A segmentation tool for neighbourhoods:

The aim of this project is to cluster similar neighbourhoods in the city of toronto together so that people can easily choose where to live based on their interests, lifestyle, and the data of each neighbourhood.
This tool can also be used by real estate agents to find the suitable location for the customer based on their interests and lifestyle.
This project clusters neighbourhoods according to the number venues, like: parks, museums, restaurants, and cafes, each one of them has.

Data used:

I used data from a wikipedia table to get the names of the neighbourhoods of Toronto. I also used the data from an online dataset to get the coordinates of the neighbourhoods based on their postal code. This is the link to the datasets that I used:
"https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M"
"https://cocl.us/Geospatial_data"
Then I used data from the foursquare API to get the number and type of venues at each neighbourhood, then used it to cluster the venues together based on their number.

Methodology:

I used the kmeans algorithm to cluster the neighbourhoods according to the number of venues they each one has. The neighbourhoods were clustered into 5 clusters with similar numbers of venues. I also added the 5 most common venue type of each neighbourhood to the final dataframe to help choose the better place for the customer according to their interests.

Results:

The results of this project is a dataframe with all the neighbourhoods of Toronto, their coordinates, number of venues, most common venue type, borough, postal code, and a cluster label indicating the cluster they belong to. This is how the dataframe looks like(many columns aren't visible due to the width of the page):

Out[16]:

	Neighborhood	1th Most Common Venue	2th Most Common Venue	3th Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bathurst Manor, Wilson Heights, Downsview North	Pizza Place	Sandwich Place	Dessert Shop	Sushi Restaurant	Coffee Shop
1	Bayview Village	Mexican Restaurant	Sushi Restaurant	Trail	Jewelry Store	Donut Shop
2	Bedford Park, Lawrence Manor East	Coffee Shop	Diner	Yoga Studio	Arts & Crafts Store	Café
3	Berczy Park	Grocery Store	Shopping Mall	Bank	Hotel	Park
4	Birch Cliff, Cliffside West	River	Women's Store	Curling Ice	Donut Shop	Doner Restaurant

The dataframe was used to draw a map with the neighbourhoods positioned on top of it and colour coded according to their cluster:

Red: cluster 0

Purple: cluster 1

Light blue: cluster 2

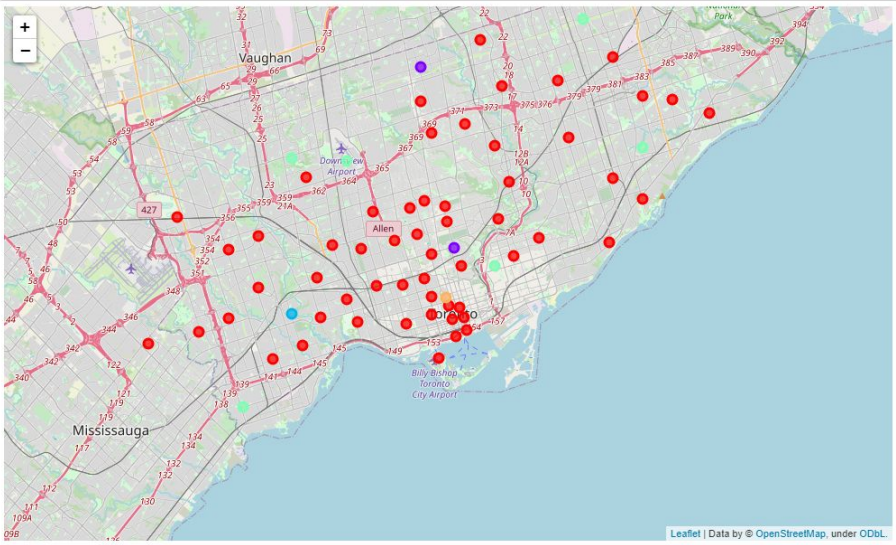
Green: cluster 3

Orange: cluster 4

This is how the data looks when visualized on the map:

```
map_clusters
```

Out[18]:



In [ ]:

Discussion:

The dataframe has some neighbourhoods in the same row because they have the same postal code as the data set I used depended on postal codes to get the coordinates of the neighbourhood. That is why sometimes some neighbourhoods are represented by the same point on the map.
I also noticed the most of the neighbourhoods were in cluster 0 which means that most neighbourhoods are similar to each other in the number of venues they have.

Conclusion:

In conclusion, this tool can be very useful to real estate agents in recommending neighbourhoods to customers or even individuals who want to find a suitable neighbourhood to live in.