Best boroughs in London to live it

Project: IBM Capstone

By: Umair Arif

Date; 24th December 2020

# 1-Introduction

No London borough Is quite the same as another. There are 32 boroughs in total, they all have different attractions, architecture styles and atmospheres which makes them feel almost like a small cluster of different towns and cities.

We have been approached by a magazine, who wants to publish an article on the best boroughs to live in their upcoming edition. They have tasked us to determine the quality of life in each borough and the most ideal boroughs for outdoor activities, shopping destinations, nightlife and dinning.

In this report we will extract the data on the transport accessibility, happiness score, pollution, crime rates and availability of green spaces in each borough to determine the quality of living, followed by using Foursquare-API to construct a model to determine the most popular venues in each boroughs.

# 2-Data

The data needed for this project is extracted from three databases i.e. the Greater London Authority Datastore, Wikipedia and Foursquare API. Once the data from these databases is imported on python interface, they will undergo the process of formatting and cleaning to remove unwanted elements and to prepare the DataFrame for in depth analysis

## 2.1Quality of life analysis

Table

Description automatically generatedThe data required for quality of life analysis is obtained from Greater London Authority Datastore, which consists of data regarding the happiness score, transport accessibility, pollution, crime rates and greenspace available in each borough. The data can be downloaded for the Datastore as an Excel format file, which can then be opened on python interface using pandas.

Figure 1: Shows how the data will look on Jupyter Notebook.

## 2.2 Population and geographical coordinates

The geographical coordinates and the population for each borough is extracted from Wikipedia page. The data is loaded on python using Beautifulsoup. Once the data is loaded on the Jupyter Notebook, the DataFrame will undergo cleaning process to remove unwanted columns and texts such as notes annotation, to split the geographical coordinates in two columns namely, latitudes and longitudes.

Graphical user interface

Description automatically generatedFigure 2: raw data from Wikipedia

Table

Description automatically generated

Figure 3: Data after undergoing formatting and cleaning.

## 2.3 Merging the DataFrame

Once both the tables are cleaned and filtered, both the tables will be merged to form one database

Table

Description automatically generated

Figure 4: Shows the merged DataFrame

## 2.4 Venues

The venues and venue category data is obtained from Foursquare API. The sample of the data on python interface is shown in figure 5.

Graphical user interface, application

Description automatically generated

Figure 5: sample of what the data extracted from Foursquare looks like.