



The Battle of Neighbourhoods

Background

London

- The United Kingdom's migrant population is concentrated in London.
- About 35% of people living in the UK who were born abroad live in the capital city.

Toronto

- In 2019, Canada welcomed more than 300,000 immigrants.
- More than one in three immigrants has chosen to settle in the Greater Toronto Area.

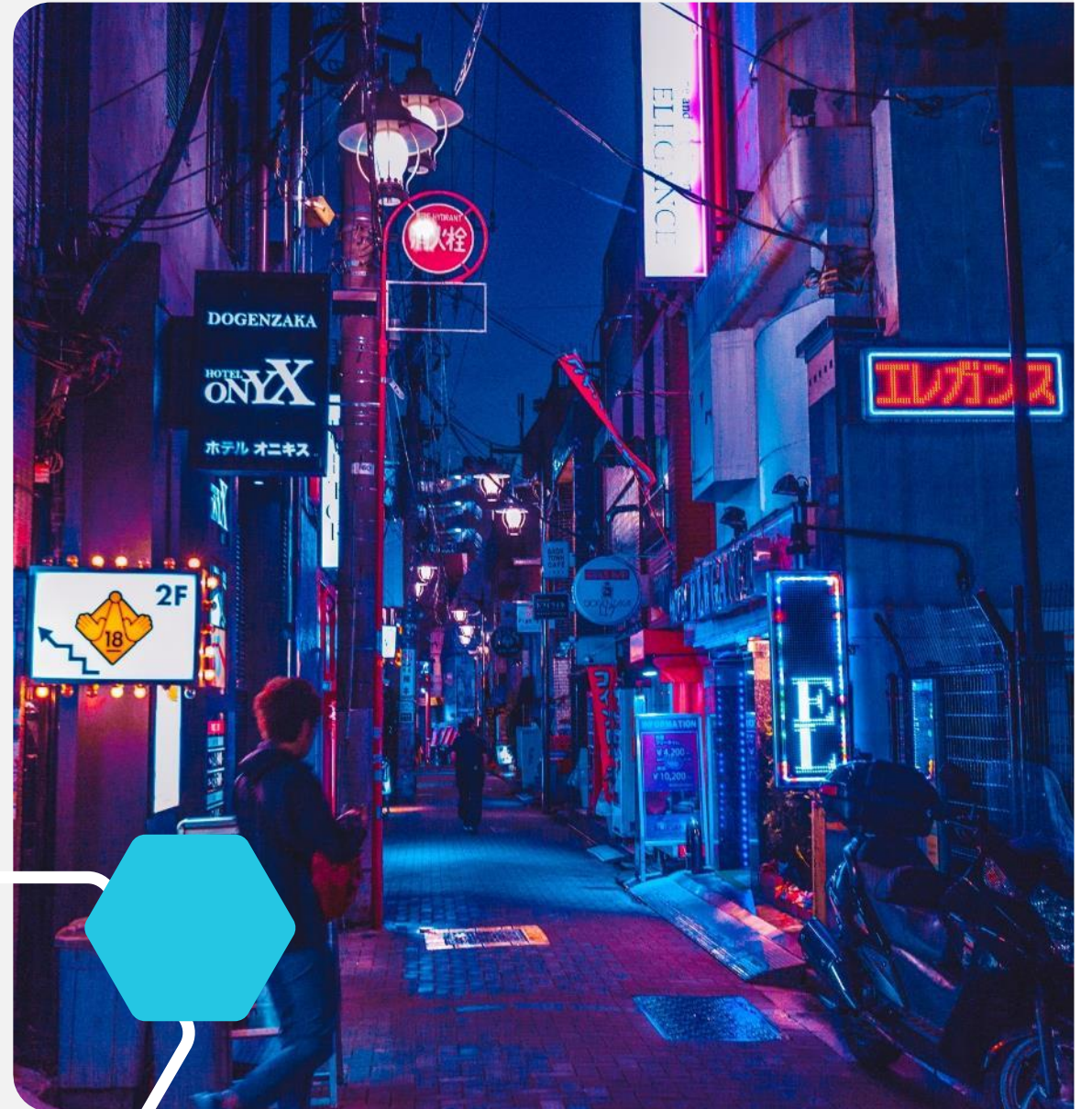


Background

Toronto and London are far away from each other

But,

It is still possible to find similarities between these two cities.



