data from Wikipedia for better locating each region, clean and save them into a table," df_edm".

|	Postalcode	Borough	Neighborhood
0	T2A	Calgary	Penbrooke MeadowsMarlborough
1	T3A	Calgary	Dalhousie, Edgemont, HamptonsHidden Valley
2	T5A	Edmonton	ClareviewLondonderry
3	T6A	Edmonton	Capilano
4	T2B	Calgary	Forest Lawn, DoverErin Woods
...	...	...	...
67	T5Y	Edmonton	Horse HillLake District
68	T6Y	Edmonton	Southeast Edmonton
69	T7Y	Spruce Grove	Parkland CountyCarvel
70	T2Z	Calgary	Douglas Glen, McKenzie Lake, CopperfieldEast S...
71	T5Z	Edmonton	Lake District

72 rows × 3 columns

Then I use the geocoder package to get the coordinate data and merge them into the previous table. Then I get the table containing all the information about postcode, borough, neighborhood, latitude and longitude.

	Postalcode	Borough	Neighborhood	latitude	longitude
0	T2A	Calgary	Penbrooke MeadowsMarlborough	53.54624	-113.49037
1	T3A	Calgary	Dalhousie, Edgemont, HamptonsHidden Valley	53.54624	-113.49037
2	T5A	Edmonton	ClareviewLondonderry	53.59443	-113.40844
3	T6A	Edmonton	Capilano	53.55094	-113.43550
4	T2B	Calgary	Forest Lawn, DoverErin Woods	53.54624	-113.49037
...	...	...	...	...	...
67	T5Y	Edmonton	Horse HillLake District	53.65196	-113.35747
68	T6Y	Edmonton	Southeast Edmonton	53.36686	-113.60763
69	T7Y	Spruce Grove	Parkland CountyCarvel	53.49069	-114.14729
70	T2Z	Calgary	Douglas Glen, McKenzie Lake, CopperfieldEast S...	53.54624	-113.49037
71	T5Z	Edmonton	Lake District	53.63227	-113.46744

