Capstone Project – The Battle of Neighborhoods

Diep, Ly Bao Long

1. Introduction/ Business Problem

1.1 Background

- It is undeniable that Canada is one of the worth-living countries all over the world because of having diverse ethnicity and offering a whole host of great opportunities for immigrants to develop. Significantly, Toronto and Vancouver have gained the reputation for being the two main economic cities in Ontario province and it is evident that these cities have a high standard of living and employment rate, making it very attractive and commonplace to inhabit down the road. Since Toronto and Vancouver are famous metropolitan areas, there has been a bunch of comprehensive analyses and surveys about them, and they will not be on the agenda of this project. As regards the ideal living areas, I will pay my attention to the other cities in Canada, particularly London city which is part of Ontario province and one of the potential candidate cities. Dissimilarly, London is a city with a moderate area of about 420.6 km2 in the southern part of Ontario. More intriguing, London is not densely populated, and it is estimated that the total population is approximately 404,699 in 2017 compared to 2.93 million in Toronto in 2017, which will contribute to reducing the severe competition in job markets. Moreover, the economic resources and wealth in London primarily derive from medical research, education, insurance, and information technology which demonstrates that London also has good social welfare to attract newcomers, and sufficiently creates a whole host of employment opportunities with prospective promotions to nourish and foster an abundant workforce. Thanks to these good social and living factors, London can be considered to meet several standard living conditions at least. The overview about London has been listed out, but another issue which is also far more important to deal with is which neighborhood in London is the best to settle down and initiate a business.

1.2 Problem

- **Problem definition:** Accommodation and start-up problem in London, Ontario.
- **Circumstance:** In this capstone, as an upcoming data scientist, I will reach out to stakeholders who are looking for a living area which is convenient and well-located with good public amenities and service in London, Ontario. Specifically, we are going to explore the neighborhoods and specify the number of ideal residential areas in London. What's more, the stakeholders also have no clue for what kind of effective business patterns they should initiate, so they need some handy suggestions for some potential businesses in the recommended areas.
- **Assumption:** It is obvious that there are a number of influential factors such as the stakeholder's affordability for the cost of living and the price of the real estate in the desired areas. Since this capstone primarily focus on using Foursquare to discover the most fantastic living area in this capstone, it will be assumed that these factors are in the available budget of the stakeholders.
- A data scientist will manipulate the power of data to generate the most feasible and promising neighborhoods based on the listed above criteria. It will be expected that the upsides and downsides will be also comprehensively laid out so that the best deliverables can be used to help the stakeholders make their final decision.

**** Agenda ****

- What neighborhood is the best to settle down in London, Ontario?
- How far is it from the living area to its surrounded venues?
- What kind of potential business pattern should be recommended?

1.3 Target Audience

- This project will significantly draw the attention of the potential stakeholders who have a desire to settle down in a second homeland and run their own business in a residential area with good living

conditions. To be more precise, the stakeholders can be tenants, new immigrants, real estate companies, international settlers, home investors, etc.

2. Data acquisition and cleaning

2.1 Data Collection

- Based on the defined circumstance, there will be a number of factors that will have impacts on the decisions and analyses:

- The number of neighborhoods that need to be taken a look at in London, Ontario.
- The distance from the living areas to the other venues within the scope of neighborhood.
- The number of available business patterns within the neighborhood.

- As listed above, the following data sources will be needed to generate the required information:

- The location and coordinates of each neighborhood in London, Ontario will be scraped from the webpage by using **Pandas function/ BeautifulSoup**.
- The number of venues and their categories within each neighborhood are extracted by using **Foursquare API**.
- **The extracted venue categories** can be used as a **separated dataset** to build the **recommender system** for the suggestion of business patterns.

2.2 Data Refinement and Formatting.

- Since the required data scraping from the available dataset on webpage is originally raw, some data transformation and refinement will have to be carried out for the easily visual analysis and exploration.
- The dataset used in this analysis should be in tabular form and consist of a number of following columns: **Postal Code, Borough, Neighborhood, Latitude, Longitude.**

Sample table:

	Postal Code	**Borough**	**Neighborhood**	**Latitude**	**Longitude**
1					
n					

- Each row contains different data values corresponding to each column.

Illustration for the original dataset

	Place	Code	Country	Admin1	Admin2	Admin3
1	London (West Huron Heights / Carling) 43.012/-81.231	N5Y	Canada	Ontario	London	
2	London (Glen Cairn) 42.966/-81.205	N5Z	Canada	Ontario	London	
3	London (East Tempo) 42.872/-81.247	N6L	Canada	Ontario	London	
4	London East (SW Argyle / Hamilton Road) 42.986/-81.182	N5W	Canada	Ontario	London	
5	London West (Central Hyde Park / Oakridge) 42.991/-81.34	N6H	Canada	Ontario	London	
6	London (Southcrest / East Westmount / West Highland) 42.955/-81.273	N6J	Canada	Ontario	London	
7	London (Jackson / Old Victoria / Bradley / North Highbury) 42.963/-81.139	N6M	Canada	Ontario	London	
8	London (YXU / North and East Argyle / East Huron Heights) 43.023/-81.164	N5V	Canada	Ontario	London	
9	London (Fanshawe / Stoneybrook / Stoney Creek / Uplands / East Masonville) 43.044/-81.239	N5X	Canada	Ontario	London	
10	London South (East Highland / North White Oaks / North Westminster) 42.958/-81.238	N6C	Canada	Ontario	London	
11	London (South White Oaks / Central Westminster / East Longwoods / West Brockley) 42.918/-81.224	N6E	Canada	Ontario	London	
12	London (Sunningdale / West Masonville / Medway / NE Hyde Park / East Fox Hollow) 43.015/-81.305	N6G	Canada	Ontario	London	
13	London (Riverbend / Woodhull / North Sharon Creek / Byron / West Westmount) 42.954/-81.342	N6K	Canada	Ontario	London	
14	London (South Highbury / Glanworth / East Brockley / SE Westminster) 42.9/-81.159	N6N	Canada	Ontario	London	
15	London (Talbot / Lambeth / West Tempo / South Sharon Creek) 42.891/-81.324	N6P	Canada	Ontario	London	
16	London Central 42.984/-81.239	N6B	Canada	Ontario	London	
17	London North (UWO) 42.998/-81.256	N6A	Canada	Ontario	London	

- As shown by the data table above, each individual row is filled with distinct data, but it is not in the desired form which is mentioned above. More significantly, the “Place” column seems to be messy because one labelled row contains not only the combination of borough and its neighborhood enclosed in parenthesis but also its latitude and longitude right in the unlabelled row below. The “Code” columns will be converted into the “Postal Code” column and the rest of the columns are not requisite and will be eliminated.
- **BeautifulSoup** library will be the efficient and easy-to-use tool in scraping this dataset. The rows containing the combination of borough and neighborhood will be separately derived and converted into two distinct columns (Borough and Neighborhood). Similarly, the coordinate values in the unlabelled rows will be also separated into two numeric columns (Latitude, Longitude).
- Finishing filtering out the above dataset, we have a complete refined data table, and it is ready to be used in the next stage of analytic process.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	N5Y	London	West Huron Heights / Carling	43.012	-81.231
1	N5Z	London	Glen Cairn	42.966	-81.205
2	N6L	London	East Tempo	42.872	-81.247
3	N5W	London East	SW Argyle / Hamilton Road	42.986	-81.182
4	N6H	London West	Central Hyde Park / Oakridge	42.991	-81.340
5	N6J	London	Southcrest / East Westmount / West Highland	42.955	-81.273
6	N6M	London	Jackson / Old Victoria / Bradley / North Highbury	42.963	-81.139
7	N5V	London	YXU / North and East Argyle / East Huron Heights	43.023	-81.164
8	N5X	London	Fanshawe / Stoneybrook / Stoney Creek / Upland...	43.044	-81.239
9	N6C	London South	East Highland / North White Oaks / North Westm...	42.958	-81.238
10	N6E	London	South White Oaks / Central Westminster / East ...	42.918	-81.224
11	N6G	London	Sunningdale / West Masonville / Medway / NE Hy...	43.015	-81.305
12	N6K	London	Riverbend / Woodhull / North Sharon Creek / By...	42.954	-81.342
13	N6N	London	South Highbury / Glanworth / East Brockley / S...	42.900	-81.159
14	N6P	London	Talbot / Lambeth / West Tempo / South Sharon C...	42.891	-81.324
15	N6B	London Central	London Central	42.984	-81.239
16	N6A	London North	UWO	42.998	-81.256