Business Problem

- Environment and culture shock are a major concern for migrants when moving to another country.
- Finding a similar neighborhood to one was used to live in or maybe a whole different neighborhood for a change of pace could be helpful when moving to another country.



Business Problem

In this project, we will introduce a machine learning technique to cluster Toronto and London areas in order to suggest neighbourhoods that are the best choice for migrants based on surrounding facilities such as school, hospital, shops, etc.



Data

Toronto

- Consists of Toronto's postcodes boroughs, neighbourhoods.
 - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- A csv file that has the geographical coordinates of each postal code for neighbourhoods in Toronto.
 - http://cocl.us/Geospatial_data

Foursquare API

- Search tool used to obtain information on neighbourhoods' venues and can be used to identify and compare Toronto and London neighbourhoods.



London

- Consists of London's postcodes, boroughs, neighbourhoods, post town.
- Data source:
https://en.wikipedia.org/wiki/List_of_areas_of_London
- Geocoder Package for geographical coordinates.



Methodology

Data Cleaning



Toronto

- Ignore any rows with a Borough that is Not assigned
- If a row has a borough but a Not assigned Neighbourhood, then the Neighbourhood will be the same as the borough
- After cleaning the dataset, we end up with **103 rows**



London

- Any rows with the same postcode are combined into one with the Borough and Neighbourhood separated with a slash.
- After cleaning the dataset, we end up with **175 rows**.

Methodology

Obtain Coordinates

Toronto

- CSV file with the geographic coordinates of each neighbourhood postal code in Toronto.

	City	Borough	Neighbourhood	Latitude	Longitude
0	Toronto	North York	Parkwoods	43.753259	-79.329656
1	Toronto	North York	Victoria Village	43.725882	-79.315572
2	Toronto	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636
3	Toronto	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763
4	Toronto	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494

London

- Use of the Geopy Library and Nominatim API to obtain London neighbourhoods coordinate.
- Any duplicate coordinates will be dropped.

	City	Borough	Neighbourhood	Latitude	Longitude
0	Toronto	North York	Parkwoods	43.753259	-79.329656
1	Toronto	North York	Victoria Village	43.725882	-79.315572
2	Toronto	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636
3	Toronto	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763
4	Toronto	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494