72 rows × 5 columns

After that, I select the Edmonton as out targeted borough and use the folium package to map all the 38 points and find that most points locating in the Loop Highway.
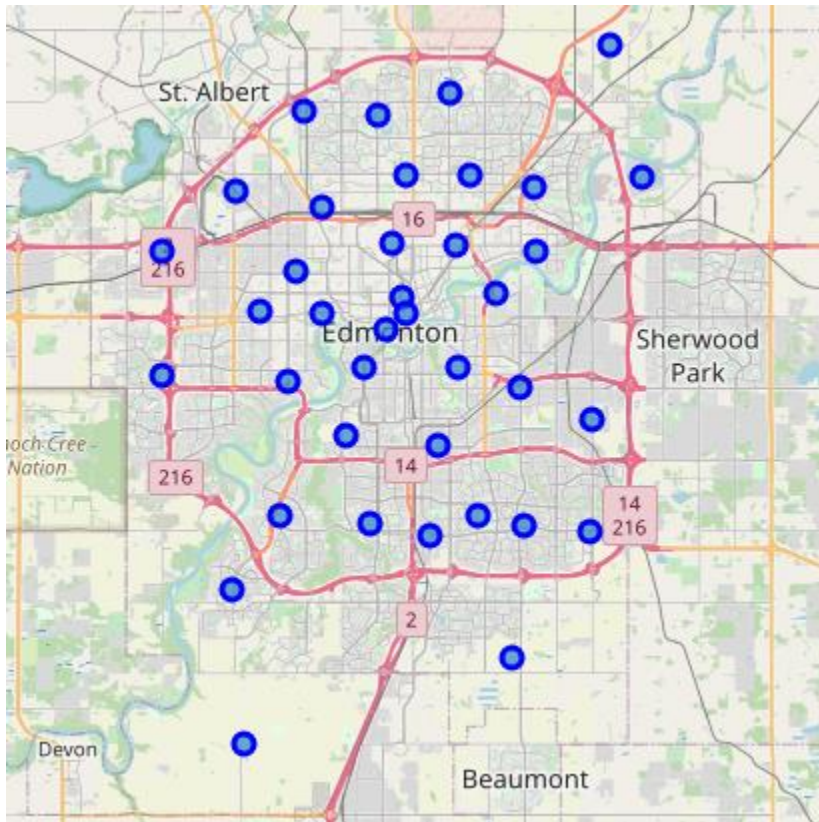
Next, I utilized the Foursquare API to explore each neighborhood, and figure out their 5th most common venues and put the data into different tables.
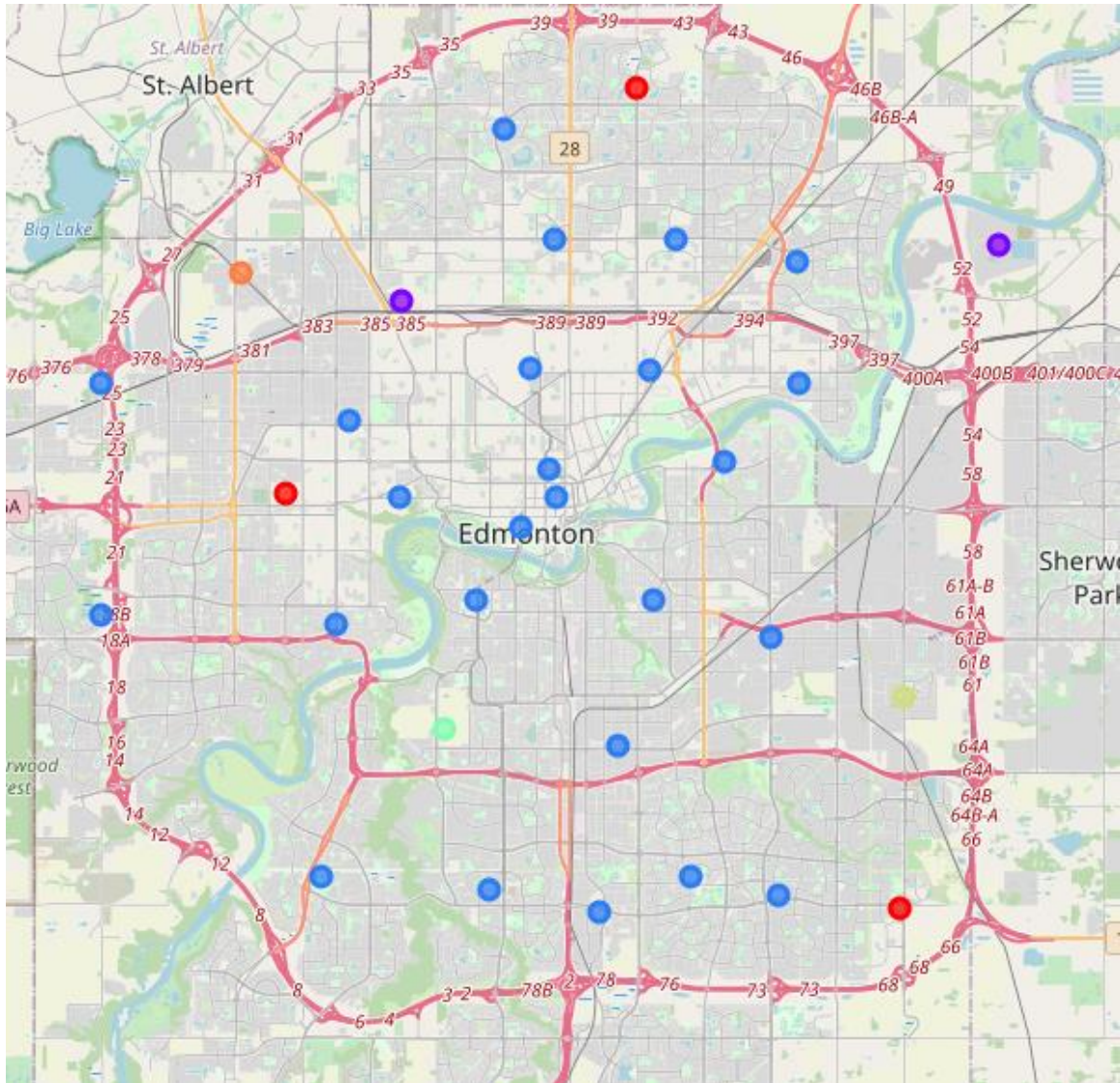
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Beverly	Sandwich Place	Fast Food Restaurant	Pizza Place	Smoke Shop	Bus Station	Burger Joint	Supermarket	Coffee Shop	Caribbean Restaurant	Light Rail Station
1	Bonnie Doon	Fast Food Restaurant	Liquor Store	Irish Pub	Theater	Eastern European Restaurant	Coffee Shop	French Restaurant	Gas Station	Grocery Store	Breakfast Spot
2	Bonnie DoonUniversity	Fast Food Restaurant	Sporting Goods Shop	Hardware Store	Storage Facility	Sandwich Place	Korean Restaurant	Vietnamese Restaurant	Business Service	Office	Noodle House
3	Capilano	Shop & Service	Construction & Landscaping	Dance Studio	Gym / Fitness Center	American Restaurant	Movie Theater	Pharmacy	Pet Store	Park	Paper / Office Supplies Store
4	Capilano, Southeast IndustrialBonnie Doon	Business Service	Indian Restaurant	Restaurant	Electronics Store	Breakfast Spot	American Restaurant	Pharmacy	Pet Store	Park	Paper / Office Supplies Store