****NOTE**:** There are **17 obtained neighborhoods** in London, Ontario in total.

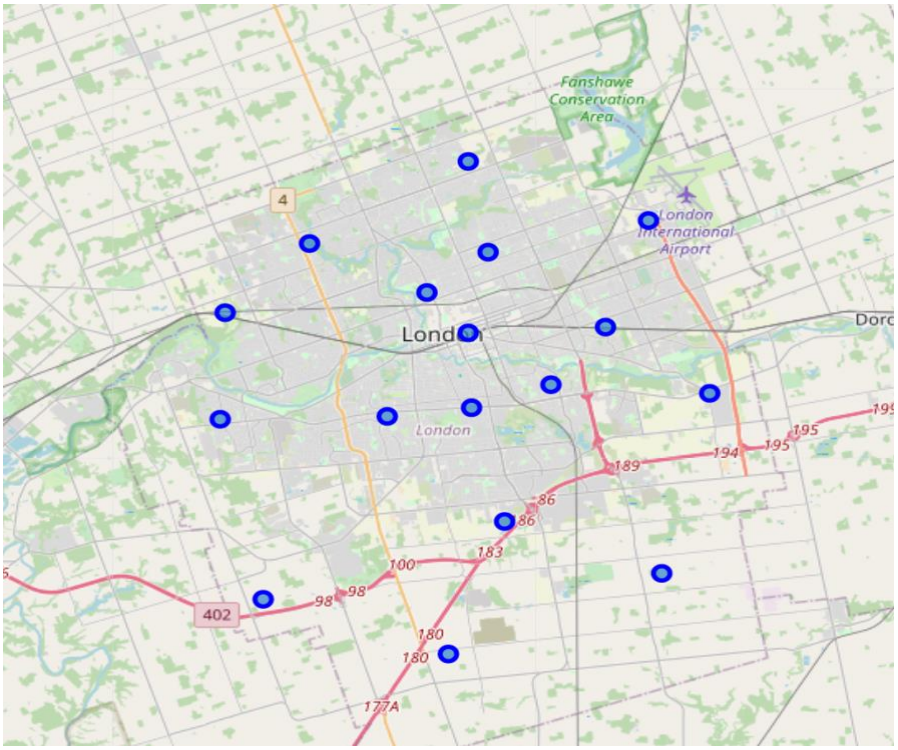
3. Methodology

- Data visualization, data exploration, data analysis and any machine learning methods utilizing for making recommendation will be comprehensively performed in this section.
- Since the required dataset for the neighborhoods in London have been thoroughly gathered **(postal code, borough, neighborhood, latitude, and longitude)**, the exploration of venues within **the radius of 1000 meters** from the default coordinate of neighborhoods can be carried out by using **Foursquare API.**
- Firstly, to make the analytic process more insightful, the **Folium package** will be utilized for the visualization of the geographic positions of all the neighborhoods in London.
- The second step of the analysis will be **the calculation and exploration of the distribution of venue categories** across different areas of London. The analysis will mainly focus on the most promising areas and within those create different clusters of neighborhoods that meet some basic requirements established in the discussion with stakeholders:
    1. Considering the areas with ****high density in the scatter of venues and the diversity of venue categories****.
    2. Presenting a geographic map of all such locations and creating clusters (applying ****k-means clustering**** of Machine Learning) of those locations to identify the candidate neighborhoods based on their similarity and dissimilarity and search for optimal areas.

- In the next step, the recommendation system for the trustworthy candidates will be built based on the exploratory venues within different neighborhoods in London. In addition, the use of the Toronto dataset as an additional reference is to make a comparison between the venue categories in both cities so that the business patterns which have never been exploited in London can be taken into consideration for the recommendation.
- Finally, the optimal solution will be ready to handle the stakeholder's problems.

3.1 Data Visualization

Map of neighborhoods in London



- The default geographic coordinates of London city are 42.9836747, -81.2496068.
- Glancing at the above map, it is remarkable that the distribution of neighborhoods becomes denser and denser when it gradually exposes to the central area of London. According to this observation, it is predicted that there will be more exploratory venues within the closer neighborhoods rather than the ones which are located in or near the outskirt of London city.

3.2 Data Exploration

- Foursquare application:

- Since the main purpose of this project is to optimize the useful application of Foursquare API which is one of the most handy and powerful Application Programming Interface tools for the geographic exploration, the utility of Foursquare will be fully exploited in the exploratory process. Now, the number of local venues in each individual neighborhood will be completely investigating.
- The neighborhood investigation only occurs within the limited radius which is set to 1000 meters from the center of the neighborhood.
- To get the venue data of each neighborhood from Foursquare, the unique client information (ID, Access Token) and the available coordinate (latitude, longitude) of each neighborhood are requisite, but the more technical part of how to create and use Foursquare API will not be mentioned here, only the final result after the neighborhood exploration. It is remarkable that the Foursquare

will basically make use of the latitude and longitude of the designated area to detect the approachable local venues within the limited radius.

- At the end of the neighborhood detection, the expected data features that have to be captured for the later analysis will be the name of venues, the venue latitude, the venue longitude, and the venue categories.

- Illustration for the exploration:

```
Exploration 1: West Huron Heights / Carling
Exploration 2: Glen Cairn
Exploration 3: East Tempo
Exploration 4: SW Argyle / Hamilton Road
Exploration 5: Central Hyde Park / Oakridge
Exploration 6: Southcrest / East Westmount / West Highland
Exploration 7: Jackson / Old Victoria / Bradley / North Highbury
Exploration 8: YXU / North and East Argyle / East Huron Heights
Exploration 9: Fanshawe / Stoneybrook / Stoney Creek / Uplands / East Masonville
Exploration 10: East Highland / North White Oaks / North Westminster
Exploration 11: South White Oaks / Central Westminster / East Longwoods / West Brockley
Exploration 12: Sunningdale / West Masonville / Medway / NE Hyde Park / East Fox Hollow
Exploration 13: Riverbend / Woodhull / North Sharon Creek / Byron / West Westmount
Exploration 14: South Highbury / Glanworth / East Brockley / SE Westminster
Exploration 15: Talbot / Lambeth / West Tempo / South Sharon Creek
Exploration 16: London Central
Exploration 17: UWO
```

- Illustration for the expected result:

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	N5Y	West Huron Heights / Carling	43.012	-81.231	Merla-Mae Ice Cream	43.010014	-81.241613	Ice Cream Shop
1	N5Y	West Huron Heights / Carling	43.012	-81.231	Jumbo Video	43.011495	-81.241666	Video Store
2	N5Y	West Huron Heights / Carling	43.012	-81.231	LCBO	43.003487	-81.227467	Liquor Store
3	N5Y	West Huron Heights / Carling	43.012	-81.231	Pho Lee	43.006657	-81.239302	Thai Restaurant
4	N5Y	West Huron Heights / Carling	43.012	-81.231	Metro	43.007334	-81.239668	Supermarket
...	...	...	...	...	...	...	...	...
341	N6A	UWO	42.998	-81.256	Archies	42.992069	-81.251477	Clothing Store
342	N6A	UWO	42.998	-81.256	The Barking Frog	42.991740	-81.251441	Bar
343	N6A	UWO	42.998	-81.256	Jade Nails	42.992917	-81.262711	Spa
344	N6A	UWO	42.998	-81.256	The Spot to Eat	42.990504	-81.251235	Breakfast Spot
345	N6A	UWO	42.998	-81.256	ISP Canada	42.992235	-81.264410	IT Services

346 rows × 8 columns

- Observably, the Foursquare has done 17 neighborhood explorations based on 17 neighborhoods and after the neighborhood investigation, there are 353 rows containing the exploratory venues of different neighborhoods and 8 columns containing the venues' corresponding features. However, this result table is messy, and it will be very arduous and time-consuming to accumulate and examine the total number of venues that one neighborhood has as there is a whole host of exploratory venues in one distinct neighborhood.

- To have a more visual overview after the exploration, each venue will be grouped by their corresponding Postal Code and Neighborhood.
  - Categorical table

	Postal Code	Neighborhood	Venue	Venue Latitude	Venue Longitude	Venue Category
0	N6B	London Central	100	100	100	100
1	N6A	UWO	56	56	56	56
2	N6C	East Highland / North White Oaks / North Westm...	54	54	54	54
3	N5Y	West Huron Heights / Carling	38	38	38	38
4	N6G	Sunningdale / West Masonville / Medway / NE Hy...	24	24	24	24
5	N6K	Riverbend / Woodhull / North Sharon Creek / By...	24	24	24	24
6	N5Z	Glen Cairn	18	18	18	18
7	N6J	Southcrest / East Westmount / West Highland	12	12	12	12
8	N6E	South White Oaks / Central Westminster / East ...	8	8	8	8
9	N5V	YXU / North and East Argyle / East Huron Heights	5	5	5	5
10	N5W	SW Argyle / Hamilton Road	5	5	5	5
11	N5X	Fanshawe / Stoneybrook / Stoney Creek / Upland...	5	5	5	5
12	N6H	Central Hyde Park / Oakridge	2	2	2	2
13	N6L	East Tempo	1	1	1	1
14	N6N	South Highbury / Glanworth / East Brockley / S...	1	1	1	1

- The above data table actually gives a more profound visualization for the neighborhood exploration. Applying coding will give the summation result of 120 unique venue categories which are approachable in London.
- Looking at the illustration above, there are a number of significant points:

1. Missing neighborhoods/ Outliers.

- There are only 15 neighborhoods with approachable local venues left after the exploratory process while 17 neighborhoods have been examined in the neighborhood exploration. There are two missing neighborhoods, which are N6M, N6P.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
6	N6M	London	Jackson / Old Victoria / Bradley / North Highbury	42.963	-81.139
14	N6P	London	Talbot / Lambeth / West Tempo / South Sharon C...	42.891	-81.324

- It is evident that there is no approachable venue that can be detected within the limited radius from the neighborhoods with the postal code N6M and N6P. These two missing neighborhoods can be considered as outliers from the others because of not providing the expected information and meeting the criteria mentioned in this project and are eliminated from the candidate list because of the inadequate standard for the next analysis.
- Only 15 remaining candidate areas will be taken into consideration and it is expected that the most promising living areas in London, Ontario can be discovered at the end.

2. Neighborhood with sparse venue density

- Some neighborhoods are extremely poor in the number of approachable local venues, which are N6E, N5V, N5W, N5X, N6H, N6L, and N6N. Nevertheless, it is advantageous that these

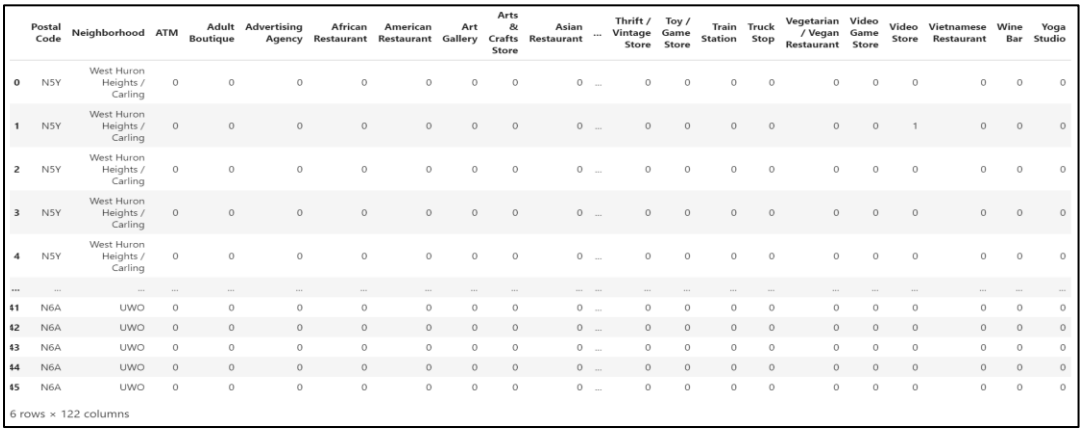
neighborhoods can be potentially got rid of the candidate list in the long run due to being unqualified, so the best ones will be unearthed.

