Introduction

Tourism is one of the biggest industries throughout the world and as long as the flights from country to country are increased, more and more people visit other countries either for leisure or business.

The capital of United Kingdom is both an economical and a touristic center and as a result London attracts around 30 million visitors from around the world every year. Consequently, the demand of accommodation is high and a large number of venues offer this convenience, from hotels to AirBnb apartments and hostels to student rooms.

Business problem

A client company which would like to expand its business to the accommodation industry has asked us to evaluate the possibility of opening a new hotel in London. Using data science methodology and analysis as well as machine learning techniques, this project aims to provide suggestions for the optimal location which a new hotel should be start welcoming visitors.

Data

Due to the fact that the greater London covers an area of approximately 1,570km, for this problem, it is necessary at first step to locate the districts/neighborhoods. The following link to the corresponding Wikipedia page has been used to get the data according to the Postal Codes of districts: https://en.wikipedia.org/wiki/List_of_areas_of_London

The above data will be scrapped in proper form and then it will be converted to a data frame in order to be suitable for processing in Jupyter notebook. Afterwards, we will reduce the data frame to London only, due to the fact that the areas of great distance from the city of London are not suggested to opening a new hotel.

Afterwards, based on the name, we will retrieve the geographical coordinates of each neighborhood using geocoder and a new data frame will be created with this information.

In order to explore the neighborhoods, Foursquare API will be used to get the latitude, longitude and location of all of the hotels in London area.