**Setting up a new restaurant in New York City - Report**

**The Problem**

The Business Problem that I wish to help solve is basically in which neighborhood of New York city should a cook/business man build his new restaurant, given the fact that Manhattan is already a place filled with many many restaurants of several cultures. Thus, the idea of a new establishment in this city would already be huge challenge to undertake and even more so for the business to thrive.

According to an article made by Nick Hines from the Vinepair website (<https://vinepair.com/booze-news/new-york-restaurants-eat-at-every-on/>) “you can’t walk a New City block without passing a restaurant”. It even states that “80 percent of restaurants fail within five years”, so it would seem very difficult to get a new restaurant business going in this city.

A study of venues in other metropolitan areas around the world will help inform which types of restaurant are less common in New York city, thus improving the possibility of success in a city already filled with so many restaurants.

**The Data**

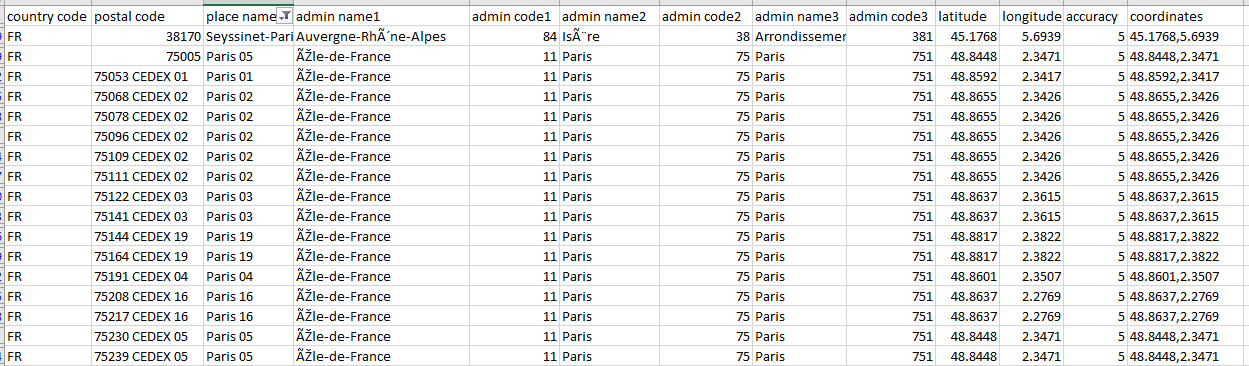
Foursquare API will be the chosen API to collect the data related to the venues for each geographical point.

To gather the information about geographical location (postal code, neighborhood, borough), ***Open Data Soft*** API (https://data.opendatasoft.com/pages/home/) was used, which is very simple to use, by simply writing down the country and city that you wish to research.

The chosen cities were New York City (which was already researched upon in the Laboratory for the Capstone section: https://cocl.us/new\_york\_dataset), Toronto (also researched as a deliverable in this final course: 'https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M'), Paris, Berlin and Porto.

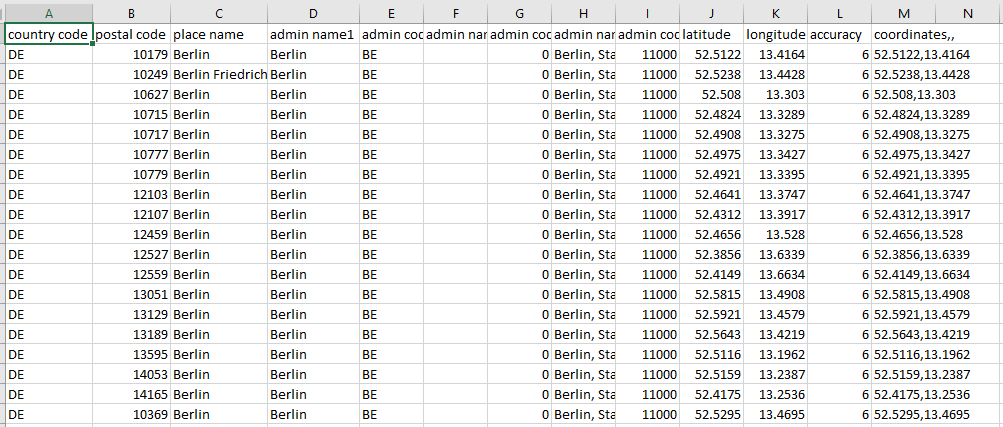
These last three were the ones that were researched through Open Data Soft. Paris and Berlin were chosen, since they definitely are known as multicultural cities with a significant diversity of venues and a considerable number of people living in them. Porto is sort of an outlier, but it’s an emergent city of Portugal, with lots of tourists coming every year and with evolving throughout time with new businesses being implemented.

The information for Toronto and New York is already in the “correct” format but in the case of the other cities the data is categorized in another format, for example, for Paris:



For these cases the Neighborhood and Borough were considered to be the same thing, which is the column “Place Name”, and of course the columns with the postal codes and the latitude and longitude geographical coordinates were kept. All the other columns were not used.

For the case of the Berlin data, the Postal Code was used as the Neighborhood name:



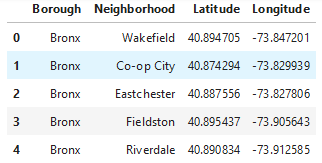
Notice how almost all of the entries have “Berlin” as the place name, so to be able to get the venues for each location, the postal code, which is unique, will be used to distinguish each place from the other.

In the case of Porto’s data, the case is a bit more complicated:



The data found is a lot more scattered throughout the entire Porto district, thus the data is not necessarily centered in the city’s core, but comprises a much larger area. But it will prove to be an interesting observation to compare this particular Portuguese city with metropolitan areas from powerful countries.

The Source data for the NY City locations should have the following format (after interpreting the JSON file):



Finally, the data for the venues to be analyzed should have the following format:

