# What and Where Invest?

Yassin Belhareth

January 2019

## 1 Introduction

To start a business project, a comprehensive study must be carried out. One of the most important things to consider is, the place of the project, There are those who prefer overcrowding, there are those who reject it. On the other hand, There are those who are hesitant among several cities, sometimes the hesitation is in the type of project. In my study, the objective is to extract the adequate information and facilitate it for the investor in order to choose the right place and type for his project. I will analysis the neighborhoods at the level of the business places and not business places that give service for people, to do this, it will be relying on a clustering algorithm in order to segment the neighborhoods in similar clusters, more details in the methodology section. The approach will be applied on two famous cities, Paris and London, they are geographically close, but each one has its specificity.

## 2 Data Description

I need in this study for some data about venues in the neighborhoods, there are many providers could give data about, for example Yelp[1], Google Places[2] and Foursquare[3].it were chosen the Foursquare API because it is quite straightforward and easy to use. It organizes its venues in categories and sub-categories, the bar chart 1 shows the number of sub-categories for each category, the food category contains the maximum number, followed

---

[1]https://www.yelp.com/developers
[2]https://cloud.google.com/maps-platform/places/
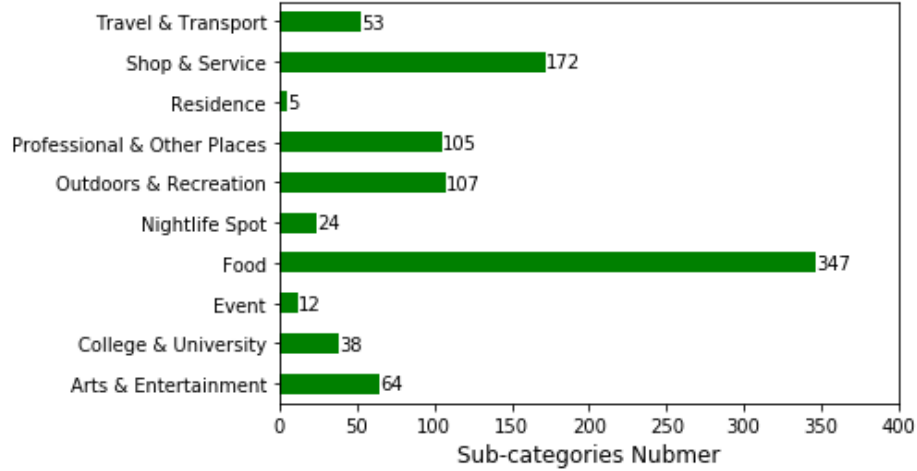[3]https://developer.foursquare.com/places-api

Figure 1: Number of Sub-categories by Categories in Foursquare

by the shop and service category, the sub-categories represent all types of categories in the world (e.g. the category food contains: French Restaurant, Japanese Restaurant, pizza place, etc.). In other hand, I should collect the neighborhoods's names of Paris and London. Wikipedia Pages ([4] and [5]) contain the required data for two cities, but the name of the neighborhood is not enough to collect venues data, we must have the latitude and longitude coordinates for each location, fortunately, if you click on any location in Wikipedia pages, you will have detailed information on it, among them its geographic coordinate (the figure 2 shows an example). For Paris, it will be used all its 20 districts, but for London, since, it is a big city ( 1500 square km ), it will be used just London post town with 32 boroughs.

---

[4]https://fr.wikipedia.org/wiki/Liste_des_quartiers_administratifs_de_Paris

[5]https://en.wikipedia.org/wiki/List_of_areas_of_London

| Location | | London borough | | Post town | |
|---|---|---|---|---|---|
| Abbey Wood | | Bexley, Greenwich [1] | | LONDON | |
| Acton | | Ealing, Hammersmith an | | | |
| Addington | | Croydon[2] | | | |
| Addiscombe | | Croydon[2] | | | |
| Albany Park | | Bexley | | | |
| Aldborough Hatch | | Redbridge[3] | | | |
| Aldgate | | [4] | | | |
| Aldwych | | Westminster[4] | | | |
| Alperton | | Brent[5] | | | |
| Anerley | | Bromley[5] | | | |
| Angel | | Islington[2] | | | |
| Aperfield | | Bromley[5] | | | |
| Archway | | Islington[6] | | | |
| Ardleigh Green | | Havering[6] | | | |
| Arkley | | Barnet[6] | | BARNET, LONDON | |
| Arnos Grove | | Enfield[6] | | LONDON | |
| Balham | | Wandsworth[7] | | LONDON | |
| Bankside | | Southwark[8] | | LONDON | |

Figure 2: Example of London Neighborhoods and How to Get Geographic Coordinate