The Battle of Neighborhoods

Periklis Kydonakis

The case and purpose

- A large restaurant company that owns several brands and operates in several European countries seeks the opportunity in expanding its business in North America and more specifically in Toronto, Canada and New York City, USA

- The purpose of this project is to analyse and compare the cities of New York and Toronto to identify the most suitable neighbourhoods for our client to expand its business. Furthermore, the purpose is to distinguish and classify the several types of restaurants currently operating in those cities

Data sets

- Neighborhood Data with geospatial information for New York and Toronto
- Venues' data from Foursquare database
- Categories and their hierarchy from Foursquare database

Data retrieval

Some data were already available and others acquired from Wikipedia and the largest amount of data from Foursquare through API calls

Data preparation

All data was formatted and prepared for further analysis with the use of Python libraries and functions

Data understanding

The focus was given mainly on the data regarding venues and categories as the geospatial data

For New York, the number of venues collected was 9.751 in total 5.078 of them being under the food category

For Toronto, the number of venues collected was 2.110 in total 1.161 of them being under the food category

TOP 10 restaurant categories in New York

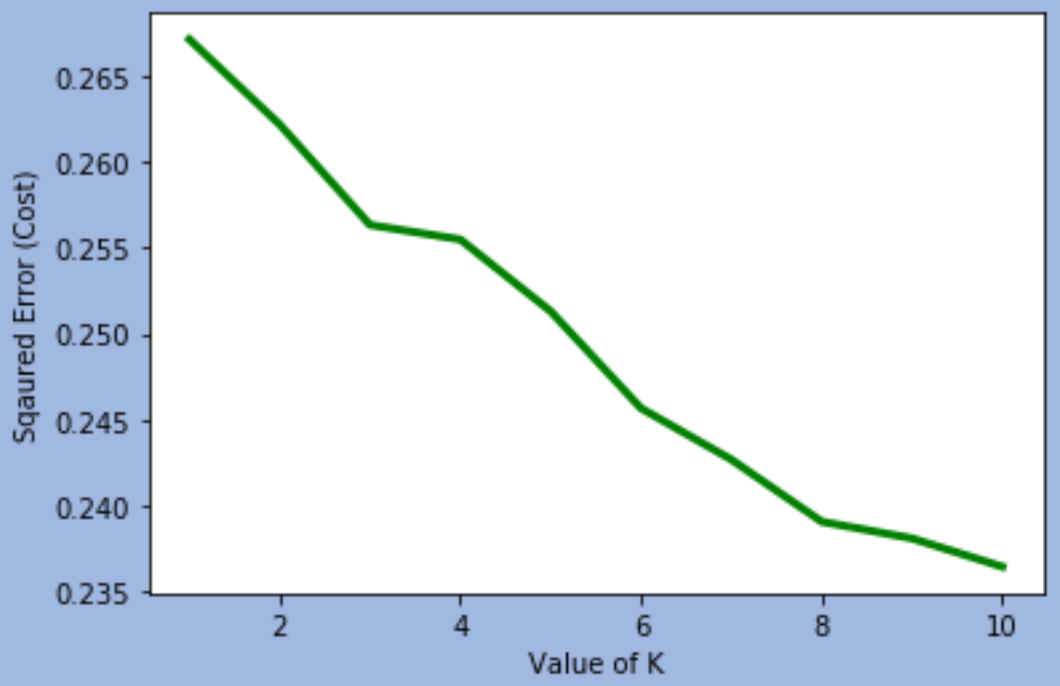
Venue Category	Venue
Pizza Place	434
Italian Restaurant	297
Coffee Shop	289
Deli / Bodega	250
Bakery	214
Chinese Restaurant	213
Sandwich Place	187
Mexican Restaurant	172
Donut Shop	163
Café	159
Count of TOP 10	**2378**
46.83% of total	

