

The Battle of Neighborhoods

Comparison between Lisbon and Toronto

1 - Introduction: Business Problem

In this project we compare the neighborhoods of the two cities and determine how similar or dissimilar they are.

The cities that we will use in this project are Toronto, Canada and Lisbon, Portugal.

We will use foursquare API to explore the most common venues in each neighborhood.

Finally we will use k-means to cluster the neighborhoods for each city and determinate the similarity or dissimilar of both cities.

2 - Loading Data and Preprocessing

For Lisbon we use a csv file with the neighborhoods and then Google API to collect the latitude and longitude.

	city	area	Latitude	Longitude
0	Lisboa	Ajuda	38.711312	-9.198212
1	Lisboa	Alcantara	38.705996	-9.182754
2	Lisboa	Alvalade	38.747756	-9.139389
3	Lisboa	Areeiro	38.744617	-9.134370
4	Lisboa	Arroios	38.731691	-9.137314

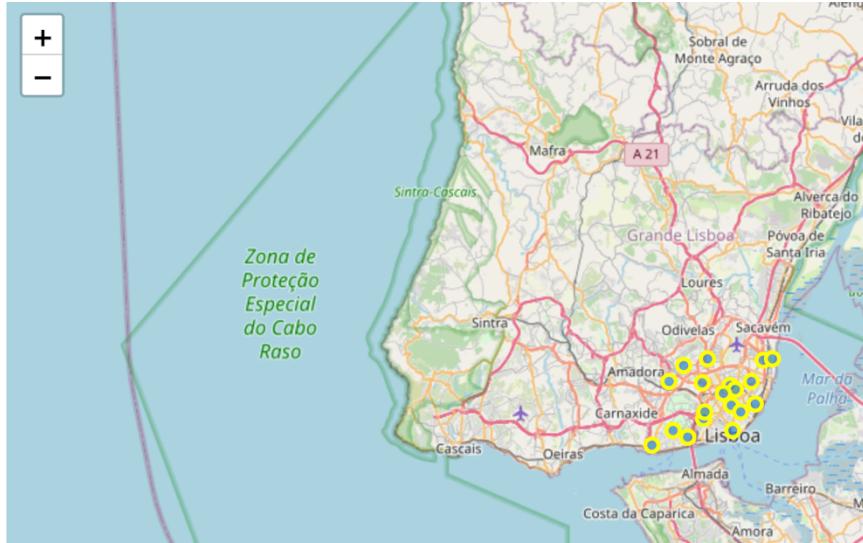
For Toronto we use BeautifulSoup4 library to import the post codes from a wiki page and then merge a csv containing the coordinates.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

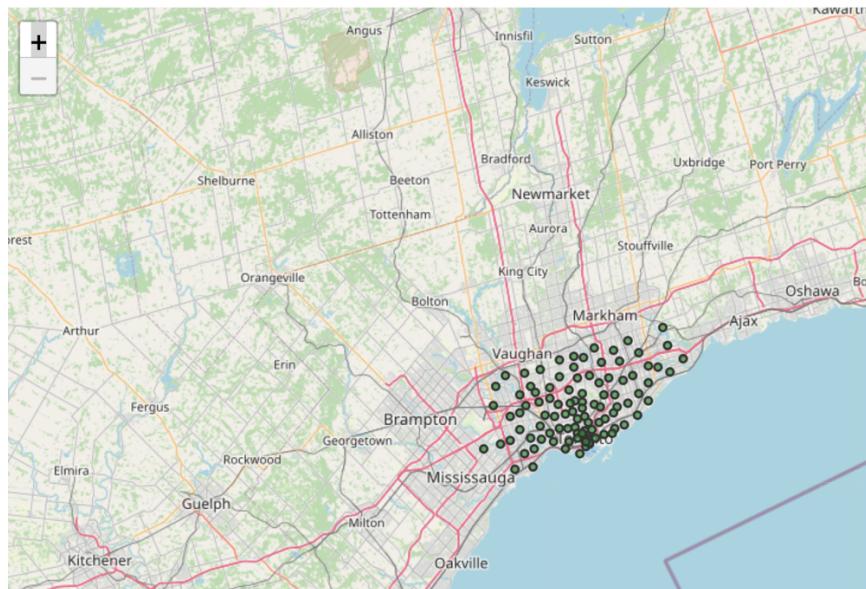
3 - Loading Maps

We use folium to display both city maps with the related neighborhoods.

Lisbon



Toronto



4 - Getting Venues

In order to get the venues and each coordinate for each neighborhood we use Foursquare API and merge the venues into one single table for each city.

