Coursera Capstone Project

Coursera IBM Data Science Certification

Kai Tan
27/12/2019

Report Content

- **Introduction Section :**
- 1.1 Discussion of the "backgroung situation" leading to the problem at hand:
- 1.2 Problem to be resolved
- 1.3 Audience for this project.
- **Data Section:**
- 2.1 Data of Current Situation (current residence place)
- 2.2 Data required to resolve the problem
- 2.3 Data sources and data manipulation
- **Methodology section :**
- 3.1 Process steps and strategy to resolve the problem
- 3.2 Data Science Methods, machine learing, mapping tools and exploratory data analysis.
- **Results section**
- Discussion of the results and how they help to take a decision.
- **Discussion section**
- Elaboration and discussion on any observations and/or recommendations for improvement.
- **Conclusion section**
- Desicison taken and Report Conclusion.

1.0 Introduction Section

- **Discussion of the business problem and the audience who would be interested in this project.**
- **1.1 Scenario and Background**
- I am a student currently residing in Downtown, Singapore. I enjoy many ammenities in the neighborhood ,such as cousine restaurants, cafes, shopping mall, food shops and entertainment. Lets imagine that I have been offered a great opportunity to work in Toronto. Although, I am very excited about it, I am a bit stress toward the process to secure a comparable place to live in Toronto. Therefore, I decided to apply the learned skills during the Coursera course to explore ways to make sure my decision is factual and rewarding. Of course, there are alternatives to achieve the answer using available Google and Social media tools, but it rewarding doing it myself with learned tools.
- **1.2 Problem to be resolved:**
- The challenge to resolve is being able to find a rental apartment unit in Toronto that offers similar characteristics and benefits to my current situation. Therefore, in order to set a basis for comparison, I want to find a renta unit subject to the following conditions:
- Apartment with min 1 bedroom with monthly rent not to exceed US$2500/month Unit located within walking distance (<=1.0 mile, 1.6 km) from a subway metro station in Toronto Area with ammenities and venues similar to the ones described for current location (See item 2.1)
- **1.3 Interested Audience**
- I believe this is a relevant project for a person or entity considering moving to Toronto or other major city, since the approach and methodologies used here are applicable in all cases. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions arisen. Lastly, this project is a good practical case toward the development of Data Science skills.

2.0 Data Section

- **Description of the data and its sources that will be used to solve the problem**
- **2.1 Data of Current Situation**
  I am currently reside in the neighborhood of Downtown in Singapore. I use Foursquare to identify the venues around the area of residence which are then shown in the Singapore map shown in methodology and execution in section 3.0 . It serves as a reference for comparison with the desired future location in Toronto.
- **2.2 Data Required to resolve the problem**
  In order to make a good choice of a similar apartment in Toronto, the following data is required: List/Information on neighborhoods form Toronto with their Geodata (latitud and longitud. List/Information about the subway metro stations in Toronto with geodata. Listed apartments for rent in Toronto area with descriptions (how many beds, price, location, address) Venues and ammenities in the Toronto neighborhoods (e.g. top 10) 2.3 sources and manipulation The list of Toronto neighborhoods is worked out during LAb exercise during the course. A csv file was created which will be read in order to create a dataframe and its mapping. The csv file 'toronto_neigh_data.csv' has the following below data structure. The file will be directly read to the Jupiter Notebook for convenience and space savings. The clustering of neighborhoods and mapping will be shown however.

3.0 Methodology

- This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.
- **The analysis and the stragegy:**
- The strategy is based on mapping the above described data in section 2.0, 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 Singapore. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.
- The procesing 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 cost of rent around a mile radius from each subway metro station?
- what is the area of Toronto with best rental pricing that meets criteria established?
- What is the distance from work place (475 yonge street) and the tentative future rental home?
- What are the venues of the two best places to live? How the prices compare?
- How venues distribute among Toronto neighborhoods and around metro stations?
- Are there tradeoffs between size and price and location?
- Any other interesting statistical data findings of the real estate and overall data.

