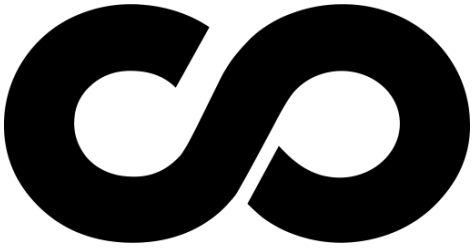

IBM Data Science Professional Certificate

Applied Data Science Capstone

London Businesses Benchmarking

Oussama KIASSI

16 July 2020

I Introduction

I.1 Background

As the capital of the United Kingdom, London attracts around 30M tourist from all over the world every year. Its tourist industry is ever-growing, enlarging the country's economy. However, some tourists face some very hard experiences. They get harassed, smuggled, aggressed and encounter other crimes. While we cannot completely control the city's safety, they naturally started to look for safe zones to get distracted and to live the English way. So, to fulfill their needs businessmen and small business owners started investing in the safest boroughs of the capital. Thus, our project plays the role of a benchmarking platform that gives the formers an idea on businesses in those safe areas. It primary aims to cluster common venues of London's safest zones.

I.2 Interest

This project will benefit tourists by displaying the common, yet safest attractions. Consequently, it will allow a greater experience for the formers, while expanding London's tourist industry. Moreover, it will be of a great help to investors. Knowing the frequent venues in the safest boroughs of London, they will start their own businesses with considerable confidence.

II Data requirements

We used in the present project the following datasets:

- London boroughs' populations from June 1991 to June 2019 (10 years spaced) scrapped from `https://www.citypopulation.de/en/uk/greaterlondon/` (34 rows, 7 columns).
- London crimes by borough in the last 24 months (Metropolitan data), downloaded from `https://data.london.gov.uk/download/recorded_crime_summary/d2e9ccfc-a054-41e3-89fb-53c2bc3ed87a/MPS%20Borough%20Level%20Crime%20%28most%20recent%2024%20months%29.csv` in a CSV format (1569 rows, 27 columns).
- London boroughs geographical borders, downloaded from `https://skgrange.github.io/www/data/london_boroughs.json` as JSON file .

- London areas, scrapped from `https://en.wikipedia.org/wiki/List_of_areas_of_London` (533 rows, 6 columns).

- London's venues requested from Foursquare API as a JSON file (15560 rows, 7 columns).

III Description

This project will be divided into two parts: the first one consists of visualizing boroughs by crime index and the second answers businessmen and small business owners' questions about business in the safest zones of London.

First and foremost, we will clean the population dataset from unnecessary columns and rows. Afterwards, we will preprocess the London crime dataset. We will only keep violations that are considered as crimes. Then we will group boroughs by those crimes. Next, we will compute crime index of each borough so that we could visualize it on a choropleth map.

After that we will scrap London areas. At that point we will only keep the safest ones to make on them a Foursquare API request. And to get the most common venues, we will compute their frequency by location. Furthermore, we will apply a K-Means clustering algorithm on the final data frame. Finally, we will display the clusters and the platform.