We also remove noise data such as building, roads, offices and Bus stops.

Lisbon

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ajuda	38.711312	-9.198212	Palácio Nacional da Ajuda	38.707653	-9.197758	Historic Site
1	Ajuda	38.711312	-9.198212	Mercado do Peixe	38.712692	-9.203276	Seafood Restaurant
2	Ajuda	38.711312	-9.198212	CEDAR (Universidade de Lisboa)	38.714426	-9.197598	Gym
3	Ajuda	38.711312	-9.198212	Ágora	38.712345	-9.194644	Café
4	Alcantara	38.705996	-9.182754	Jardim do Alto de Santo Amaro	38.704570	-9.184478	Garden

Toronto

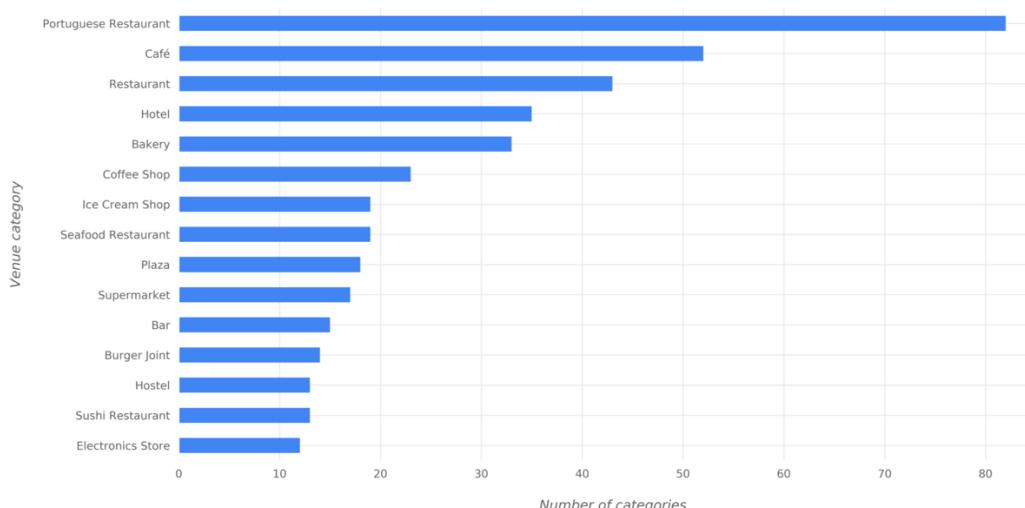
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Scarborough Historical Society	43.788755	-79.162438	History Museum
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store

5 - Common Categories

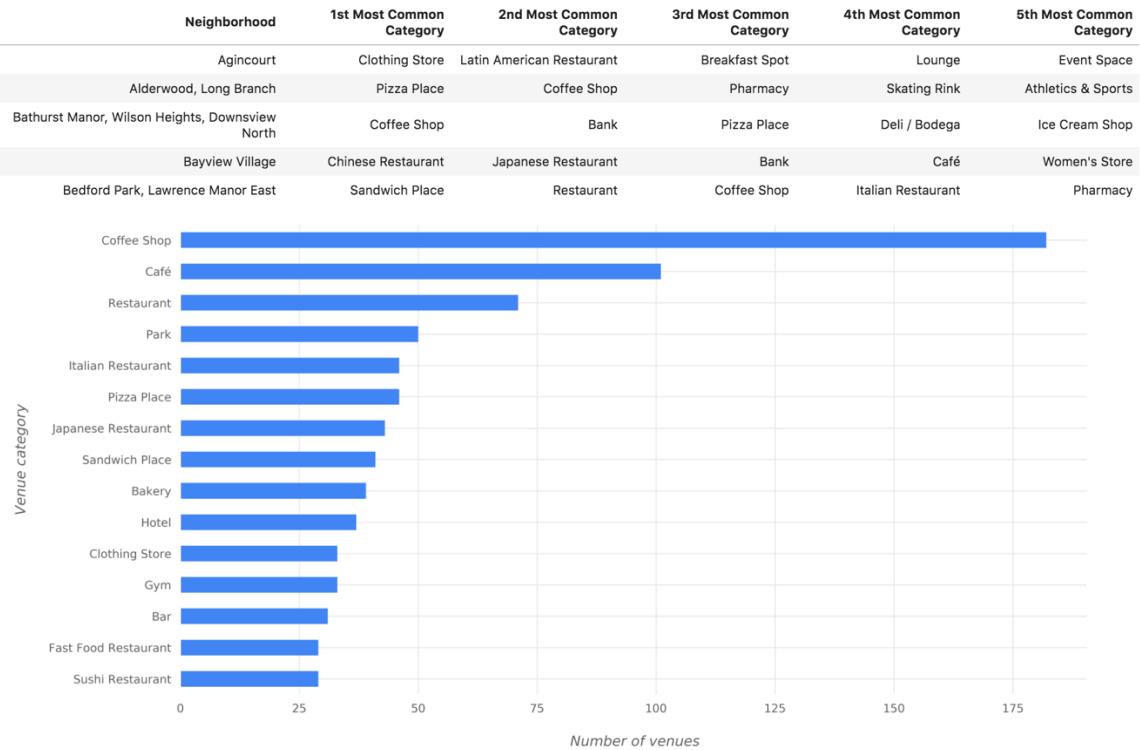
We display the 5 most common categories for each neighborhood as a support for the next step, clustering. We also have checked the most common categories in each city.

