# Final project - The Battle of Neighborhoods

## Coursera | Applied Data Science Capstone

by Martin Kovarik

## Week 4

### Assignment

1) A description of the problem and a discussion of the background. (15 marks)

2) A description of the data and how it will be used to solve the problem. (15 marks)

### PART 1 - A description of the problem and a discussion of the background

As a use case for this project, I chose the young family looking for a 2 bedroom apartment to rent in London. As London offer immense pool of apartments to rent, it is a good idea to:

[i] either limit the search with certain criteria,

[ii] or have a tool at a disposal which would help the family quickly evaluate the rental opportunity.

In the first case, the criterion to limit the search can be simply the monthly rent rate. Especially a young family, which may need to change the apartment with the increasing family size, will hesitate to pay high rent for an apartment which is just temporary for them. We can help them by categorizing the London boroughs by a median of local rents and visualize those on the map of London. This way they may limit their search more effectively to boroughs which better correspond to their rent rates expectations.

In second case, we can help them evaluate the rental opportunity by describing the given borough typical venues. Or even provide them set of clusters of boroughs which share similar characteristics. This way they can immediately get a basic idea what to expect in the given borough where the apartment is situated.

***PART 2 CONTINUES ON NEXT PAGE.***

### PART 2 - A description of the data and how it will be used to solve the problem

To achieve the goal described in the *description* section, we will use 3 sources of data.

1. First we will gather list of London's boroughs and their coordinates from Wikipedia web page. (<https://en.wikipedia.org/wiki/List_of_London_boroughs>).

This will allow us to visualize the boroughs on the map of London.

*Example of final output from this section:*

*Table

Description automatically generated*

2. Secondly, we will use Foursquare API to extract information about venues available in given London's boroughs. Applying the one hot encoding along with data transformation into easily readable dataframe we create a simple overview of the typical venues in each borough. Then using clustering method K-means, we will group the boroughs into clusters with similar features. All of those actions will help us get an additional insight about nature of each borough.

*Example of final output from this section:*

Table

Description automatically generated

3. Thirdly, we will use the use the public data about rental rates in London. (<https://www.ons.gov.uk/peoplepopulationandcommunity/housing/adhocs/12435privaterentalmarketinlondonoctober2019toseptember2020>)  
We will use this to create an easily understandable, visual overview of the rental rates in London, to allow our user group refine their decisions about where to focus their apartment search.

*Example of final output from this section:*

Table

Description automatically generated