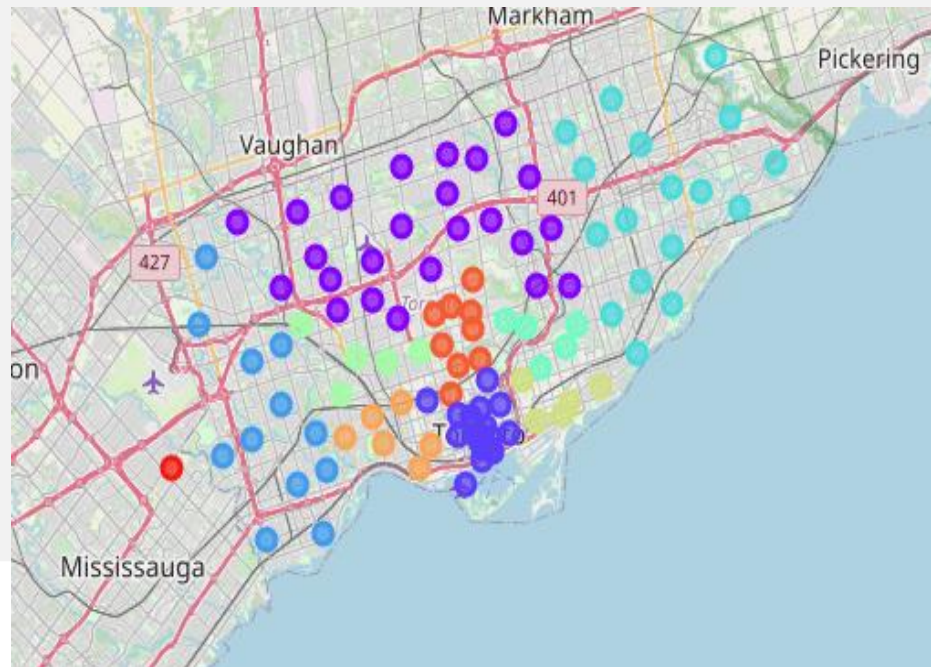
Methodology

Data Visualisation

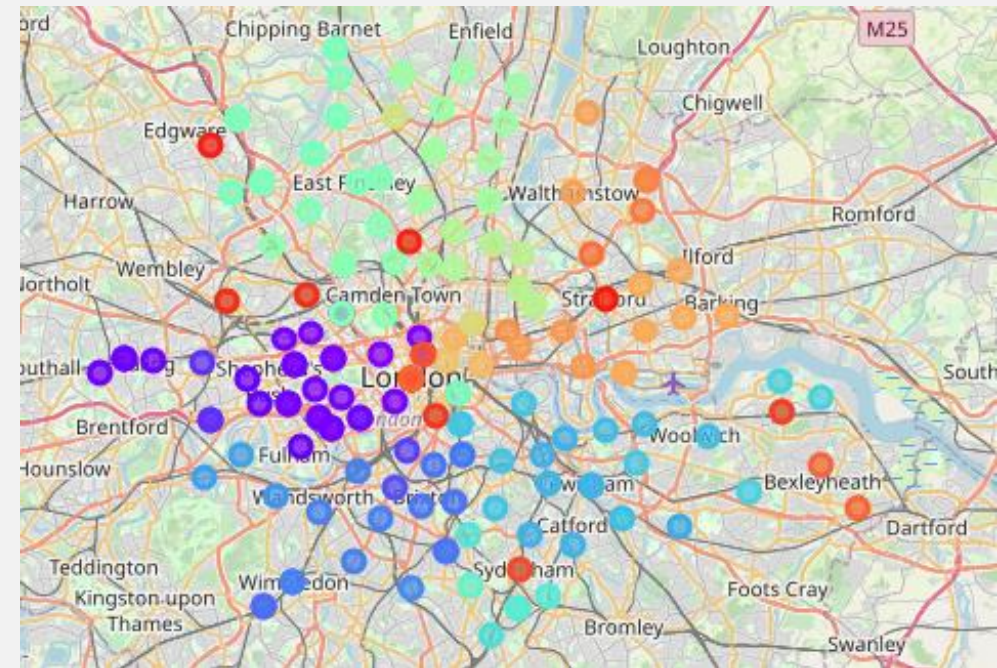
The Folium Library is used to plot a map of all neighbourhoods in each city. The neighbourhoods that are part of the same borough and are drawn with the same colour.



Toronto



London



Methodology

Foursquare API

Foursquare API is used to locate nearby venues within 500 meters of each neighbourhood in Toronto and London.



Toronto

	City	Borough	Neighbourhood	Latitude	Longitude	Venue	Venue Category
0	Toronto	North York	Parkwoods	43.753259	-79.329656	Brookbanks Park	Park
1	Toronto	North York	Parkwoods	43.753259	-79.329656	649 Variety	Convenience Store
2	Toronto	North York	Parkwoods	43.753259	-79.329656	Variety Store	Food & Drink Shop
3	Toronto	North York	Victoria Village	43.725882	-79.315572	Victoria Village Arena	Hockey Arena
4	Toronto	North York	Victoria Village	43.725882	-79.315572	Tim Hortons	Coffee Shop

London

	City	Borough	Neighbourhood	Latitude	Longitude	Venue	Venue Category
0	London	Westminster / Camden	Covent Garden / Charing Cross / Aldwych / St G...	51.51651	-0.11968	Scarves Bar	Hotel Bar
1	London	Westminster / Camden	Covent Garden / Charing Cross / Aldwych / St G...	51.51651	-0.11968	Rosewood London	Hotel
2	London	Westminster / Camden	Covent Garden / Charing Cross / Aldwych / St G...	51.51651	-0.11968	The Hoxton Holborn	Hotel
3	London	Westminster / Camden	Covent Garden / Charing Cross / Aldwych / St G...	51.51651	-0.11968	Sir John Soane's Museum	History Museum
4	London	Westminster / Camden	Covent Garden / Charing Cross / Aldwych / St G...	51.51651	-0.11968	Lincoln's Inn Fields	Park