Lisbon

	Neighborhood	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category
0	Ajuda	Gym	Historic Site	Café	Seafood Restaurant	Yoga Studio
1	Alcantara	Restaurant	Café	Mediterranean Restaurant	Bakery	Coffee Shop
2	Alvalade	Bar	Portuguese Restaurant	Restaurant	Hotel	Café
3	Areeiro	Café	Bakery	Portuguese Restaurant	Plaza	Restaurant
4	Arroios	Portuguese Restaurant	Hotel	Hostel	Café	Supermarket



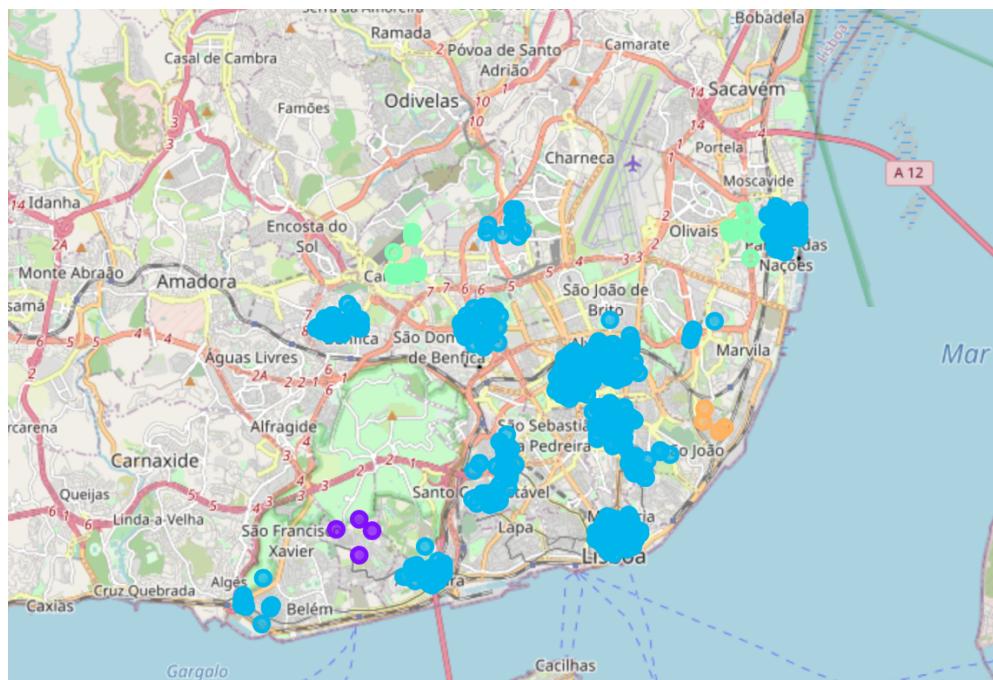
Toronto



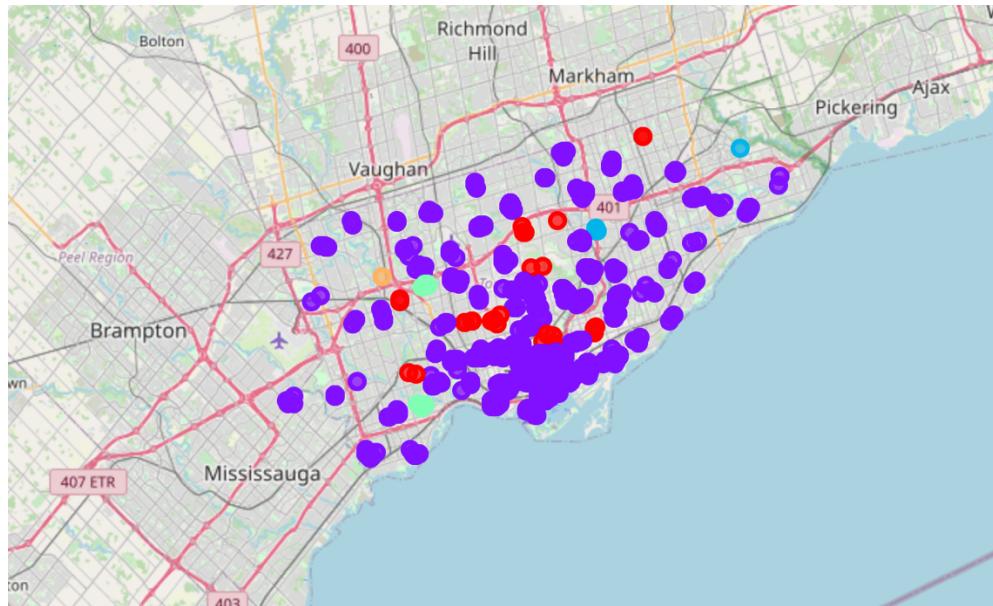
6 - Clustering

We use K-means as an algorithm to define our clusters and we use 5 cluster for each city to find the clusters for each city.

Lisbon



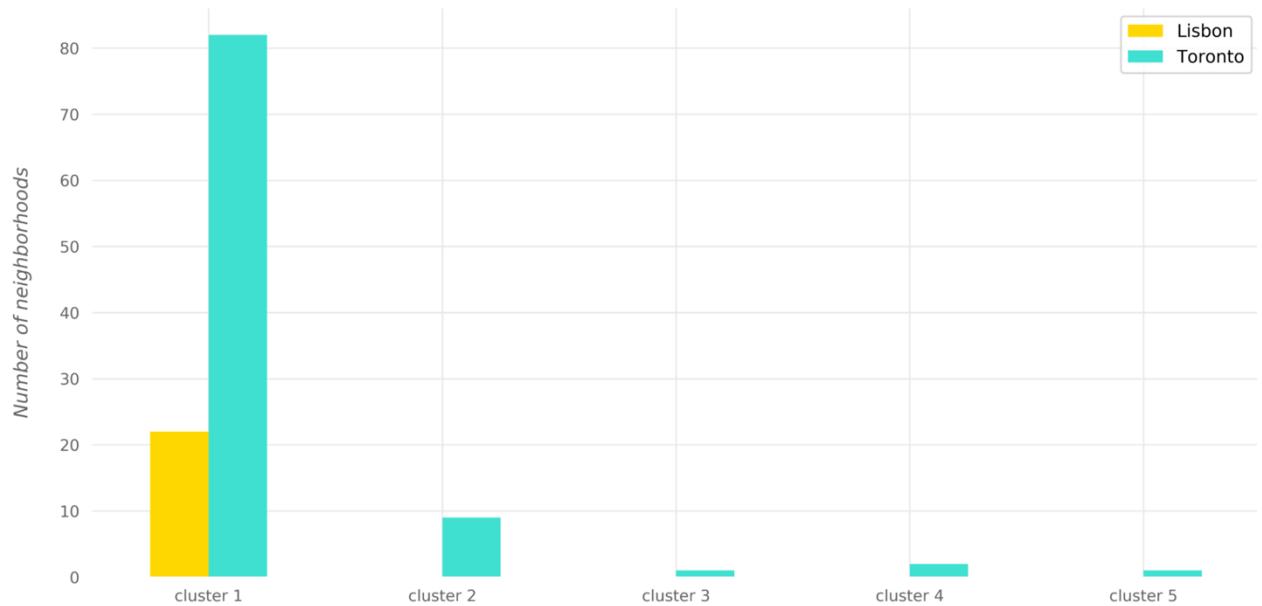
Toronto



7 – Results

We group the neighborhoods for each city into one single table and build a new K-means to see the similarity between the neighborhoods of each city.

We investigate each cluster and then display a plot to better understand the similarity or dissimilarity.



8 – Conclusions

The main goal of this project is compare two cities, Toronto and Lisbon and evaluate their similarly.

In conclusion, we can observe according to our clusters (5) that they differ from each other significantly. However cluster 1 is the one which contains more neighborhoods for each city. One reason can be related with the common categories in both cities such as coffee shops and restaurants however Lisbon is much smaller and has for example typical Portuguese restaurants which Toronto doesn't have (common to have Japanese and Italian instead).