4.0 Execution and Results

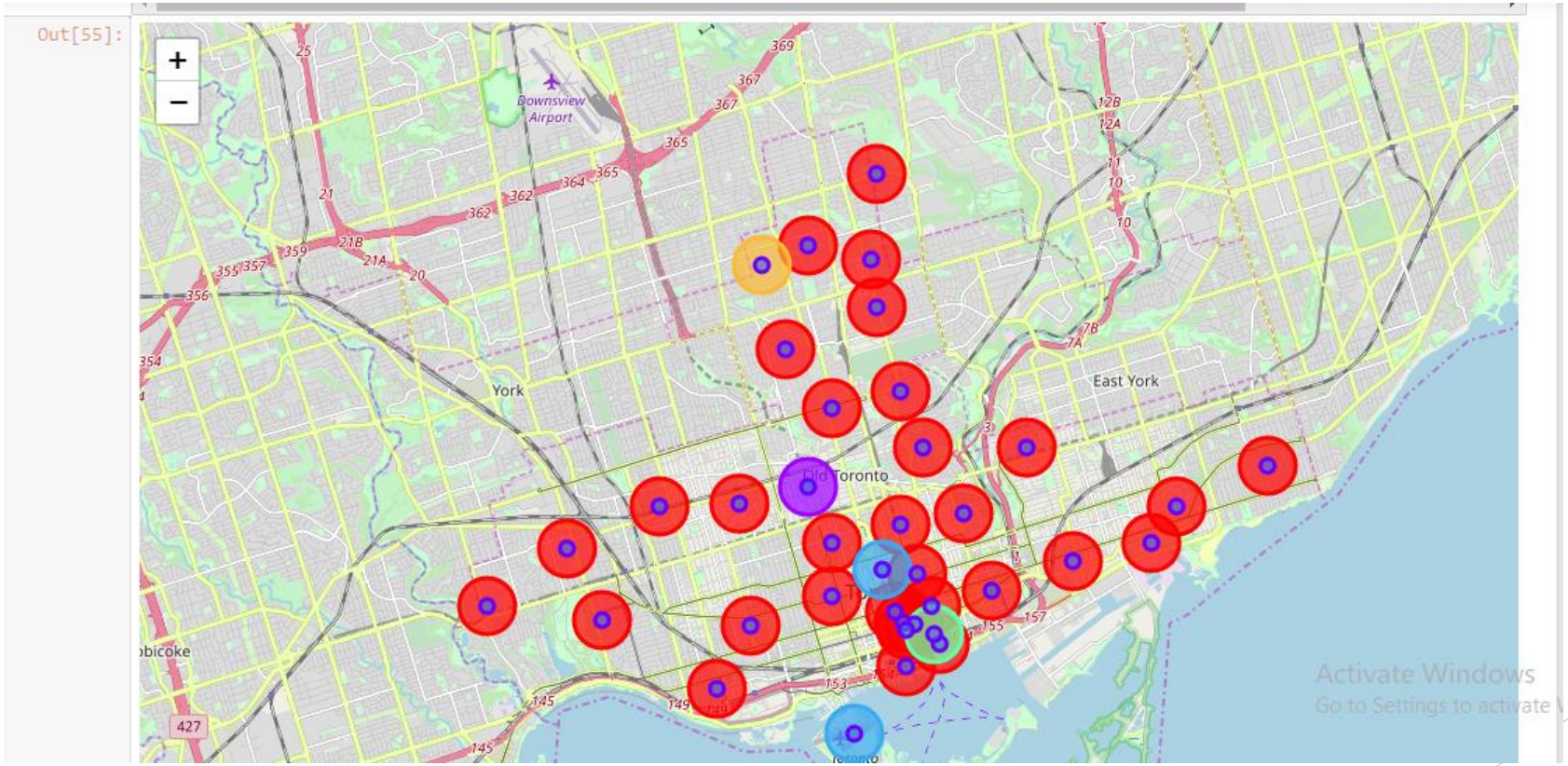
Venues around Neighbourhood

In [46]:
```
# Venues near current Singapore residence place
SGnearby_venues.head(10)
```

Out[46]:

	name	categories	lat	lng
0	PARKROYAL on Pickering	Hotel	1.285754	103.846169
1	Song Fa Bak Kut Teh 松發肉骨茶 (Song Fa Bak Kut Teh)	Soup Place	1.285002	103.844522
2	Orchid Club	Hotel Bar	1.285569	103.846394
3	Wink Hostel	Hostel	1.284132	103.844241
4	Din Tai Fung 鼎泰豐 (Din Tai Fung)	Dumpling Restaurant	1.284975	103.844558
5	28 HongKong Street	Cocktail Bar	1.287737	103.846934
6	Outram Park Fried Kway Teow Mee	Noodle House	1.285446	103.845922
7	LINS Smoodees	Juice Bar	1.283741	103.846945
8	Fill a Pita	Middle Eastern Restaurant	1.284031	103.847624
9	FOC by Nandu Jubany	Spanish Restaurant	1.287460	103.847503