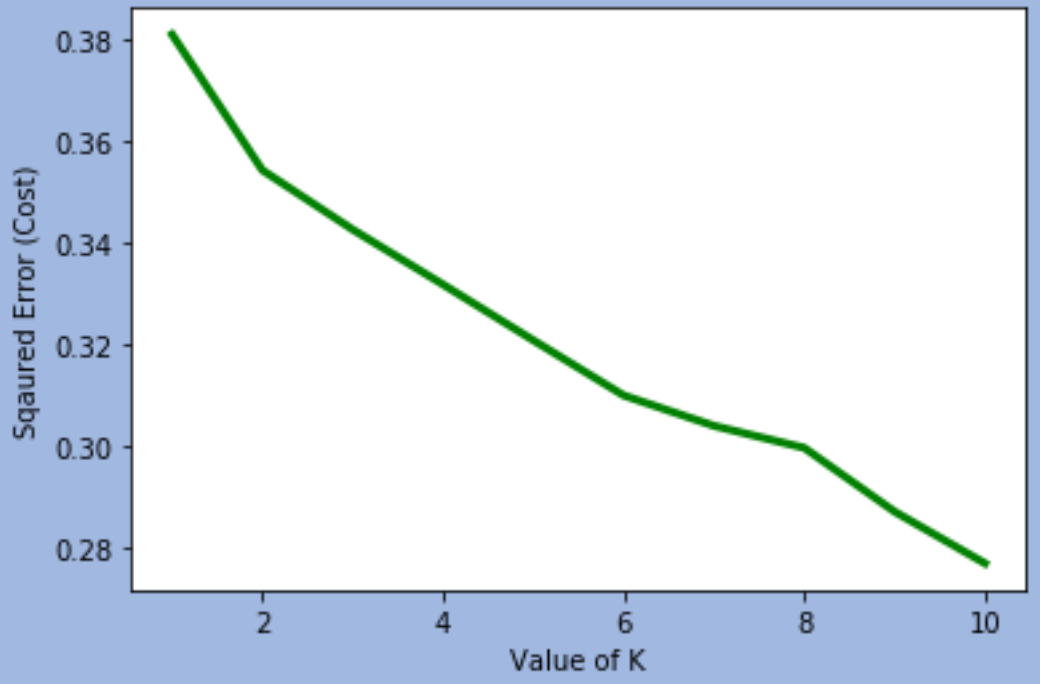
TOP 10 restaurant categories in Toronto

Venue Category	Venue
Coffee Shop	186
Café	98
Restaurant	68
Pizza Place	46
Italian Restaurant	46
Japanese Restaurant	42
Bakery	40
Sandwich Place	39
Sushi Restaurant	29
American Restaurant	28
Count of TOP 10	**622**
53.57% of total	