3.3 Data Analysis

- In this section, the analytic techniques, mathematic calculations, and machine learning algorithms served to perform statistic testing for the analysis will be comprehensively elaborated. To be more specific, the key methods include the one-hot coding technique, K-Means clustering (ML), and 2D Cartesian distance calculation.

a) One-hot coding

- It is sometimes unknown that the "One-hot Coding" is one of the most efficient and easy-to-use techniques which can perform the conversion between the multiple categories or sorts which are in string values and categorical values which are in numeric format (1 if the category is available, otherwise 0).
- As indicated by the categorical table above, there are numerous venue categories detected in one neighborhood, so the application of "One-hot Coding" method will give an overall observation of venue categories which are accessible in a neighborhood.



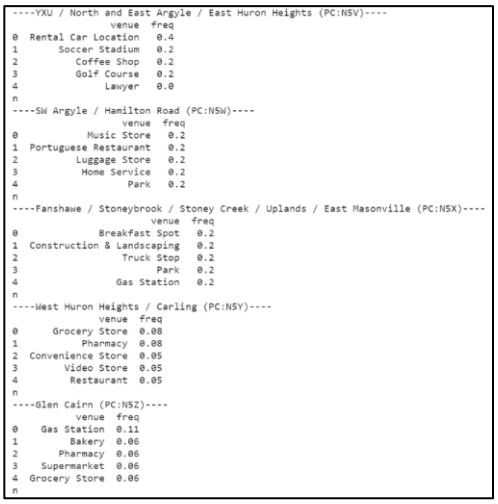
	Postal Code	Neighborhood	ATM	Adult Boutique	Advertising Agency	African Restaurant	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	...	Thrift / Vintage Store	Toy / Game Store	Train Station	Truck Stop	Vegetarian / Vegan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant	Wine Bar	Yoga Studio
0	N5Y	West Huron Heights / Carling	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	N5Y	West Huron Heights / Carling	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	1	0	0	0
2	N5Y	West Huron Heights / Carling	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	N5Y	West Huron Heights / Carling	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	N5Y	West Huron Heights / Carling	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
341	N6A	UWO	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
342	N6A	UWO	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
343	N6A	UWO	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
344	N6A	UWO	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
345	N6A	UWO	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

6 rows × 122 columns

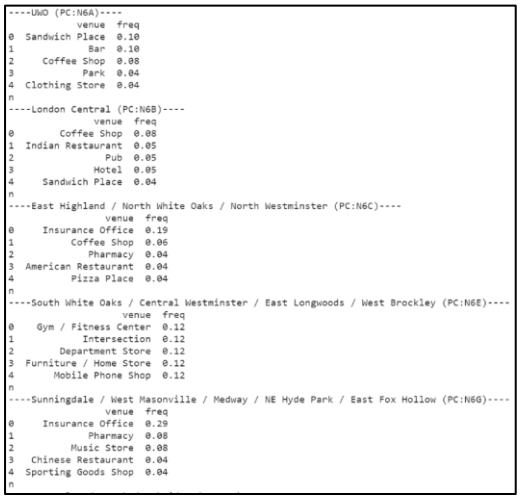
- The "One-hot Coding" table depicts 346 rows standing for different available categorical values of each venue and 122 columns will be the categorical features.
- At this point, the inquiry of how frequently the presence of a venue category will be in a neighborhood will be what have to be resolved. In other words, the frequency/ frequency distribution scores of each venue category need to be computed. The calculation of frequency score is fundamentally the mean/ average of the total categorical values of one venue category divided by the total number of venue categories in one neighborhood. This mean ratio indicates the frequent proportion (the quantity/ the distribution) of one venue category within its neighborhood. The greater the ratio is, the greater number of one venue category in a neighborhood.
  - Frequency/ Frequency distribution score.

```
----YXU / North and East Argyle / East Huron Heights (PC:N5V)----
                venue  freq
0  Rental Car Location   0.4
1       Soccer Stadium   0.2
2          Coffee Shop   0.2
3          Golf Course   0.2
4               Lawyer   0.0
n
```

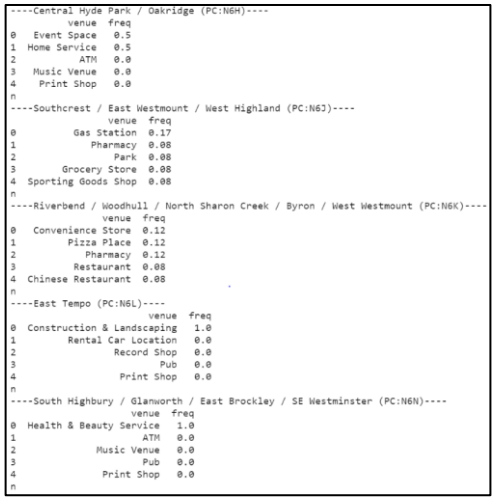
- Looking at the first row of the venue (row 0), the venue category with “Rental Car Location” has the highest frequency distribution score, demonstrating that there will be lots of venues that provide rental car services in YXU / North and East Argyle / East Huron Heights neighborhood. It is significant that the lowest score is equal to 0, implying that there is no lawyer office located in the neighborhood with the postal code, N5V. In general, the area with the postal code, N5V, has only 4 approachable venue categories, which reveals the minor number in the distribution of venue categories in this area.



```
----YXU / North and East Argyle / East Huron Heights (PC:N5V)----
                venue  freq
0  Rental Car Location   0.4
1       Soccer Stadium   0.2
2          Coffee Shop   0.2
3          Golf Course   0.2
4               Lawyer   0.0
n
----SW Argyle / Hamilton Road (PC:N5W)----
                venue  freq
0          Music Store   0.2
1  Portuguese Restaurant   0.2
2         Luggage Store   0.2
3          Home Service   0.2
4                  Park   0.2
n
----Fanshawe / Stoneybrook / Stoney Creek / Uplands / East Masonville (PC:N5X)----
                venue  freq
0           Breakfast Spot   0.2
1  Construction & Landscaping   0.2
2                Truck Stop   0.2
3                      Park   0.2
4             Gas Station   0.2
n
----West Huron Heights / Carling (PC:N5Y)----
                venue  freq
0      Grocery Store  0.08
1           Pharmacy  0.08
2  Convenience Store  0.05
3        Video Store  0.05
4         Restaurant  0.05
n
----Glen Cairn (PC:N5Z)----
          venue  freq
0   Gas Station  0.11
1        Bakery  0.06
2      Pharmacy  0.06
3   Supermarket  0.06
4  Grocery Store  0.06
n
```



```
----UWO (PC:N6A)----
             venue  freq
0  Sandwich Place  0.10
1             Bar  0.10
2     Coffee Shop  0.08
3            Park  0.04
4  Clothing Store  0.04
n
----London Central (PC:N6B)----
             venue  freq
0      Coffee Shop  0.08
1  Indian Restaurant  0.05
2              Pub  0.05
3            Hotel  0.05
4    Sandwich Place  0.04
n
----East Highland / North White Oaks / North Westminster (PC:N6C)----
                venue  freq
0      Insurance Office  0.19
1           Coffee Shop  0.06
2              Pharmacy  0.04
3  American Restaurant  0.04
4           Pizza Place  0.04
n
----South White Oaks / Central Westminster / East Longwoods / West Brockley (PC:N6E)----
                venue  freq
0    Gym / Fitness Center  0.12
1            Intersection  0.12
2       Department Store  0.12
3  Furniture / Home Store  0.12
4       Mobile Phone Shop  0.12
n
----Sunningdale / West Masonville / Medway / NE Hyde Park / East Fox Hollow (PC:N6G)----
                venue  freq
0      Insurance Office  0.29
1              Pharmacy  0.08
2          Music Store  0.08
3   Chinese Restaurant  0.04
4  Sporting Goods Shop  0.04
n
```



```
----Central Hyde Park / Oakridge (PC:N6H)----
          venue  freq
0   Event Space   0.5
1  Home Service   0.5
2           ATM   0.0
3   Music Venue   0.0
4    Print Shop   0.0
n
----Southcrest / East Westmount / West Highland (PC:N6J)----
                venue  freq
0         Gas Station  0.17
1            Pharmacy  0.08
2                Park  0.08
3       Grocery Store  0.08
4  Sporting Goods Shop  0.08
n
----Riverbend / Woodhull / North Sharon Creek / Byron / West Westmount (PC:N6K)----
                venue  freq
0   Convenience Store  0.12
1         Pizza Place  0.12
2            Pharmacy  0.12
3          Restaurant  0.08
4  Chinese Restaurant  0.08
n
----East Tempo (PC:N6L)----
                venue  freq
0  Construction & Landscaping   1.0
1         Rental Car Location   0.0
2                Record Shop   0.0
3                        Pub   0.0
4                 Print Shop   0.0
n
----South Highbury / Glanworth / East Brockley / SE Westminster (PC:N6N)----
                venue  freq
0  Health & Beauty Service   1.0
1                  ATM   0.0
2           Music Venue   0.0
3                  Pub   0.0
4           Print Shop   0.0
n
```

- Thoroughly observing the above the frequency distribution score of all venue categories of the neighborhoods, some remarkable points are found:
  + A number of potential candidate areas with a variety of venue categories.
  + Unqualified neighborhoods which are hardly diverse in the distribution of venue categories.
  + The top popular venue categories are score sorted in descending order based on the frequency distribution.
- Subsequently, only the most 10 commonplace venues categories are extracted for the candidate examination. This is a sub-step before getting into the more mathematic analysis.