Toronto Map – Neighbourhoods and Cluster of Venues

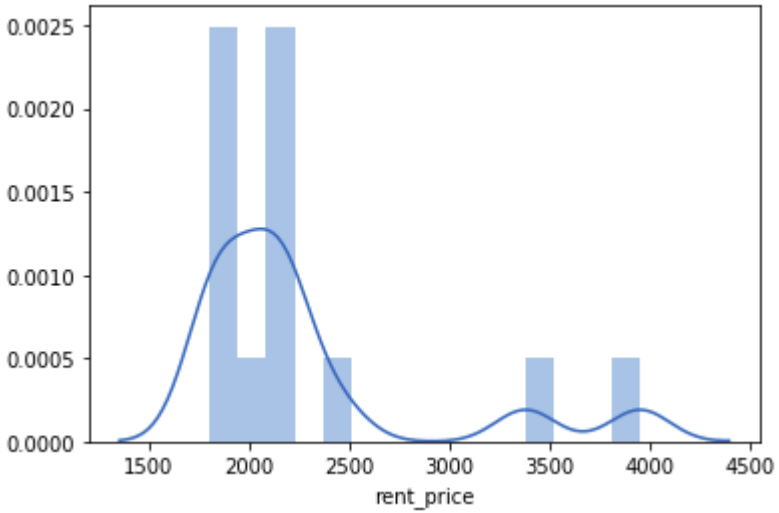


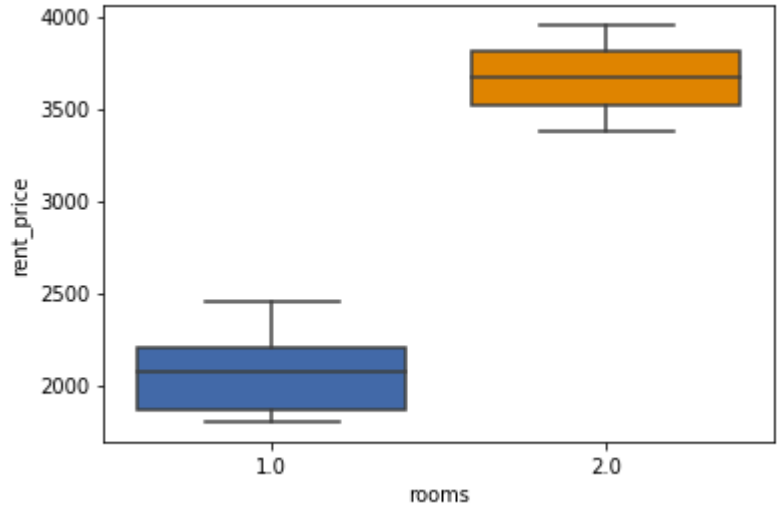
GeoData Toronto apartment for rent

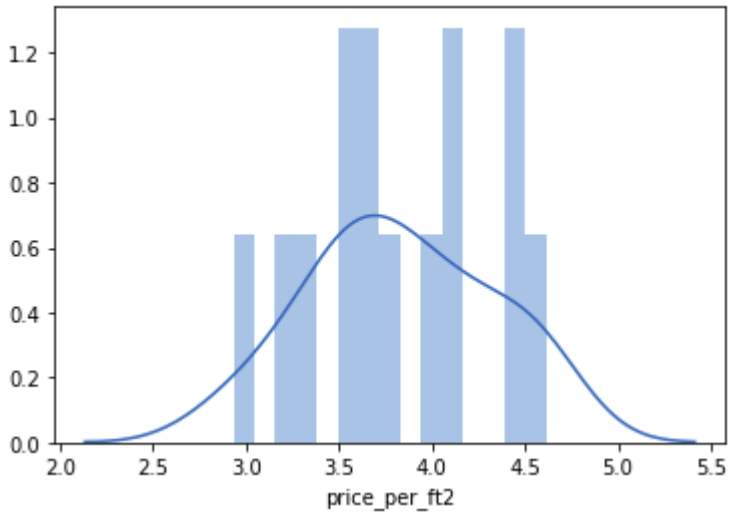
Out[74]:

	address	price_per_ft2	rooms	area_ft2	rent_price	lat	long
0	80 Wellesley Street, East	4.092672	1.0	464.0	1899.0	51.026808	-3.099185
1	80 Wellesley Street, East	3.241590	1.0	654.0	2120.0	51.026808	-3.099185
2	1101 Bay Street, Toronto	3.750000	1.0	500.0	1875.0	43.667761	-79.388314
3	1080 Bay Street, Toronto	4.030612	2.0	980.0	3950.0	43.667127	-79.388498
4	955 Bay Street, Toronto	4.108216	1.0	499.0	2050.0	43.664556	-79.387125
5	33 Bay Street, Toronto	2.933333	1.0	750.0	2200.0	43.642563	-79.377210
6	832 Bay Street, Toronto	4.480652	1.0	491.0	2200.0	43.662272	-79.386536
7	955 Bay Street, Toronto	3.700000	1.0	500.0	1850.0	43.664556	-79.387125
8	763 Bay Street, Toronto	3.712121	1.0	660.0	2450.0	43.659990	-79.385216
9	70 Temperance Street	3.600000	1.0	500.0	1800.0	43.650532	-79.382105
10	887 Bay Street, Toronto	3.380000	2.0	1000.0	3380.0	43.662973	-79.386335
11	955 Bay Street, Toronto	4.500000	1.0	400.0	1800.0	43.664556	-79.387125
12	33 Harbour Square, toronto	3.505843	1.0	599.0	2100.0	43.640683	-79.377000
13	70 Temperance Street, Toronto	4.612159	1.0	477.0	2200.0	43.650532	-79.382105

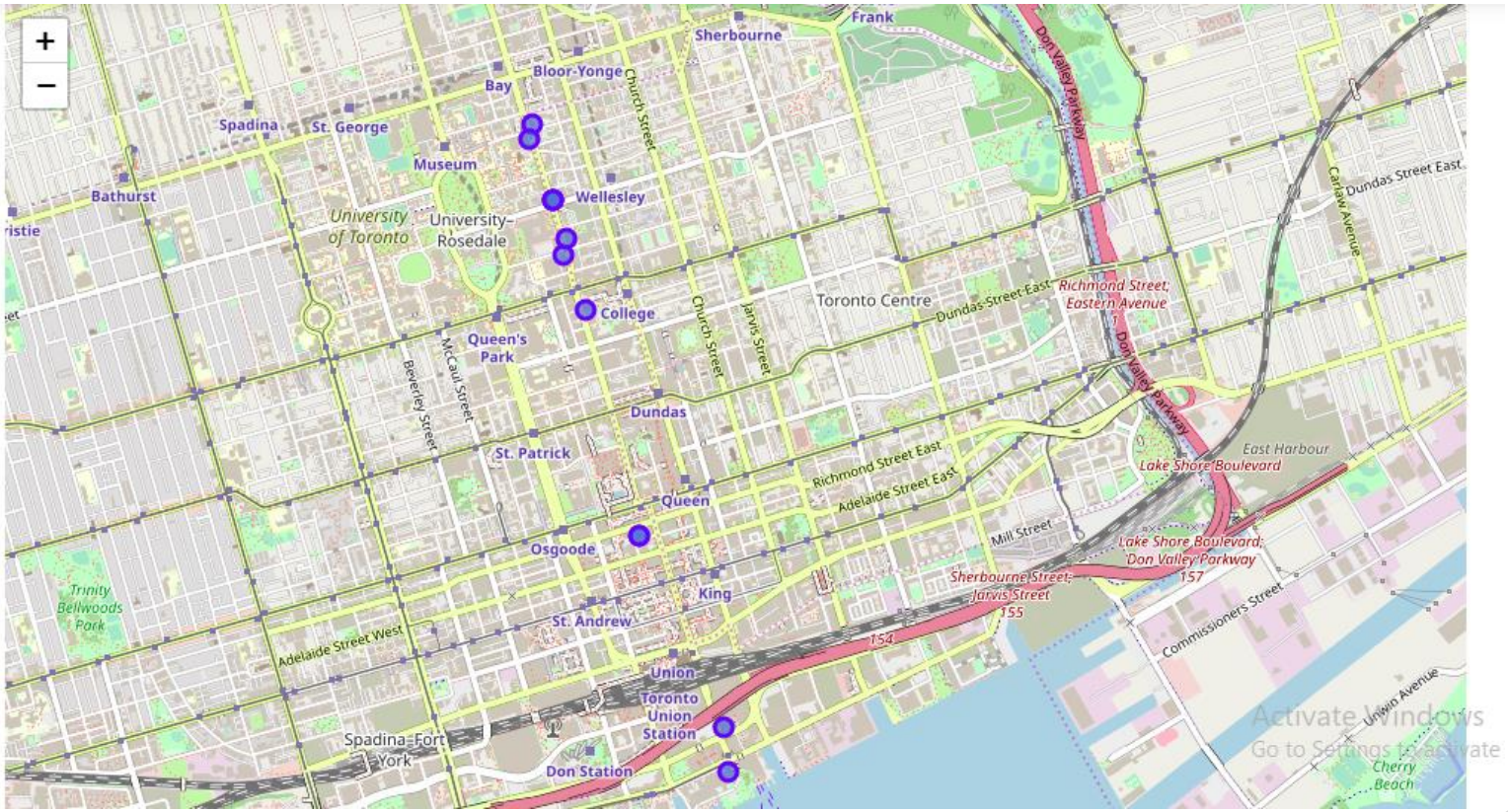
Rental Price Statistics Toronto Apartments