Modelling

New York and Toronto Elbows and Silhouette scores



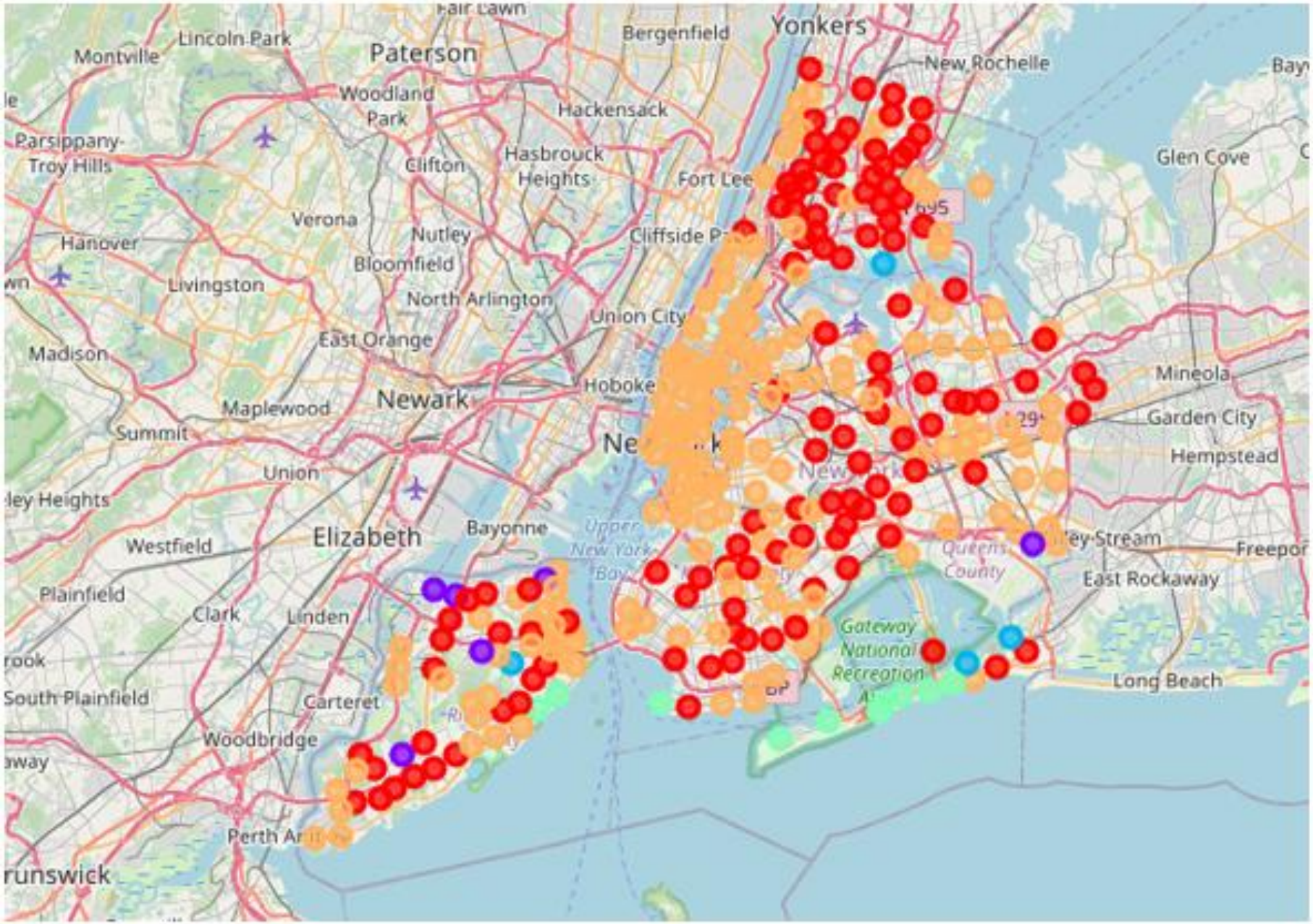



```
For n_clusters = 2, silhouette score is 0.5236801422751508)
For n_clusters = 3, silhouette score is 0.09529805887971099)
For n_clusters = 4, silhouette score is 0.5265485356597408)
For n_clusters = 5, silhouette score is 0.030516352395696653)
For n_clusters = 6, silhouette score is 0.033134899470598195)
For n_clusters = 7, silhouette score is 0.033259823939917985)
For n_clusters = 8, silhouette score is 0.03500437773899958)
For n_clusters = 9, silhouette score is 0.03699078931763086)
For n_clusters = 10, silhouette score is 0.03175896567437805)
```