After that, it is time to classify them into different clusters. The most common method is K-means, which will cluster all the neighborhood into specific K clusters. While the key point is to find out what is the most proper K.

Elbow Method is the most common method to decide the K when clustering data, and I picked the elbow of the curve,7, as the number of clusters to use.
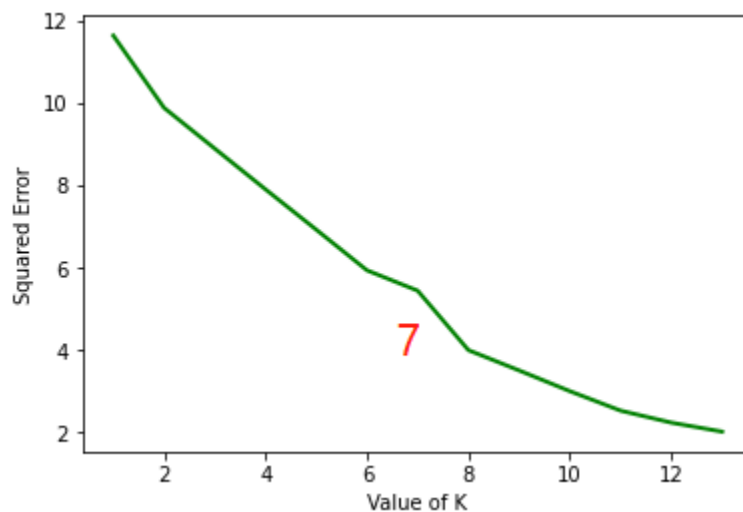
Then, I use K-means method to divide all the regions/ neighborhoods into 7 different groups and may them in different colors.