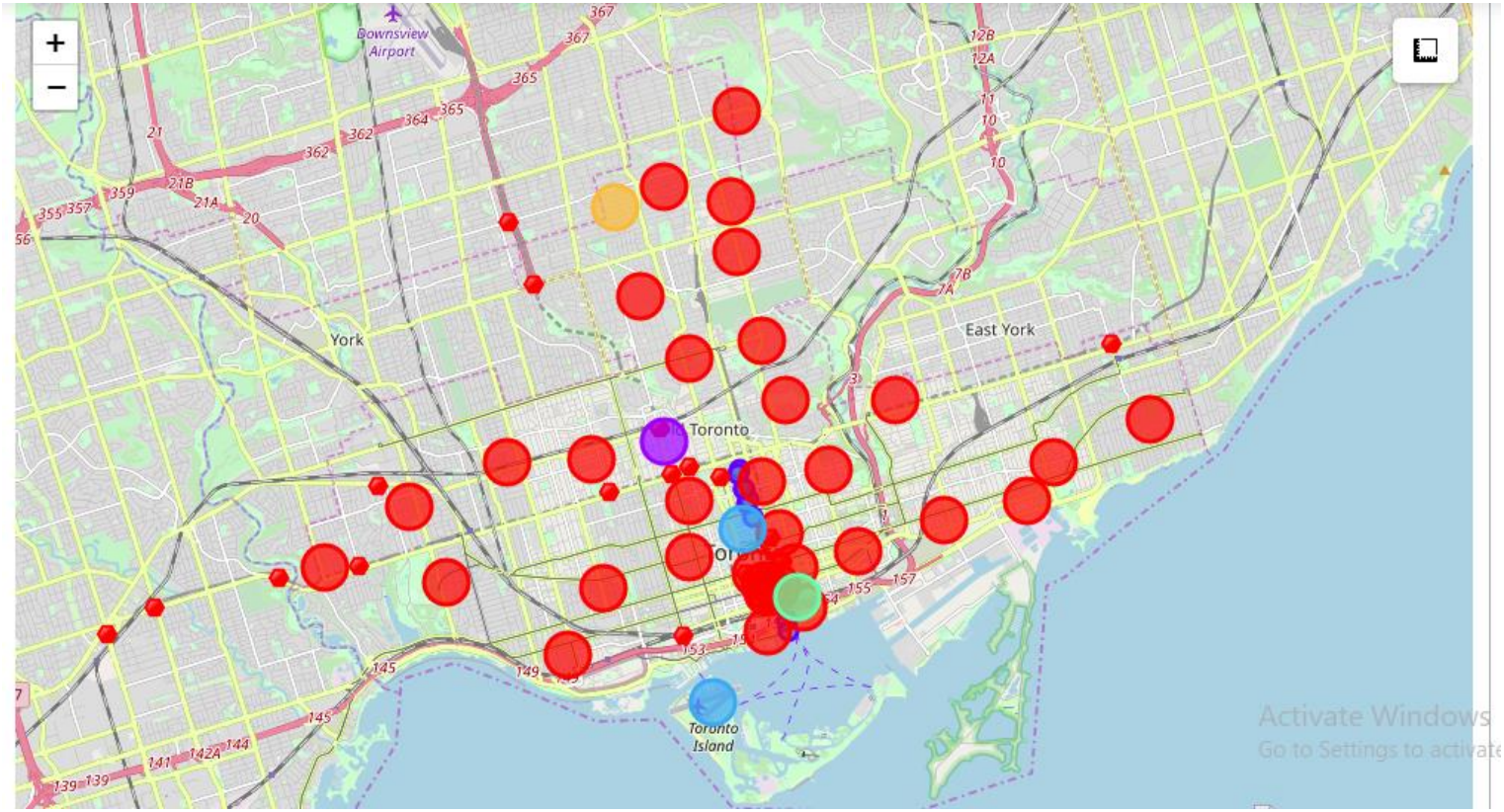
Apartment for rent in Toronto



Toronto subway station data

	sub_station	sub_address	lat	long
0	Museum	75 Queen's Park Toronto, Ontario Canada	43.666727	-79.392753
1	union	65 Front Street West, Toronto, Ontario Canada	43.644903	-79.381836
2	Osgoode	181 Queen Street West, Toronto, Ontario Canada	43.650596	-79.386919
3	Queen	3 Queen Street East Toronto, Ontario Canada	43.652554	-79.378199
4	Dupont	263 Dupont Street Toronto, Ontario Canada	43.674776	-79.406824
5	St. Andrew	173 King Street West, Toronto, Ontario Canada	43.647492	-79.385118
6	King	3 King Street East, Toronto, Ontario Canada	43.649067	-79.377801
7	Kipling	950 Kipling Avenue Toronto, Ontario Canada	43.640273	-79.534708
8	Bathurst	819 Bathurst Street Toronto, Ontario Canada	43.640079	-79.401135
9	Oldmill	2672 Bloor Street West Toronto, Ontario Canada	43.649666	-79.495176
10	Spadina	6 Spadina Road Toronto, Ontario Canada	43.667134	-79.404179
11	Main Street	315 Main Street Toronto, Ontario Canada	43.689089	-79.302061
12	Runnymede	265 Runnymede Road Toronto, Ontario Canada	43.651672	-79.476356
13	Christie	5 Christie Street Toronto, Ontario Canada	43.664111	-79.418405
14	St George	139 St. George Street Toronto, Ontario Canada	43.668315	-79.400014
15	Dundas	3 Dundas Street East Toronto, Ontario Canada	43.656537	-79.381022
16	Glencairn	785 Glencairn Avenue Toronto, Ontario Canada	43.709339	-79.441752
17	High Park Subway Station	35 Quebec Avenue Toronto, Ontario Canada	43.665098	-79.471943
18	Eglinton West	1300 Eglinton Avenue West Toronto, Ontario Canada	43.698903	-79.435998

Consolidated Map (Includes Neighbourhood, clusters of venue, apartments and subway stations)



Apartment Selection

- **Using the "one map" above, I was able to explore all possibilities since the popups provide the information needed for a good decision.**
- Apartment 1 rent cost is US2450 slightly below the US2500 budget. Apartment 1 is located very near to subway station and near to cluster 2 which is in Central Bay Street Neighbourhood. I can walk to work place and use subway to go other places.
- Based on current Singapore venues, I feel that Cluster 2 type of venues is a closer resemblance to my current place. That means that APARTMENT 1 is a good choice for me.

Discussion

- This project has certainly helps me to improve my data science skillset through using varies API, Modules and toolkits in order to accomplish a certain task.
- I have learned how to use Folium, Nominatim and other modules that certainly helps a lot in completing a given task.

Conclusion

- In Conclusion, I feel that this is a very good exercise to take what I have learned in Data Science to apply in real world project. Thank you.
- I have also realized that data science is an ever moving field that requires aspiring data scientist to keep learning and experimenting.