

Exploring & Profiling Adjacent Neighborhoods

Week 01 - Submissions

Coursera Capstone Project for
IBM Data Science Specialization

Christos Glymidakis – enaeuro@yahoo.com



Introduction – Business Problem

Problem Overview

It is often the case that someone would like to know “what is around” a certain location of explicit interest. For example, when we visit a new place we only know the address of the hotel we have booked, or of the house of the friends who are going to accommodate us. However, most commonly we will not know anything about the neighborhoods around this single point. To overcome this challenge, in this capstone project I have proposed the following advisory solution:

- Select the single pin point on a map, around which we would like to do some profile exploration. For this single point we will need either a valid and complete address or the exact coordination (latitude and longitude).
- Automatically generate a grid of 5x5 points (this can be shrank or expanded, depending on how far away we would like to explore), which have a distance of approximately 1-2 km from each other. This grid essentially covers an area of roughly 16-64 square kilometers (rectangular with almost 4-8 km each side).
- Using the Foursquare API fetch the most common venues for each area and consider their venue type.
- Apply a clustering algorithm with an average of 5 distinct profiles (i.e. $k=5$) and therefore, group the areas into similar profiles.
- In this way we can tell if the neighborhoods around the focal point are similar to the one where we plan to stay, and towards which direction we might have alternative patterns, landscapes and type of venues.

Data Overview

Data Overview

To solve this challenge we will need the following data:

- An initial,=]
- valid and complete address with which we can call the Google geocode API and retrieve the exact coordinate of the point of interest.
- The shape and size of the grid we would like to explore around the central point. As a rule of thumb, we can use the following steps:

```
AT LATITUDE 40 DEGREES (NORTH OR SOUTH)
One degree of latitude = 111.03 km
One minute of latitude = 1.85 km
One degree of longitude = 85.39 km
One minute of longitude = 1.42 km
```

- Knowing the points of the grid allows us to use the Google API and retrieve the names of the places on the grid, like the street, the neighborhood, the city and the country. Among the various addresses that might fall into the same neighborhood, we simply keep only the first. Anyhow, this is mostly for informative purposes.
- For each one of the points on the virtual grid, we call the Foursquare API and retrieve the containing venues, with a limit of 100 per request.
- Using the venues type, we create a profile for each neighborhood and finally we run a clustering algorithm to derive the basic, distinct profiles of all the neighborhoods around our main point of interest.

Example of a virtual grid on a map

Example of a
virtual grid on a
real map