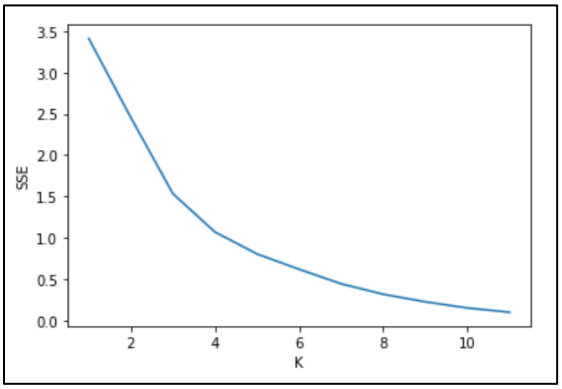
	Postal Code	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	N5V	YXU / North and East Argyle / East Huron Heights	Rental Car Location	Coffee Shop	Soccer Stadium	Golf Course	NaN	NaN	NaN	NaN	NaN	NaN
1	N5W	SW Argyle / Hamilton Road	Portuguese Restaurant	Home Service	Park	Luggage Store	Music Store	NaN	NaN	NaN	NaN	NaN
2	N5X	Fanshawe / Stoneybrook / Stoney Creek / Upland...	Gas Station	Truck Stop	Park	Breakfast Spot	Construction & Landscaping	NaN	NaN	NaN	NaN	NaN
3	N5Y	West Huron Heights / Carling	Grocery Store	Pharmacy	Video Store	Restaurant	Pizza Place	Convenience Store	ATM	Pet Store	Breakfast Spot	Sandwich Place
4	N5Z	Glen Cairn	Gas Station	Insurance Office	Bank	Pharmacy	Convenience Store	Coffee Shop	Discount Store	Fast Food Restaurant	Skating Rink	Photography Lab
5	N6A	UWO	Sandwich Place	Bar	Coffee Shop	Yoga Studio	Breakfast Spot	Café	Clothing Store	Japanese Restaurant	Park	Restaurant
6	N6B	London Central	Coffee Shop	Pub	Indian Restaurant	Hotel	Pizza Place	Sandwich Place	Bookstore	Italian Restaurant	Restaurant	Theater
7	N6C	East Highland / North White Oaks / North Westm...	Insurance Office	Coffee Shop	Pizza Place	American Restaurant	Pharmacy	Restaurant	Ice Cream Shop	Hotel	Print Shop	Café
8	N6E	South White Oaks / Central Westminster / East ...	Intersection	Hotel	Mobile Phone Shop	Men's Store	Gym / Fitness Center	Furniture / Home Store	American Restaurant	Department Store	NaN	NaN
9	N6G	Sunningdale / West Masonville / Medway / NE Hy...	Insurance Office	Music Store	Pharmacy	Coffee Shop	Mobile Phone Shop	Spa	Sporting Goods Shop	Pizza Place	Gas Station	Baby Store
10	N6H	Central Hyde Park / Oakridge	Home Service	Event Space	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	N6J	Southcrest / East Westmount / West Highland	Gas Station	Ice Cream Shop	Grocery Store	Pharmacy	Park	Business Service	Sporting Goods Shop	Pool	Bank	Storage Facility
12	N6K	Riverbend / Woodhull / North Sharon Creek / By...	Pizza Place	Pharmacy	Convenience Store	Restaurant	Chinese Restaurant	Bank	Park	Pub	Sandwich Place	Middle Eastern Restaurant
13	N6L	East Tempo	Construction & Landscaping	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	N6N	South Highbury / Glanworth / East Brockley / S...	Health & Beauty Service	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Personal observation: it is discernible that several neighborhoods have a diverse distribution in venue categories, but it is also easy to notice that some areas which contain NaN values (unavailable data) are sparse in the number of accessible venue categories such as N6L, N6N, N6H, N5V, N5W, and N5X, which do not meet the required criteria. Undoubtedly, these unqualified neighborhoods can be potentially got rid of from the candidate list later. To be more exact, a clustering method, K-Means, which is an algorithm of Machine Learning will be taken into practice to identify the potential candidate.

b) K-Means clustering

- K-Means is an unsupervised learning method of machine learning. The primary purpose of utilizing K-Means is to partition the neighborhoods into a number of k clusters based on the similarity and dissimilarity so those trustworthy candidates can be detected. However, the flaw of K-Means is that it requires a number of k initial random centroids (clusters) for the features of the given dataset, but it is very challenging to get the best k number for the optimal results.
- Fortunately, the elbow method can solve all of this problem. Firstly, the estimated interval for the k initial clusters will be set from 1 to 11 for the elbow testing. A visual graph will be needed in this circumstance.

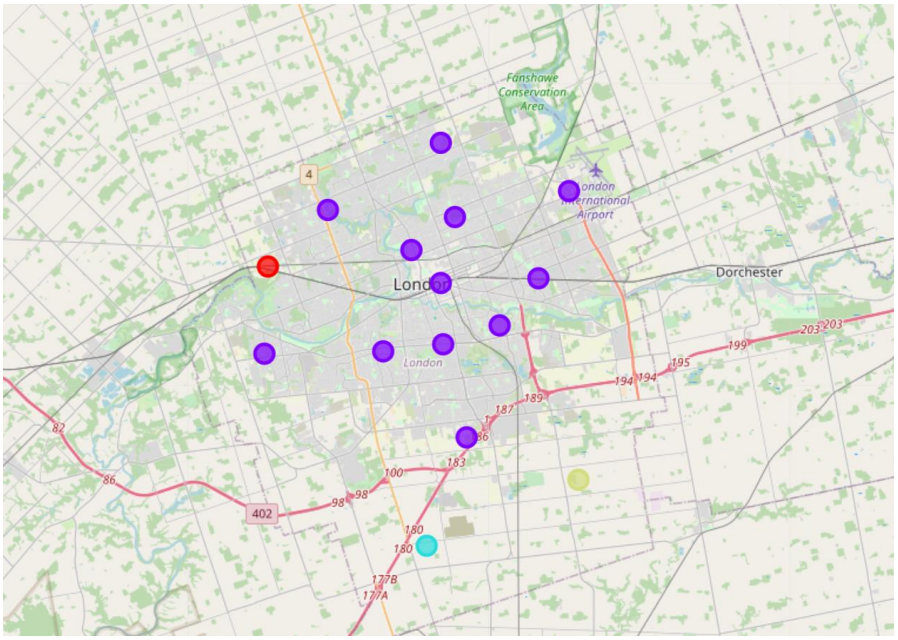
Elbow diagram



- In the elbow diagram above, the horizontal axis represents the number of k from 1 to 11, and the vertical axis indicates the Sum of Squared Error (SSE). Observing the sketched plot above, it is discernible that the Sum of Squared Error keeps decreasing while the K number is increasing. The point at which the drawn line is greatly bent is equal to 4. This is the elbow of the line at the point with k = 4. This also interprets the functionality of how the elbow method. The dataset containing potential candidate areas will be divided into 4 clusters. Then, the feature data that is the frequency distribution score of all venue categories of each neighborhood will be plugged into the training process of K-Means. As a result, 15 labeled clusters are obtained after the training corresponding to 15 neighborhoods. Now, each different area in London has its own label. The labels actually reveal the clusters/ groups to which the neighborhoods with similarity in venue categories belong. The neighborhoods with the same labels will be partitioned into the same clusters.

- **Cluster Label Table**

Postal Code	N5Y	N5Z	N6L	N5W	N6H	N6J	N5V	N5X	N6C	N6E	N6G	N6K	N6N	N6B	N6A
Cluster labels	0	0	1	0	3	0	0	0	0	0	0	0	2	0	0



Cluster 0: 11 neighborhoods **(purple)**. **Cluster 1**: 1 neighborhood **(teal)**. **Cluster 2**: 1 neighborhood **(green)**. **Cluster 3:** 1 neighborhood **(red)**.

- As indicated by the cluster label table, there are 4 unique cluster labels in total, and cluster 0 is the most outstanding cluster among the others with a great number of neighborhoods, promising that the prospective candidates will be in cluster 0. To be more precise, the neighborhoods in each different cluster will be examined individually.

- **Cluster analysis**

Cluster 3

	Postal Code	Neighborhood	Latitude	Longitude	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
4	N6H	Central Hyde Park / Oakridge	42.991	-81.34	Home Service	Event Space	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Cluster 2

	Postal Code	Neighborhood	Latitude	Longitude	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
12	N6N	South Highbury / Glanworth / East Brockley / S...	42.9	-81.159	Health & Beauty Service	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Cluster 1

	Postal Code	Neighborhood	Latitude	Longitude	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
2	N6L	East Tempo	42.872	-81.247	Construction & Landscaping	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Cluster 0

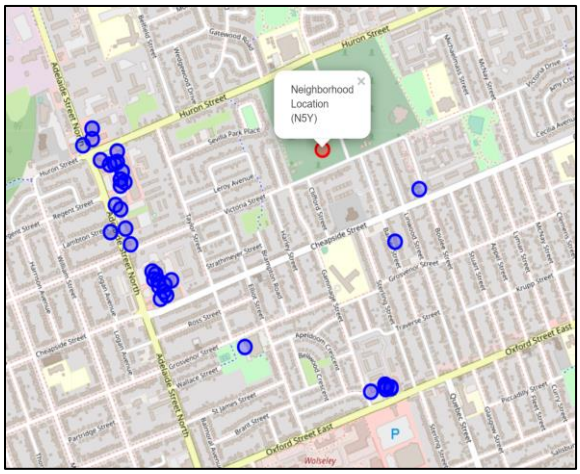
	Postal Code	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	N5Y	West Huron Heights / Carling	43.012	-81.231	Grocery Store	Pharmacy	Video Store	Restaurant	Pizza Place	Convenience Store	ATM	Pet Store	Breakfast Spot	Sandwich Place
1	N5Z	Glen Cairn	42.966	-81.205	Gas Station	Insurance Office	Bank	Pharmacy	Convenience Store	Coffee Shop	Discount Store	Fast Food Restaurant	Skating Rink	Photography Lab
3	N5W	SW Argyle / Hamilton Road	42.986	-81.182	Portuguese Restaurant	Home Service	Park	Luggage Store	Music Store	NaN	NaN	NaN	NaN	NaN
5	N6J	Southcrest / East Westmount / West Highland	42.955	-81.273	Gas Station	Ice Cream Shop	Grocery Store	Pharmacy	Park	Business Service	Sporting Goods Shop	Pool	Bank	Storage Facility
6	N5V	YXU / North and East Argyle / East Huron Heights	43.023	-81.164	Rental Car Location	Coffee Shop	Soccer Stadium	Golf Course	NaN	NaN	NaN	NaN	NaN	NaN
7	N5X	Fanshawe / Stoneybrook / Stoney Creek / Upland...	43.044	-81.239	Gas Station	Truck Stop	Park	Breakfast Spot	Construction & Landscaping	NaN	NaN	NaN	NaN	NaN
8	N6C	East Highland / North White Oaks / North Westm...	42.958	-81.238	Insurance Office	Coffee Shop	Pizza Place	American Restaurant	Pharmacy	Restaurant	Ice Cream Shop	Hotel	Print Shop	Café
9	N6E	South White Oaks / Central Westminster / East ...	42.918	-81.224	Intersection	Hotel	Mobile Phone Shop	Men's Store	Gym / Fitness Center	Furniture / Home Store	American Restaurant	Department Store	NaN	NaN
10	N6G	Sunningdale / West Masonville / Medway / NE Hy...	43.015	-81.305	Insurance Office	Music Store	Pharmacy	Coffee Shop	Mobile Phone Shop	Spa	Sporting Goods Shop	Pizza Place	Gas Station	Baby Store
11	N6K	Riverbend / Woodhull / North Sharon Creek / By...	42.954	-81.342	Pizza Place	Pharmacy	Convenience Store	Restaurant	Chinese Restaurant	Bank	Park	Pub	Sandwich Place	Middle Eastern Restaurant
13	N6B	London Central	42.984	-81.239	Coffee Shop	Pub	Indian Restaurant	Hotel	Pizza Place	Sandwich Place	Bookstore	Italian Restaurant	Restaurant	Theater
14	N6A	UWO	42.998	-81.256	Sandwich Place	Bar	Coffee Shop	Yoga Studio	Breakfast Spot	Café	Clothing Store	Japanese Restaurant	Park	Restaurant

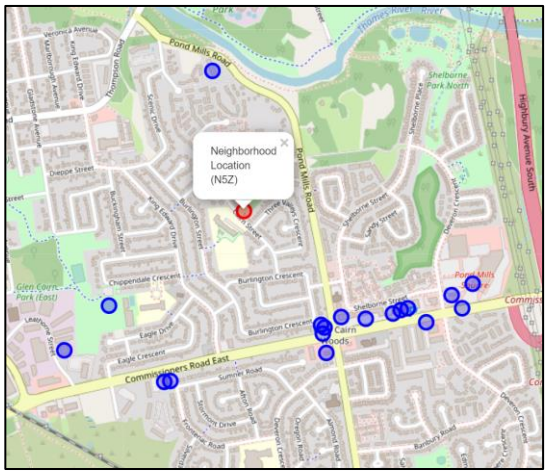
- According to the illustration above, it is evident that the neighborhood distribution of the purple cluster (cluster 0) outweighs the other clusters. More crucially, most of the neighborhoods in cluster 0 have more than 4 accessible venue categories while the other clusters mark the neighborhoods which are extremely restricted venue categories. This empowers the analysts to not only get rid of the neighborhoods with poor venue distribution but also achieve a number of potential areas which have a variety of venue categories.
- Undeniably, K-Means has contributed to filtering out the cluster with promising candidates, particularly cluster 0. However, the final objectives in this analysis mainly aim to the neighborhoods which has as many diverse venue categories as possible. Hence, the neighborhoods lacking the diversity in venue distribution will be out of the scope. The complete refined candidate list contains only 8 neighborhoods.

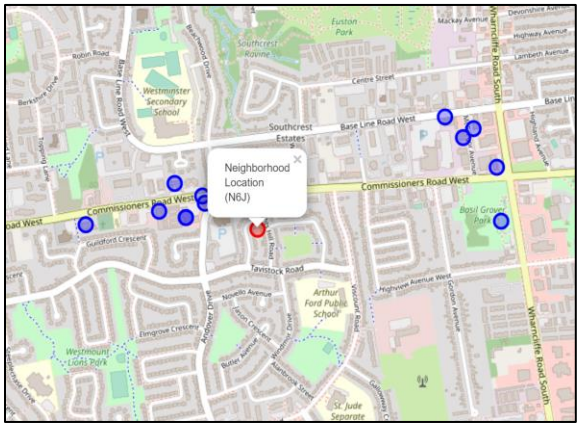
Candidate list

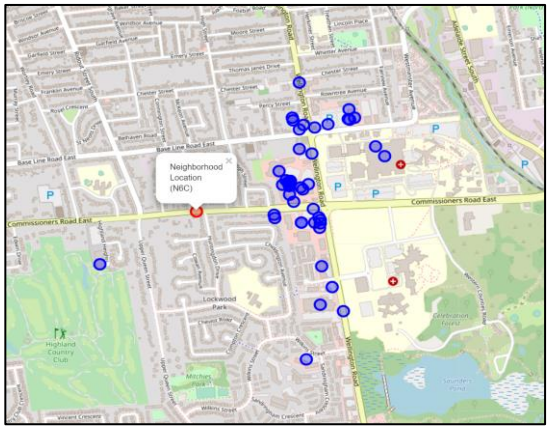
	Postal Code	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	N5Y	West Huron Heights / Carling	43.012	-81.231	Grocery Store	Pharmacy	Video Store	Restaurant	Pizza Place	Convenience Store	ATM	Pet Store	Breakfast Spot	Sandwich Place
1	N5Z	Glen Cairn	42.966	-81.205	Gas Station	Insurance Office	Bank	Pharmacy	Convenience Store	Coffee Shop	Discount Store	Fast Food Restaurant	Skating Rink	Photography Lab
5	N6J	Southcrest / East Westmount / West Highland	42.955	-81.273	Gas Station	Ice Cream Shop	Grocery Store	Pharmacy	Park	Business Service	Sporting Goods Shop	Pool	Bank	Storage Facility
8	N6C	East Highland / North White Oaks / North Westm...	42.958	-81.238	Insurance Office	Coffee Shop	Pizza Place	American Restaurant	Pharmacy	Restaurant	Ice Cream Shop	Hotel	Print Shop	Café
10	N6G	Sunningdale / West Masonville / Medway / NE Hy...	43.015	-81.305	Insurance Office	Music Store	Pharmacy	Coffee Shop	Mobile Phone Shop	Spa	Sporting Goods Shop	Pizza Place	Gas Station	Baby Store
11	N6K	Riverbend / Woodhull / North Sharon Creek / By...	42.954	-81.342	Pizza Place	Pharmacy	Convenience Store	Restaurant	Chinese Restaurant	Bank	Park	Pub	Sandwich Place	Middle Eastern Restaurant
13	N6B	London Central	42.984	-81.239	Coffee Shop	Pub	Indian Restaurant	Hotel	Pizza Place	Sandwich Place	Bookstore	Italian Restaurant	Restaurant	Theater
14	N6A	UWO	42.998	-81.256	Sandwich Place	Bar	Coffee Shop	Yoga Studio	Breakfast Spot	Café	Clothing Store	Japanese Restaurant	Park	Restaurant

