

## Data science capstone project – The battle of the Neighborhoods:

### **Introduction/Problem definition:**

A businessman has lived in New York, United states and now lives in Toronto, Canada. He has been asked to work for Fictive incorporation in Europe. They have offices in Paris and Berlin. He has a beautiful wife and two children. In New York and Toronto they have lived in nice neighborhoods. He does want to work in Europe, but in a similar Neighbourhood. The problem that we try to solve is to find a similar neighborhood for our businessman in either Berlin or Paris.

### **Data that will be used:**

Data containing boroughs of new York:

[https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

Data containing boroughs of Toronto:

[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

Data containing boroughs of Berlin:

[https://en.wikipedia.org/wiki/Boroughs\\_and\\_neighborhoods\\_of\\_Berlin#External\\_links](https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin#External_links)

Data containing boroughs of Paris:

[https://en.wikipedia.org/wiki/Arrondissements\\_of\\_Paris](https://en.wikipedia.org/wiki/Arrondissements_of_Paris)

### **Additional Data:**

Additional data will be gathered by using the Foursquare API.

### **Use of data:**

In this project we will gather venue information about the different Boroughs and neighbourhoods in the four different cities. This will be done by using the foursquare API. Some of the borough data has already been collected, other data will be collected by use of Beautifulsoup and/or pandas.