Methodology

Foursquare API: Cleaning

- After merging the Toronto and London venues dataset into a single dataframe.
- Similar venues names; for example, 'Bar' and 'Hotel bar' that are in the same category 'Bar' are cleaned up.
- A new dataframe is created containing the top 10 venues for each neighbourhood



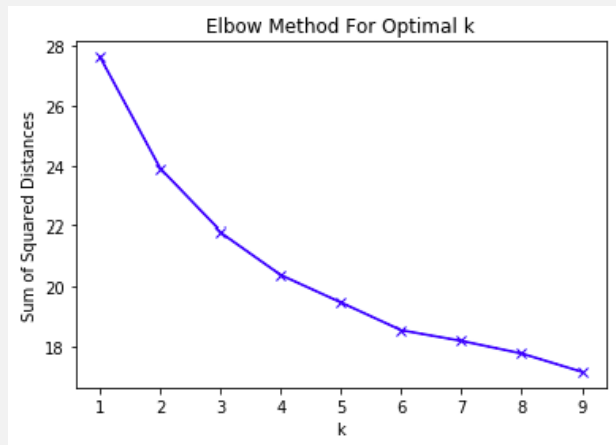
	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Acton	Restaurant	Cafeteria	Pub	Grocery Store	Bakery	Bookstore	Supermarket	Park	Sandwich Place	Organic Grocery
1	Agincourt	Lounge	Restaurant	Breakfast Spot	Skating Rink	Clothing Store	Flea Market	Field	Film Studio	Fish & Chips Shop	Fish Market
2	Alderwood / Long Branch	Restaurant	Gym	Pub	Cafeteria	Pharmacy	Sandwich Place	Skating Rink	Pool	Fabric Shop	Farmers Market
3	Anerley / Penge	Supermarket	Hotel	Grocery Store	Restaurant	Convenience Store	Fish Market	Farmers Market	Field	Film Studio	Fish & Chips Shop
4	Angel / Hackney	Cafeteria	Restaurant	Bar	Pub	Breakfast Spot	Burger Joint	Yoga Studio	Costume Shop	Donut Shop	School
5	Archway / Upper Holloway	Restaurant	Cafeteria	Grocery Store	Pub	Sandwich Place	Farmers Market	Park	Gastropub	Hotel	Bar



Methodology

Machine Learning: K-Means

- K-Means is used to segment and cluster all the neighbourhoods in Toronto and London based on the similarity of the nearby venue types.
- The Elbow method is used to determine the optimal value of K

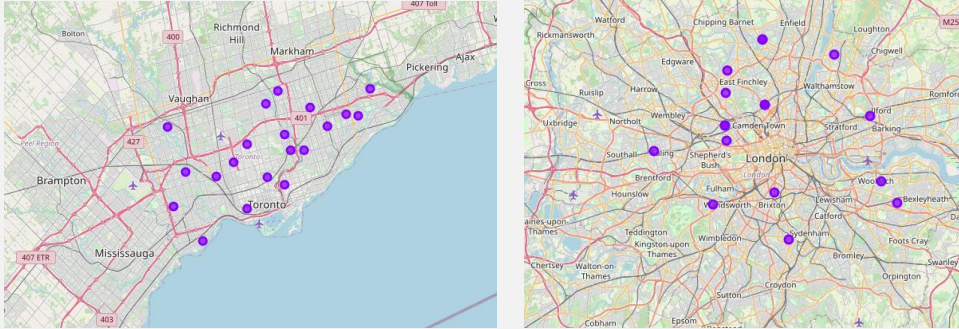


- After testing 4 and 6 as values of K, we get one to two clusters out of the six with one single neighbourhood which isn't useful in the scope of this project when K=6. In conclusion the optimal value of K is 4.