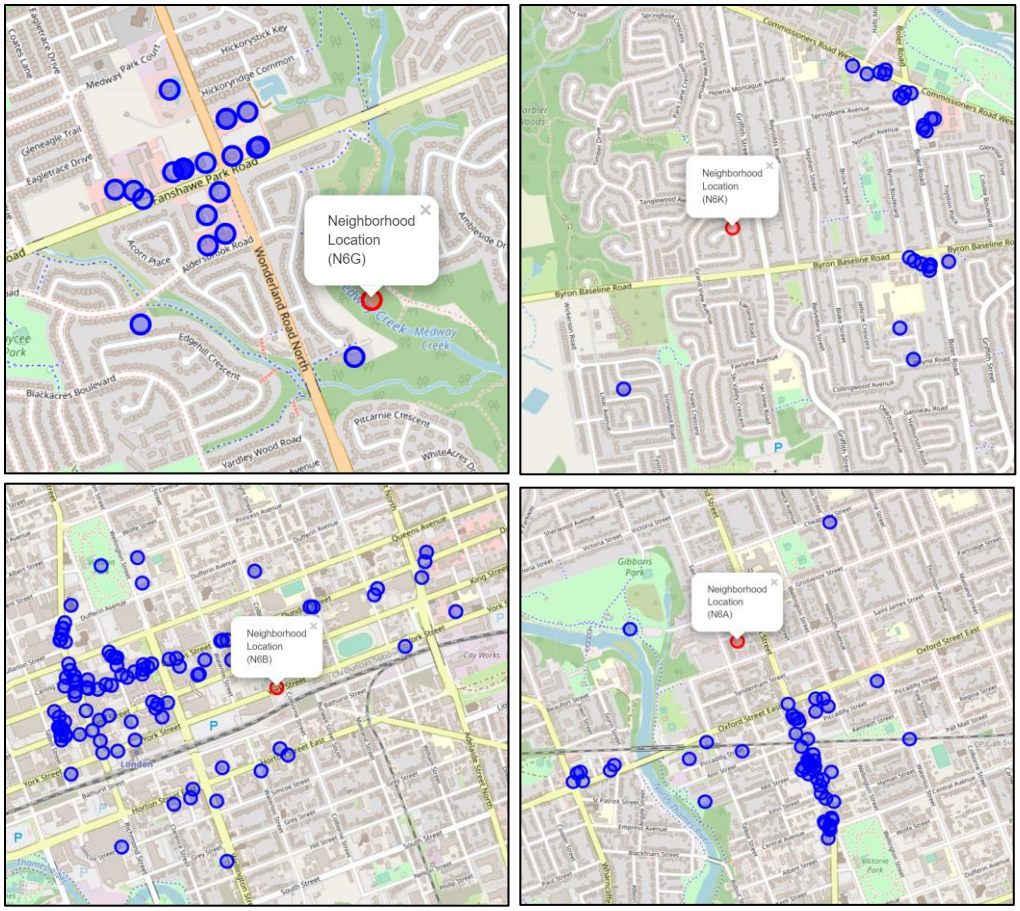
- Currently, the prospective candidates have been determined. In the next step, I will carry out a thorough candidate analysis so that the upsides and downsides of each candidate will be listed out. Visualizing the local venues of each neighborhood on a geographical map will boost analytical efficiency.











- What needed to concentrate on:
  - The distribution of venue categories within a neighborhood.
  - How dense the number of venues in one neighborhood should be.
  - The proximity between the venues and their neighborhood.
- If visually looking at the geographical map of different neighborhoods above, it is feasible to point out that N6B is densely surrounded by the local venues. This candidate is exactly the living area that the stakeholders are looking forward to inhabiting.
- Nonetheless, these visual analyses are not enough to determine the final result since the other potential ones will run the risk of being biased and missed out at the end. Moreover, the visual maps are incapable of showing the total number of various venue categories within a neighborhood and there is no clue or statistical data for the proximity from the venues to the neighborhood's center. In other words, a more persuasive, realistic, and rational demonstration will be required.

Ranking Criteria

- The top position in the recommendation list should start with the candidate that shows the diversity of venue categories and is densely surrounded by venues within the area.
- The candidate should be also close to different venues within the radius of 1000 meters from the center of neighborhood.

c) 2D Cartesian distance calculation and venue category distribution.

- By applying the 2D cartesian coordinate (x, y) instead of spherical coordinates (latitude, longitude), it is totally easy to calculate the distance from the venues to their neighborhood's center.

- Firstly, the latitude and longitude of each neighborhood will be converted into x, and y and this mathematic task can be perfectly done by calling the "pyproj" module in Python. Again, the concern about how to technically perform these coordinate transformations even using "pyproj" module will certainly not be interpreted here. Having all the 2D Cartesian coordinates will benefit the distance computation from the venues to the central neighborhood.

	Venue	Venue Latitude	Venue Longitude	Venue Category	Labels	x	y	Distance from Neighborhood center (meters)
0		42.984000	-81.239000	Center	1	-5.388038e+06	1.073775e+07	0.000000
1	Residence Inn by Marriott	42.985535	-81.239570	Hotel	0	-5.387783e+06	1.073777e+07	256.945605
2	YMCA	42.985099	-81.241463	Gym / Fitness Center	0	-5.387817e+06	1.073801e+07	341.681444
3	Canadian Tire Auto Service Centre	42.981576	-81.238838	Auto Garage	0	-5.388428e+06	1.073779e+07	391.951997
4	The Morrissey House	42.985865	-81.241693	Pub	0	-5.387691e+06	1.073801e+07	438.881536
...	...	...	...	...	...	...	...	...
94	New Delhi Deli	42.982431	-81.250865	Indian Restaurant	0	-5.388066e+06	1.073918e+07	1429.194590
95	Labatt Brewery	42.977674	-81.247433	Brewery	0	-5.388889e+06	1.073889e+07	1429.440840
96	Marky's Crepe Cafe	42.986349	-81.250646	Dessert Shop	0	-5.387446e+06	1.073905e+07	1431.764030
97	Tim Hortons	42.980562	-81.250161	Coffee Shop	0	-5.388378e+06	1.073914e+07	1434.902569
98	Blu Duby	42.983046	-81.251073	Restaurant	0	-5.387965e+06	1.073918e+07	1439.480090

❖ **Distance formula:** $d = \sqrt{(x^2 - x^1)^2 + (y^2 - y^1)^2}$

- All the required distances can be mathematically computed by applying the basic distance formula. These distances will be classified and only the ones which are within 1000 meters are accumulated for the later analysis. Besides, as listed in the "Ranking Criteria", only the neighborhoods having a variety of venue categories will be qualified, so let gather this requisite data.

	index	Total Venue Category	Total Venue with distance < 1000m
0	N6B	61	39
1	N6C	40	38
2	N6A	33	29
3	N5Y	30	2
4	N6G	17	10
5	N5Z	17	8
6	N6K	16	0
7	N6J	11	6

- This will be the complete table for the final analysis. It is not too unexpected when the neighborhood with postal code, N6B is on the top rank, which is plausible because it is actually the London Central where takes place the primary economic activities in London city. In the next sub-section, the more detailed analysis about each individual neighborhood will be clearly stated out.

d) Candidate Analysis

- In this section, I will examine the advantages and disadvantages of each individual neighborhood candidate based on the number of accessible venue categories and the most feasible proximity between the local venues and its neighborhood center and then finish the ranking system.

❖ **<u>Neighborhood analysis</u>**

****1. N6B****

+ 1st candidate
+ Total Venue Category: 63

+ Total Venue with distance < 1000m: 40

➢ **Personal Evaluation:** abundant quantity in venue categories. As shown by the geographical map, the distribution of venue categories is greatly dense within the neighborhood of N6B. It is significant that N6B is the center of London city in Ontario, so it is absolutely a decent and top-notch living area to settle down in the future because it has perfectly met all of the requisite criteria.

****2. N6C****

+ 2nd candidate
+ Total Venue Category: 42
+ Total Venue with distance < 1000m: 39

➢ **Personal Evaluation:** diverse quantity in venue categories. The second recommendation ought to be N6C. Although N6C does not have as many venue categories as N6B, it is still a well-located area for the stakeholders whose budget is restricted. It is quite obvious that the cost of living in an area being further from the central area tend to be cheaper.

****3. N6A****

+ 3rd candidate
+ Total Venue Category: 29
+ Total Venue with distance < 1000m: 24

➢ **Personal Evaluation:** Moderate quantity in venue categories. Overall, N6A has a restriction in venue categories (only 29), but one of its striking features is that there are up to 24 venues which are close to the local center (red point), which enables local inhabitants to save money for transportation.

****4. N5Y****

+ 4th candidate
+ Total Venue Category: 29
+ Total Venue with distance < 1000m: 2

➢ **Personal Evaluation:** Moderate quantity in venue categories. Although both N5Y and N6A have the same number of venue categories, most of the venues' distances within N5Y are greater than 1000 meter(s) since we are trying our best to maximize the proximity. This tend to create a number of hindrances for the elderly or children who are not car owners or have to deal with traffic problems. In general, N5Y is still considered as a trustworthy area to live.

****5. N6G****

+ 5th candidate
+ Total Venue Category: 17
+ Total Venue with distance < 1000m: 10

➢ **Personal Evaluation:** Sparse quantity in venue categories. N6G do not has different types of venue, but it can still provide the local inhabitants with basic daily supplement.

****6. N5Z****

+ 6th candidate
+ Total Venue Category: 17
+ Total Venue with distance < 1000m: 8

➢ **Personal Evaluation:** Sparse quantity in venue categories. There is nothing much different between N6G and N5Z.

****7. N6K****

+ 7th candidate
+ Total Venue Category: 16
+ Total Venue with distance < 1000m: 0

- **Personal Evaluation:** Sparse quantity in venue categories. N6K is the only neighborhood which has the worst proximity between venues and the center of their neighborhood.

****8. N6J****

\+ 8th candidate

\+ Total Venue Category: 11

\+ Total Venue with distance < 1000m: 6

- **Personal Evaluation:** poor venue categories. An inferior candidate among the others.

❖ **<u>Neighborhood to neighborhood analysis</u>**

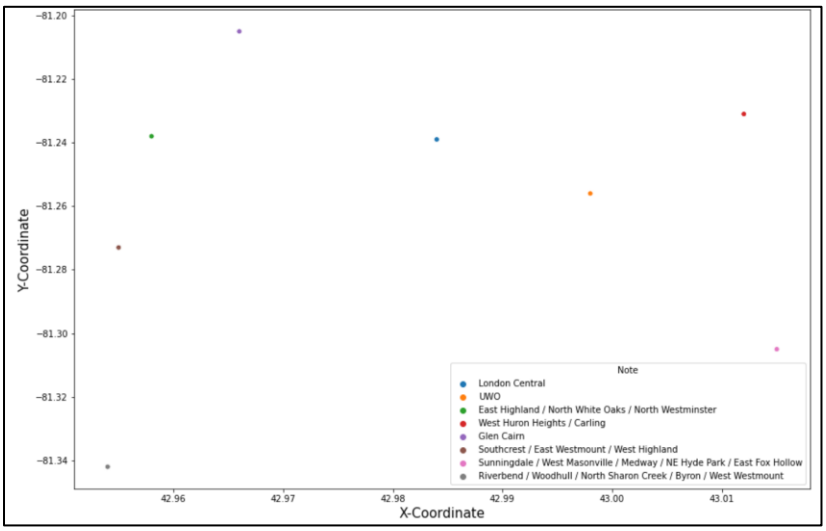
- In case, the stakeholders cannot afford their accommodation in cosmopolitan and bustling areas like London Central, they definitely have another alternative choice. Specifically, the cost of real estate or even housing prices is completely changeable in different local areas. The cost of living near the city center will prohibitively expensive because of the high standard of living, but isolating from the center will greatly help stakeholders save a bunch of money. Obviously, the greater distance from the center, the cheaper the living expense will be.
- Observing the geographical maps of the every-single area, it is noticeable that N6B is greatly crowded in venue distribution and the proximity between the neighborhood center (in red) and its venues is greatly dense. Therefore, this candidate will be chosen as the central base to measure the distance among the other neighborhoods.