By analyzing the cluster 1 in purple, I found that the most common venue for each neighborhood is fast food restaurants. That makes senses because these two regions are closed to industrial area, which may be good rental regions for people working in factories but not students.

```
# The far industrial area for purple
Edm_merged.loc[Edm_merged['Cluster Labels'] == 1]
```

|	Postalcode	Borough	Neighborhood	latitude	longitude	Cluster Labels	1st Most Common Venue
16	T5L	Edmonton	Westmount, CalderMistatim	53.585900	-113.554240	1.0	Fast Food Restaurant
26	T6S	Edmonton	Clover Bar	53.598235	-113.334579	1.0	Fast Food Restaurant

By analyzing the cluster 2 in blue, I found that the most common venues for each neighborhood are restaurants, supermarkets and Pharmacy, which are the obvious signals for the core living area, which means I should select the rentals only from neighborhoods in this cluster.

```
#The core living area for blue
#near the restaurant
Edm_merged.loc[Edm_merged['Cluster Labels'] == 2]
```

|	Postalcode	Borough	Neighborhood	latitude	longitude	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
0	T5A	Edmonton	ClareviewLondonderry	53.594430	-113.408440	2.0	American Restaurant	Bus Station	Asian Restaurant	Discount Store	Pharmacy	Arts & Crafts Store
1	T6A	Edmonton	Capilano	53.550940	-113.435500	2.0	Shop & Service	Construction & Landscaping	Dance Studio	Gym / Fitness Center	American Restaurant	Movie Theater
2	T5B	Edmonton	Beverly	53.570765	-113.462861	2.0	Sandwich Place	Fast Food Restaurant	Pizza Place	Smoke Shop	Bus Station	Burger Joint
3	T6B	Edmonton	Capilano, Southeast IndustrialBonnie Doon	53.512370	-113.418280	2.0	Business Service	Indian Restaurant	Restaurant	Electronics Store	Breakfast Spot	American Restaurant
4	T5C	Edmonton	Londonderry	53.599420	-113.453110	2.0	Department Store	Restaurant	Coffee Shop	Sandwich Place	Fast Food Restaurant	American Restaurant

For convenience, I do not want to continue on analyze the other clusters because I have got the information I need.

However, as showing above, I still have to many options in blue. In order to simplify the questions, I assume the incoming students prefer the regions where they just need less than 10 minuses on the way to university. After felting, only two regions come into my view: the left-side one is University/Strathcona Place and the right-side one is Downtown/Downtown Fringe.
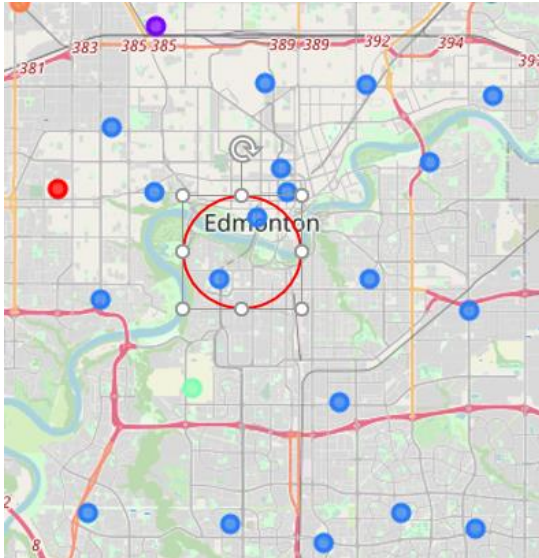
4. Results section:

After manually collecting price data for one-bedroom rentals from the website, Rental.ca, I got the boxplot for comparing the price levels in different regions. I found that the average price in Downtown/Downtown Fringe have higher average prive,1115 CAD/month compared that of 980 CAD/month in University/Strathcona Place. However, the lowest rental price in Downtown/Downtown Fringe is only 730 CAD/month, while the lowest rental price in University/Strathcona Place is about 800 CAD/month.