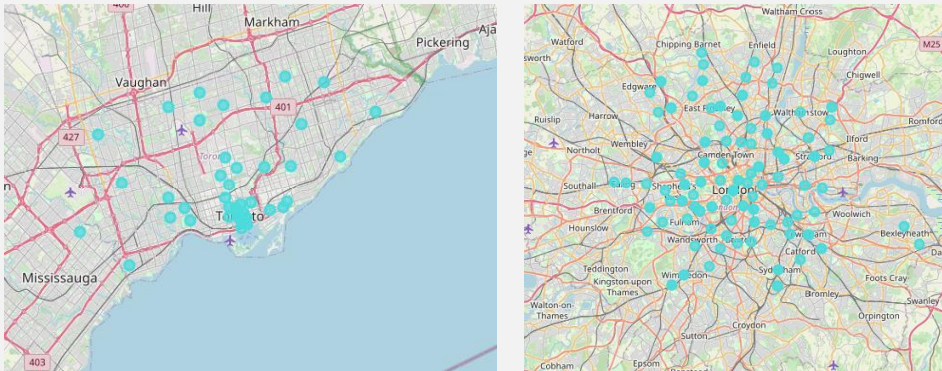


Results

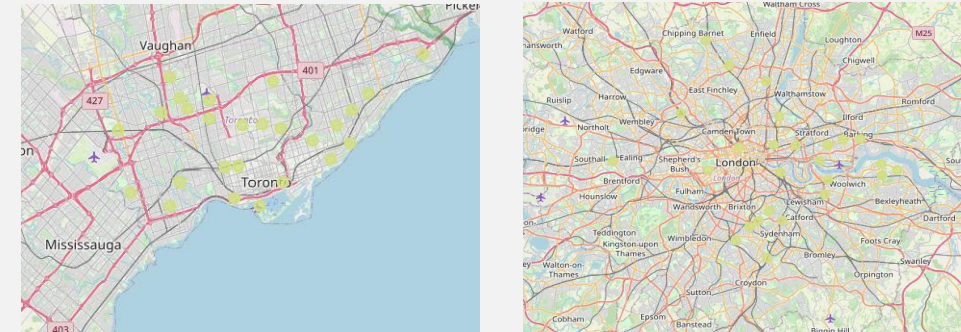
- The number of neighbourhoods in cluster 1 is 37



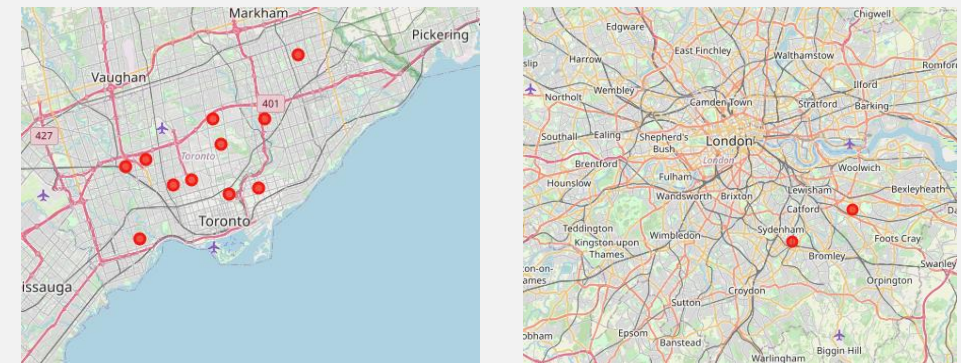
- The number of neighbourhoods in cluster 2 is 170