	Postal Code	Neighborhood	Lat	Long	x	y	Distance from London Central
0	N6B	London Central	42.984	-81.239	-5.388038e+06	1.073775e+07	0.000000
1	N6A	UWO	42.998	-81.256	-5.385488e+06	1.073938e+07	3027.931136
2	N6C	East Highland / North White Oaks / North Westm...	42.958	-81.238	-5.392204e+06	1.073829e+07	4200.914811
3	N5Y	West Huron Heights / Carling	43.012	-81.231	-5.383724e+06	1.073610e+07	4618.439279
4	N5Z	Glen Cairn	42.966	-81.205	-5.391544e+06	1.073422e+07	4970.773897
5	N6J	Southcrest / East Westmount / West Highland	42.955	-81.273	-5.392025e+06	1.074247e+07	6180.506523
6	N6G	Sunningdale / West Masonville / Medway / NE Hy...	43.015	-81.305	-5.381858e+06	1.074467e+07	9283.563478
7	N6K	Riverbend / Woodhull / North Sharon Creek / By...	42.954	-81.342	-5.390879e+06	1.075058e+07	13141.269967

- Distance from UWO to London Central = 3.03km
- Distance from East Highland / North White Oaks / North Westminster to London Central = 4.2 km
- Distance from West Huron Heights / Carling to London Central = 4.62km
- Distance from Glen Cairn to London Central = 4.97km
- Distance from Southcrest / East Westmount / West Highland to London Central = 6.18km
- Distance from Sunningdale / West Masonville / Medway / NE Hyde Park / East Fox Hollow to London Central = 9.28km
- Distance from Riverbend / Woodhull / North Sharon Creek / Byron / West Westmount to London Central = 13.14km

- In this second analysis, it is observable that UWO is the first recommended neighborhood because of being closest to London Central. Moreover, according to the neighborhood analysis, UWO is voted as the 3rd candidate with a good number of venue categories and having the best proximity to most of its local venues.
- These calculations indicate the computed distances from each individual neighborhood to London Central. Based on this calculation, the stakeholders will have a second optional choice for their future inhabitation and the final decision can be made.

- In conclusion, the stakeholders should take both of the candidate analysis into account so that the best optimal decision can be finally made relying on the satisfied conditions or personal demands.



- The scatter plot above just illustrates the location of each neighborhood in 2D Cartesian coordinate in order to give the stakeholders a visual observation of how far the neighborhoods will be from each other.

e) Recommendation

- Although the first tough problem has been completely solved, the last one is quite challenging since the stakeholders also desire to get more informative recommendations for the business they can get down to after settling down, and obviously, there will have to be a lot of essential analysis such as economic research for the business tendency in London city. As mentioned in the objectives of this analysis, the application of Foursquare API must be optimized, so the handy functionality of Foursquare API can be used to deal with this issue in lieu of putting a lot of effort into figuring out the external source information.
- There is no denying that London is not a big and densely populated city as Toronto, resulting in the restriction in business models. Fortunately, this enables the analyst to expand the scope of the exploration for prospective business models which are approachable in Toronto and have not been exploited in London. For instance, I will reuse the Toronto dataset which is available in the previous lab sessions as an external reference and make comparisons between the number of unique business patterns in both cities so that untouched business patterns can be identified and recommended.
- Applying the similar steps in the “data acquisition and cleaning” part on the Toronto dataset, the required data for the comparison of business patterns is entirely gathered and only the boroughs which are the sub-parts of “Toronto” such as 'Downtown Toronto', 'East Toronto', 'West Toronto', 'East York East Toronto', and 'Central Toronto' will be thoroughly explored.

- **London vs Toronto**

- The illustration below indicates the neighborhood exploration in Toronto city. The process of comparing will mainly be carried out on the “Venue Category” feature which is available in both London and Toronto dataset. Based on the Toronto datasets, the sole venue categories or business patterns that are not accessible or unavailable in London are fully extracted to make recommendations in the later part.

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M5A	['Regent Park', 'Harbourfront']	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
1	M5A	['Regent Park', 'Harbourfront']	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
2	M5A	['Regent Park', 'Harbourfront']	43.654260	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	M5A	['Regent Park', 'Harbourfront']	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
4	M5A	['Regent Park', 'Harbourfront']	43.654260	-79.360636	The Distillery Historic District	43.650244	-79.359323	Historic Site
...	...	...	...	...	...	...	...	...
3285	M7Y	Business reply mail Processing Centre 969 East...	43.662744	-79.321558	TTC Stop #03057	43.663314	-79.330099	Light Rail Station
3286	M7Y	Business reply mail Processing Centre 969 East...	43.662744	-79.321558	Rogers Authorized Dealer - Lifestyle Communica...	43.658084	-79.328486	Mobile Phone Shop
3287	M7Y	Business reply mail Processing Centre 969 East...	43.662744	-79.321558	Booty Camp Fitness	43.668830	-79.326700	Gym / Fitness Center
3288	M7Y	Business reply mail Processing Centre 969 East...	43.662744	-79.321558	Greenwood Dog Park	43.667854	-79.329160	Dog Run
3289	M7Y	Business reply mail Processing Centre 969 East...	43.662744	-79.321558	Breakfast Club	43.662811	-79.310174	Breakfast Spot

3290 rows × 8 columns

- **Guideline:** After the extraction of essential data, it should be guaranteed that the final data is efficiently refined without being duplicated or containing any interchangeable duplicates in the category of venues. For example, the term “food market” or “food store” may be different as markets normally tend to be bigger in size and have a variety of food items for selling rather than stores, but in fact, both of it serves for food supply. Moreover, the business organizations being strictly and solely under the responsibility and control of governmental or national organizations, public services being extensively and freely opened for public and community purposes, and commonplace destinations being rampantly found in any places with a high standard of living will be filtered out from the list of business recommendation including museums, zoos, airports, subway, aquarium, natural landscapes, library, gym, supermarket, school, college, theater, stadium, boat & ferry, etc.

The list of original extracted data before refining contains 190 recommendations.

```
['Distribution Center', 'Historic Site', 'Chocolate Shop', 'Performing Arts Venue', 'Tech Startup', 'Shoe Store', 'Animal Shelter', 'Diner', 'German Restaurant', 'Karao
ke Bar', 'Gym Pool', 'Dance Studio', 'Museum', 'Comic Shop', 'Plaza', 'Ramen Restaurant', 'Shopping Mall', 'Tanning Salon', 'Miscellaneous Shop', 'Steakhouse', 'Bubble
Tea Shop', 'Lounge', 'Neighborhood', 'Botanical Garden', 'Garden', 'Poke Place', 'Monument / Landmark', 'Art Museum', 'BBQ Joint', 'Creperie', 'Speakeasy', 'Gay Bar',
'Beer Bar', 'Cocktail Bar', 'Jazz Club', 'Fountain', 'Tailor Shop', 'Optical Shop', 'Movie Theater', 'Fish Market', 'Cheese Shop', 'Bagel Shop', 'Trail', 'Indie Movie T
heater', 'Health Food Store', 'Beach', 'Nail Salon', 'Cupcake Shop', 'Caribbean Restaurant', 'Mattress Store', "Women's Store", 'Dive Bar', 'Dry Cleaner', 'Basketball S
tadium', 'Comfort Food Restaurant', 'Salad Place', 'Lake', 'Molecular Gastronomy Restaurant', 'Gourmet Shop', 'Deli / Bodega', 'Office', 'University', 'Smoke Shop', 'So
uvlaki Shop', 'Opera House', 'South American Restaurant', 'Korean Restaurant', 'Comedy Club', 'Taco Place', 'Gift Shop', 'School', 'Ethiopian Restaurant', 'Candy Stor
e', 'Rock Climbing Spot', 'Jewish Restaurant', 'Brazilian Restaurant', 'Dog Run', 'Wine Shop', 'Camera Store', 'Other Repair Shop', 'Turkish Restaurant', 'Hostel', 'Don
ut Shop', 'Dim Sum Restaurant', 'Jewelry Store', 'Roof Deck', 'Aquarium', 'Scenic Lookout', 'Baseball Stadium', 'Track', 'Fried Chicken Joint', 'Cuban Restaurant', 'Bou
tique', 'Malay Restaurant', 'Mac & Cheese Joint', 'Argentinian Restaurant', 'Tapas Restaurant', 'Whisky Bar', 'Dumpling Restaurant', 'Butcher', 'Fruit & Vegetable Stor
e', 'Tibetan Restaurant', 'Frozen Yogurt Shop', 'Fish & Chips Shop', 'Churrascaria', 'Hawaiian Restaurant', 'Beach Bar', 'Food', 'Climbing Gym', 'Chiropractor', 'Hotel
Bar', 'Stadium', 'Accessories Store', 'Poutine Place', 'Pakistani Restaurant', 'Skate Park', 'Snack Place', 'Indian Chinese Restaurant', 'Indie Theater', 'Baseball Fiel
d', 'Garden Center', 'Food & Drink Shop', 'Stationery Store', 'Coworking Space', 'Curling Ice', 'Pide Place', 'Antique Shop', 'Flea Market', 'Falafel Restaurant', 'Coll
ege Quad', 'College Gym', 'Gaming Cafe', 'Persian Restaurant', 'Organic Grocery', 'Cajun / Creole Restaurant', 'Wings Joint', 'Supplement Shop', 'Castle', 'Pastry Sho
p', 'Music School', 'Bike Shop', 'Pool Hall', 'Lingerie Store', 'Amphitheater', 'Other Great Outdoors', 'Zoo', 'Light Rail Station', 'Cambodian Restaurant', 'Sports Clu
b', 'Photography Studio', 'Indonesian Restaurant', 'Cemetery', 'Syrian Restaurant', 'Doner Restaurant', 'Belgian Restaurant', 'College Theater', 'Cantonese Restaurant',
'Noodle House', 'Arepa Restaurant', 'Food Court', 'Filipino Restaurant', 'Street Art', 'Udon Restaurant', 'Sri Lankan Restaurant', 'Rental Service', 'Airport', 'Harbor
/ Marina', 'Sculpture Garden', 'Airport Lounge', 'Tunnel', 'Airport Food Court', 'Airport Terminal', 'Airport Service', 'Boat or Ferry', 'Pier', 'Bed & Breakfast', 'Eye
Doctor', 'Bike Rental / Bike Share', 'Pie Shop', 'Farm', 'Taiwanese Restaurant', 'General Entertainment', 'Market', 'Library', 'Recording Studio', 'Theme Restaurant',
'Martial Arts School', 'Escape Room', 'Sake Bar', 'Hardware Store']
```

