

Data Understanding

There will be two types of data that need to be merged for a master data set. The project needs the names and locations of each borough in London, UK. Additionally, population and surface area of each borough will play an important role in the final decision.

On the other hand, it needs the characteristics of each borough, such as the number of unique businesses. By collecting the data, boroughs can be clustered and the number of locations for the home improvement business can be reduced from total number of boroughs to a few greater areas.

The data about boroughs and related information will be acquired through web scrapping from Wikipedia. The following page has all the relevant information about London, UK: https://en.wikipedia.org/wiki/List_of_London_boroughs

On the other hand, the data about the characteristics of each borough will be acquired from Foursquare API.

Data Preparation

In this stage, the data about London will be acquired, processed and transformed into a tabular data. Moreover, the optimum distance among boroughs will be calculated to find the radius for querying Foursquare API.

In the next step, the data for the characteristics of each borough will be generated based on the coordinates and radius information.

Finally, the boroughs will be clustered using an unsupervised machine learning algorithm.

Evaluation

In the evaluation stage, each cluster will be thoroughly examined and the optimal location will be decided based on the data on each cluster, the total population and the surface area of each cluster.