- The number of neighbourhoods in cluster 3 is 55



- The number of neighbourhoods in cluster 4 is 13

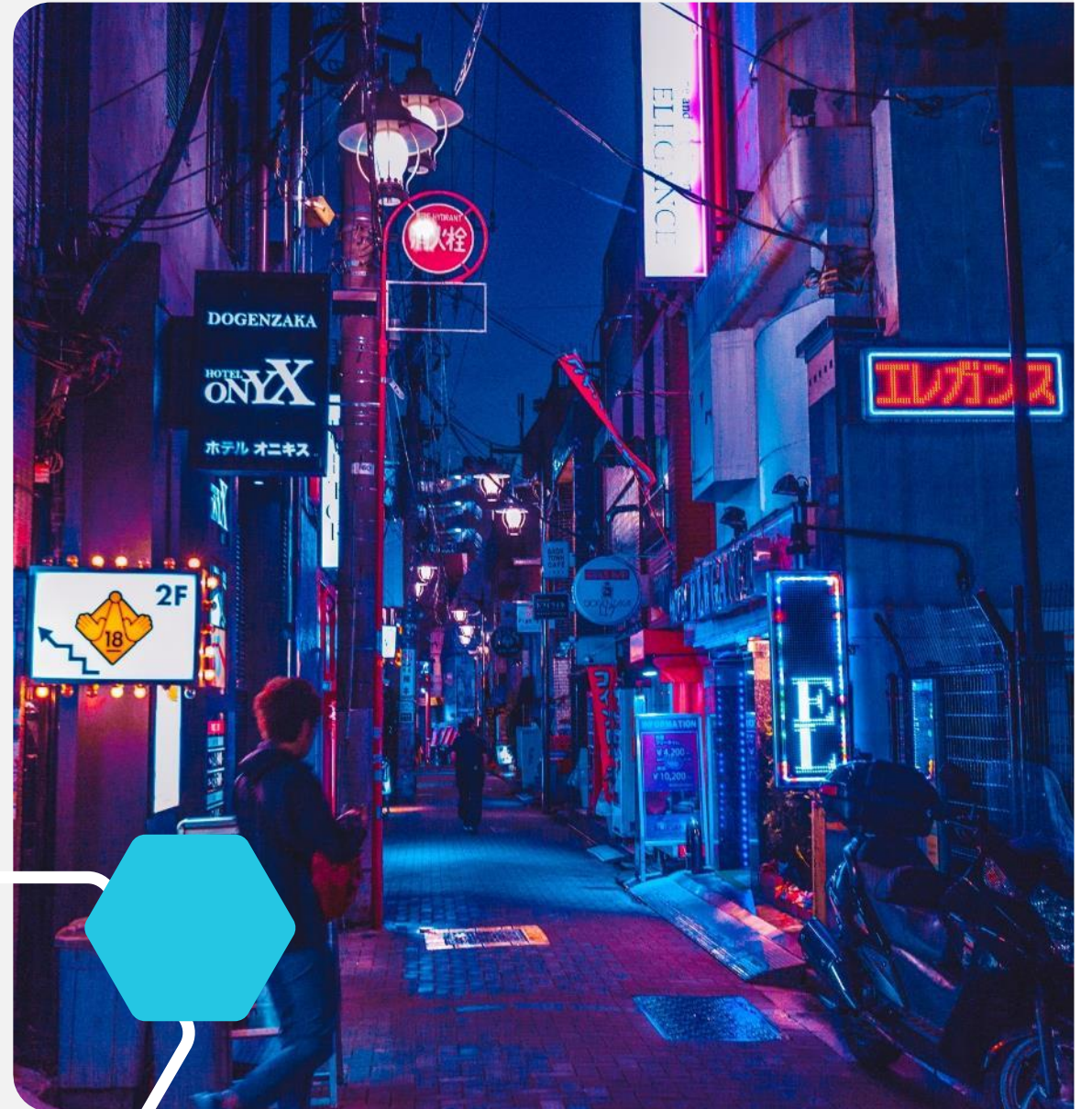


Results

- Clusters 1 and 3 are for those who prefer to settle down in a residential area surrounded by parks, grocery stores, pharmacy and restaurants,
- Cluster 2 is for those who prefer to live in a more crowded area and have access to a variety of venues.

However,

- The majority of the neighbourhoods in lie into cluster 2 which represents 62% of the dataset.
- Cluster 4 represents less than 5% with only 2 neighbourhood from London.
- This problem is due to the unreliable locations provided by the Geocoder package



Conclusion

- We believe that this model will perform better if the location data for London is more accurate.
- This system can be applied to other cities.
- Furthermore, this research can be useful for real estate business if combined with Price Paid data, or city security solution if combined with crime data, etc.





Thank You!