The list of extracted data after refining contains 113 recommendations. Perceptually, 77 recommendations have been eliminated.

```
List of recommendation before removing: 190

************************ Removing unqualified venue categories ************************
Airport, Airport Lounge, Airport Food Court, Airport Service, Airport Terminal, Skate Park, University, Museum, Art Museum, Castle, Boat or Ferry, Aquarium, Zoo, Fount
ain, Beach, Beach Bar, Botanical Garden, Garden Center, Sculpture Garden, Trail, Lake, Track, Distribution Center, Basketball Stadium, Baseball Stadium, Stadium, Neigh
borhood, Library, Stationery Store, Light Rail Station, Gym Pool, Climbing Gym, College Gym, Movie Theater, Indie Movie Theater, Indie Theater, College Theater, Fish M
arket, Flea Market, Market, Shopping Mall, Health Food Store, Comfort Food Restaurant, Food, Food & Drink Shop, Food Court, Organic Grocery, Plaza, Monument / Landmar
k, School, Music School, Martial Arts School, College Quad, Pool Hall, Scenic Lookout, Farm, Diner, Other Great Outdoors, Sports Club, Rock Climbing Spot, Dog Run, Per
forming Arts Venue, Street Art, Historic Site, Animal Shelter, Opera House, Tunnel, Curling Ice, Garden, Baseball Field, Amphitheater, Harbor / Marina, Cemetery, Cowor
king Space, Bike Rental / Bike Share, Pier, Eye Doctor,

List of recommendation after removing: 113
```

- 113 business recommendations will be classified into each distinct group based on their properties.

- There are 5 categories for the business recommendations: restaurant, shop, bar, store, and remains.

Recommendation table

Restaurant	Shop	Bar & Club	Store	Remains
German Restaurant	Chocolate Shop	Karaoke Bar	Shoe Store	Tech Startup
Ramen Restaurant	Comic Shop	Gay Bar	Mattress Store	Dance Studio
Caribbean Restaurant	Miscellaneous Shop	Beer Bar	Women's Store	Tanning Salon
Molecular Gastronomy Restaurant	Bubble Tea Shop	Cocktail Bar	Candy Store	Steakhouse
South American Restaurant	Tailor Shop	Jazz Club	Camera Store	Lounge
Korean Restaurant	Optical Shop	Dive Bar	Jewelry Store	Poke Place
Ethiopian Restaurant	Cheese Shop	Comedy Club	Fruit & Vegetable Store	BBQ Joint
Jewish Restaurant	Bagel Shop	Whisky Bar	Accessories Store	Creperie
Brazilian Restaurant	Cupcake Shop	Hotel Bar	Lingerie Store	Speakeasy
Turkish Restaurant	Gourmet Shop	Sake Bar	Hardware Store	Nail Salon
Dim Sum Restaurant	Smoke Shop			Dry Cleaner
Cuban Restaurant	Souvlaki Shop			Salad Place
Malay Restaurant	Gift Shop			Deli / Bodega
Argentinian Restaurant	Wine Shop			Office
Tapas Restaurant	Other Repair Shop			Taco Place
Dumpling Restaurant	Donut Shop			Hostel
Tibetan Restaurant	Frozen Yogurt Shop			Roof Deck
Hawaiian Restaurant	Fish & Chips Shop			Fried Chicken Joint
Pakistani Restaurant	Antique Shop			Boutique
Indian Chinese Restaurant	Supplement Shop			Mac & Cheese Joint
Falafel Restaurant	Pastry Shop			Butcher
Persian Restaurant	Bike Shop			Churrascaria
Cajun / Creole Restaurant	Pie Shop			Chiropractor
Cambodian Restaurant				Poutine Place
Indonesian Restaurant				Snack Place
Syrian Restaurant				Pide Place
Doner Restaurant				Gaming Cafe
Belgian Restaurant				Wings Joint
Cantonese Restaurant				Photography Studio
Arepa Restaurant				Noodle House
Filipino Restaurant				Rental Service
Udon Restaurant				Bed & Breakfast
Sri Lankan Restaurant				General Entertainment
Taiwanese Restaurant				Recording Studio
Theme Restaurant				Escape Room

- The recommendation table above will enable the stakeholders to make their own choices for what kind of business pattern they are able to initiate after settling down.
- **Personal evaluation**: More crucially, restaurant is one of the most striking recommendations with a great number of various options, which can be potentially deployed in London down the road. According to this, the stakeholders ought to pay their attention to this first recommendation because investing money in opening restaurants which have not been existed before in the local area can bring a whole host of tremendous benefits. The other business models such as bars, shops, and

stores can also be deployed in London, but it is advised that the stakeholders should be cautious in making their decisions to get down to business likes bar or clubs because these entertaining places are strictly ruled by the government in order to restrict social crime. Lastly, the uncategorized recommendation/ remains, as the name implies, is a combination of different kinds of business patterns.

4. Result and Discussion

- In this analysis, I have accomplished the main tasks in helping the stakeholders to figure out which part of London is worth to inhabit. Eventually, I have achieved the required goal and give the stakeholders a number of valuable offers:

1. **8 worth-living neighborhoods** in London and their advantages and disadvantages.
2. **5 recommendations** for the choice of business patterns.

- Please, taking **NOTE** that the default coordinates/ center of the neighborhood which is mentioned throughout the capstone hypothetically represents the actual geographic location which the stakeholders choose for their accommodation in the future, so the calculated distance from the default coordinates/ center of the neighborhood to the surrounded venues is just estimated based on the available dataset scraping from the webpage and positional estimation. In fact, these distance calculations can be changeable whenever the center of the neighborhood or the actual position moves.
- In the Candidate section, I have mentioned a number of potential living areas which will meet the capstone's listed criteria. More crucially, some neighborhoods which are **N6H, N6N, N6L, N5W, N5V, and N5X** have been eliminated from the recommendation because of the low density of venue categories within the local area, but these neighborhoods can also be in the preference of the stakeholders. In reality, not all of the stakeholders want to inhabit in a hustling and bustling area. This is mainly because the cost of living in the metropolitan areas is prohibitively expensive. Moreover, some people have a desire to live in a peaceful residential area where they can find their inner peace and immerse themselves in the natural surroundings without being affecting by the fast-living pace of contemporary cities.

- **Notice:** Another critical issue that should be considered in the "Business Recommendation" section is that these recommendations are only **used for references**. There are a lot of internal and external factors which have great impacts on start-up. To be more specific, it relies on the **stakeholder's personal profile, strength, experience, and orientation**. Hence, please **take this notice into account** before or after examining the "Business Recommendation" section.

5. Conclusion

- To recap, in this capstone. I have successfully explored the neighborhoods in London, thoroughly specified several ideal residential areas, and optimally maximized the proximity from the neighborhood to the local venues by using the most efficient geographical tool, Foursquare API, and applying the distance calculation based on 2D Cartesian coordinates. In addition, the suggestion for the potential business patterns have also been accomplished as the stakeholder's requirements.
- More crucially, I have given out the most feasible answers for the 3 main questions which are listed as the final goal in this capstone and achieved the main goals in the an ultimate way so as to fulfill the stakeholder's desires.
- This analysis is just an objective guideline. Hence, the final decision still relies on the stakeholder's perspective in considering the different characteristics and the pros and cons of each neighborhood and business pattern. It is advised that the stakeholders should take into account a number of key factors such as the local security, the number of available accommodations, the best proximity to the workplace, the employment availability, and the user's demands so that the optimal determination can be implemented in settling down and initiating a business.

Reference

London, Ontario. (2021, April 08). Retrieved April 12, 2021, from https://en.wikipedia.org/wiki/London,_Ontario

The definitive guide to living In London, Ontario. (n.d.). Retrieved April 12, 2021, from https://www.homesinlondonontario.ca/living-london-ontario/

Postal codes Ontario, Canada. (n.d.). Retrieved April 12, 2021, from http://www.geonames.org/postalcode-search.html?q=london&country=CA&adminCode1=ON&fbclid=IwAR2XipWkuSm3F9YSjjVvFqp7SfYCPl9_XaxiehoPnn-7XmsjtnJBrbKh31g

Beautifulsoup4. (n.d.). Retrieved April 12, 2021, from https://pypi.org/project/beautifulsoup4/

Foursquare. (2021, April 08). Retrieved April 12, 2021, from https://foursquare.com/