The rental price for Downtown Fringe and University
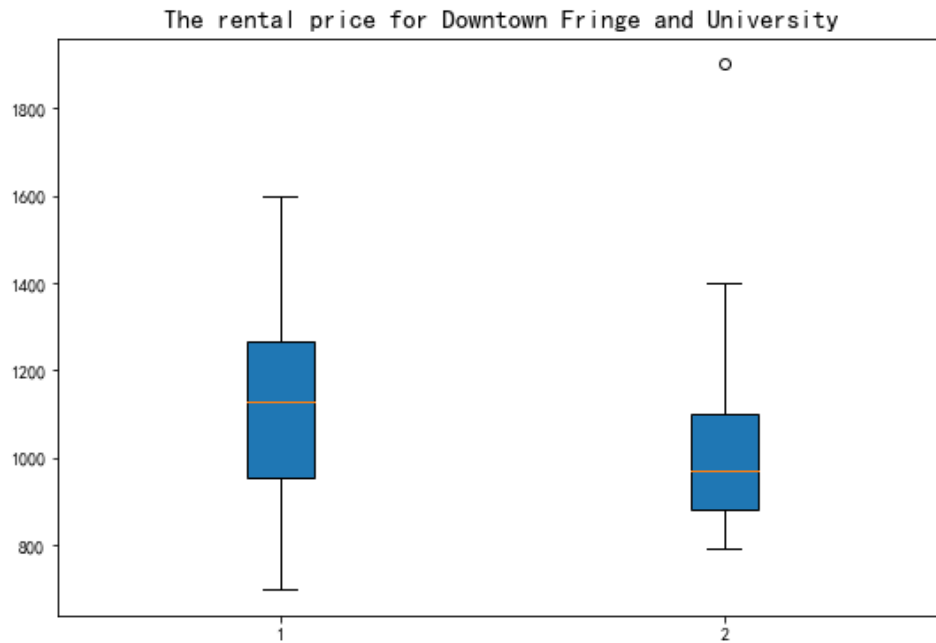
5. Discussion section:

The first thing I have to point out is that unlike other big cities such as Toronto and New York, the cluster points are really simple. 25 blues points cluster in the middle of the Edmonton which represent the core living area. Then, other 2 or 3 points representing industrial area or airport area scatter in the corner of the city. As a result, even though I change the parameter of K into other numbers, the overall clustering situation does change. However, it still brings an issue that the range is too broad to find the proper region for rentals. My recommendation is to select to rent a room/apartment not only in the core living area but also close to your studying or working area.

The second thing I want to say is that most websites do not allow you to directly scrape data, because they treat the price levels for rentals as the trade secret. Thus, I have to collect the data from Rental.ca manually, which is slow and incomplete. If I can get the whole data including the price levels and regions, I can even give suggestions for any people new to Edmonton rather than just international students. What is more, the collected data does not contain the furnished condition of each rental.

All the issues above are caused by the lack of data or information. From this project, I really felt the significance for data. If you do not have adequate and precise data, you even do not know where to rent a room/apartment.

6. Conclusion section:

In conclusion, I strongly recommend incoming international students to search their first rentals in University/Strathcona Place, because compared with the Downtown/Downtown Fringe region, the average price is much lower. Another advantage is that you can live as closed as to the university, which can save your time and cost on the way.

However, if you really have time and limited budget, you can still have chance to find low price rentals in Downtown/Downtown Fringe. While, , and you will not live very comfortable since the age for the house or apartment may be very old and all the facilities may be decay.

7. References:

A table with the data about the borough: "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T

A table with the data about each borough's coordinates data: http://python-visualization.github.io/folium/

A table with the data about the existing restaurant's types, location, evaluation etc.: https://foursquare.com/

Two tables contain the rental price for Downtown/Downtown Fringe region and University/Strathcona Place region: https://rentals.ca/edmonton?baths=1&beds=1&types=apartments&types=apartment&types=studio&types=bachelor&types=basement&types=duplex&types=loft&types=condo&types=houses&types=house&types=town-house&types=multi-unit&types=cabin&types=cottage&types=rooms&types=private-room&types=shared-room&bbox=-114.11694,53.35588,-113.21743,53.68166