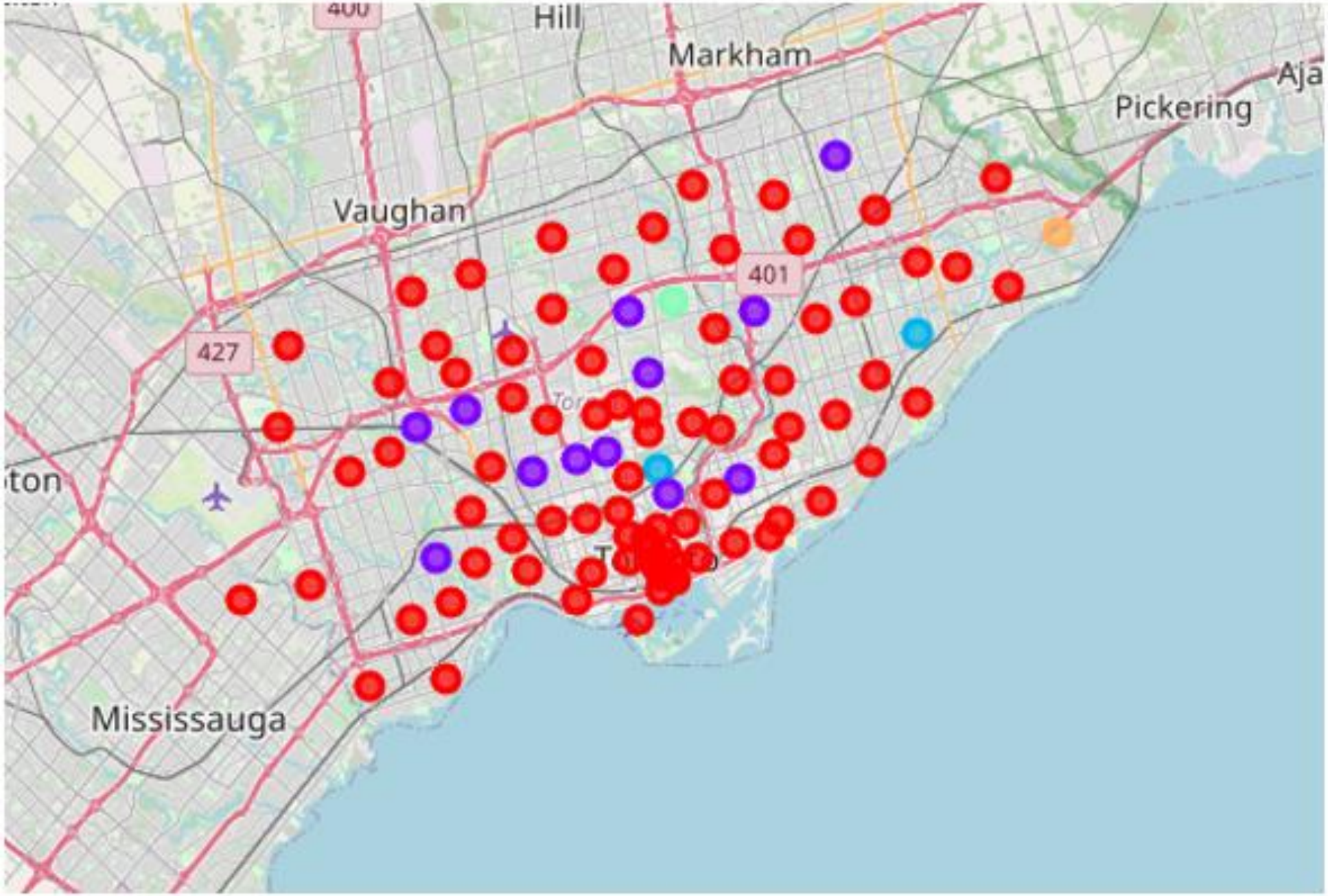
```
For n_clusters = 2, silhouette score is 0.3018523882819826)
For n_clusters = 3, silhouette score is 0.2916324955210721)
For n_clusters = 4, silhouette score is 0.3003125718298839)
For n_clusters = 5, silhouette score is 0.30596654368247284)
For n_clusters = 6, silhouette score is 0.315103384915657)
For n_clusters = 7, silhouette score is 0.15138692180971608)
For n_clusters = 8, silhouette score is 0.10512880922234978)
For n_clusters = 9, silhouette score is 0.14205795339101646)
For n_clusters = 10, silhouette score is 0.11994526058467776)
```

The final number of clusters was decided to be 5 for both New York and Toronto

New York's dominant clusters 1 and 5 appear in red and orange below.



Toronto's dominant clusters 1 and 2 appear in red and purple below



Clustering Analysis

- It can be clearly seen that in both cities neighborhoods present wide similarities as both have two dominant clusters which are 1 and 5 for New York and 1 and 2 for Toronto so let us examine those clusters in a little bit more depth

- New York's clusters 1 and 5 present a wide variety in venues both for restaurants and other categories. We have included for those clusters the first 10 rows each row having the 10 most common categories on that cluster

Results

The client would like to open restaurants that present high levels of popularity in both cities as its brands are strong and they would like to target the largest proportion of population possible. Thus, the restaurant types that match the requirements are Pizza Places, Italian restaurants and Coffee Shops. Pizza Place and Italian Restaurant are the two categories that were chosen

As far as the second question is concerned, we found that the client has I wide variety of neighborhoods to choose from in order to pick the place for the new restaurants. As a matter of fact, they can choose many different neighborhoods to open the required number of places. For New York Bronx and Manhattan should be included and for Toronto Downtown is the strongest candidate

Conclusion

Overall, it as a great experience. We had the opportunity to work as data scientists on a project that gave us the chance to experiment on a large dataset and use a wide variety of tools for data gathering, formatting and analysis. Furthermore, we used mapping libraries and bar charts for visualisation. To sum up the findings of that project can be used as a basis for further analysis and / or in similar cases. Both New York and Toronto are great cities to expand business in a wide range of activities. Their multiculturality is apparent that is combined with the large numbers of population and visitors. I hope that you enjoyed reading this and that it will help someone in taking that some steps forward.