**IBM Data Science Professional Certificate**

**IBM Capstone Project: Applied Data Science Capstone**

**Final report**

**The Battle of the Neighborhoods**

**1. Introduction/Business Problem section**

***1.1. Background and description:***

A company would like to open a Russian restaurant in Greater London.

There are a lot of immigrants from former USSR in London that like Russian Food. Many Russians like the UK and London is one the favorites cities for tourism. Also some native citizens are interested in Russian Food. It’s quite exotic for them.

It means Greater London has good conditions for opening a Russian restaurant potentially.

Greater London is organized into 33 government districts:

* 32 boroughs;
* the City of London.

It’s a huge region covers about 1600 square km and has population about 9 million people. All boroughs are very different.

Therefore, it’s very important to understand what place is the best to develop a new branch of business (to open Russian restaurant).

***1.2. Target Audience:***

Customer is a company, that want to start a Russian restaurant. We should help to make the best choice, to recommend the correct borough and explain why this place is better than others.

***1.3. Success Criteria:***

The success criteria of the project will be a good recommendation of borough for a restaurant.

**2. Data section**

We will use the next data sources:

* Foursquare API.

<https://ru.foursquare.com/>  
This service is used to collect Foursquare location data, explore each borough of London and acquire information about venues, especially restaurants, café, bars, pubs, etc. We will use this data to understand better how many restaurants in boroughs. Also it will help us to create maps.

* The London Datastore

<https://data.london.gov.uk/>

The London Datastore has been created by the Greater London Authority. It has a lot of free data for every year about boroughs, populations, squares, etc. in Greater London. The most useful for us are a list of borough names and borough population for 2020 year:

<https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough>

* TripAdvisor

<https://www.tripadvisor.com/>

Unfortunately, Foursquare does not have information about all Russian restaurants and Russian Food, therefore we need to collect and check some information manually, particularly on TripAdvisor.

**3. Methodology section**

***3.1. Business Understanding:***

Our main goal is to find optimum borough for a new restaurant.

***3.2. Analytic approach:***

We are going to use two approaches (high-level view):

1st Approach:

Our goal is to find boroughs where there are a lot of restaurants. We want to use them as “a center of attraction”. At the same time, we have to exclude places where there are Russian restaurants. We don’t want to compete with Russian Restaurant but we believe we can attract a part of clients where there are a lot of restaurants with others kind of dishes:

* Find Top 5 boroughs where are located maximum of restaurants, café, pubs, etc;
* Find boroughs with Russian restaurants;
* Delete from Top 5 boroughs with Russian restaurants.

2nd Approach:

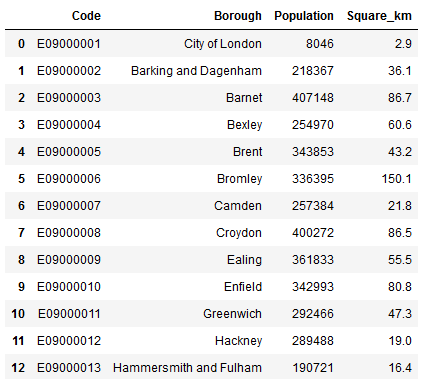
Our goal is to select clusters similar to boroughs with Russian restaurant. If in one of boroughs there is a Russian restaurant, that continues to work and has a lot of good reviews, then we should find borough with similar infrastructure.

Therefore, we will use k-means algo to cluster boroughs.

***3.3. Exploratory Data Analysis:***

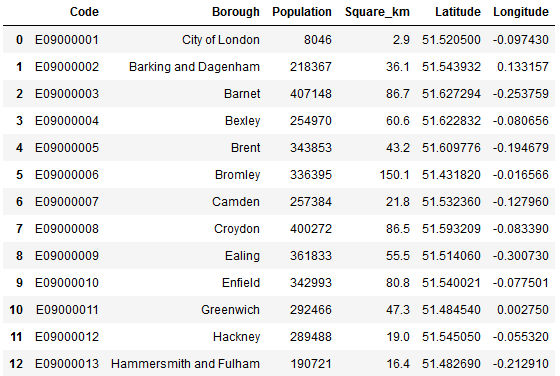
1. Download csv-file from <https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough> and Convert csv-file to DataFrame by pandas.

2. Choose data for 2020 year only and choose useful columns for us.



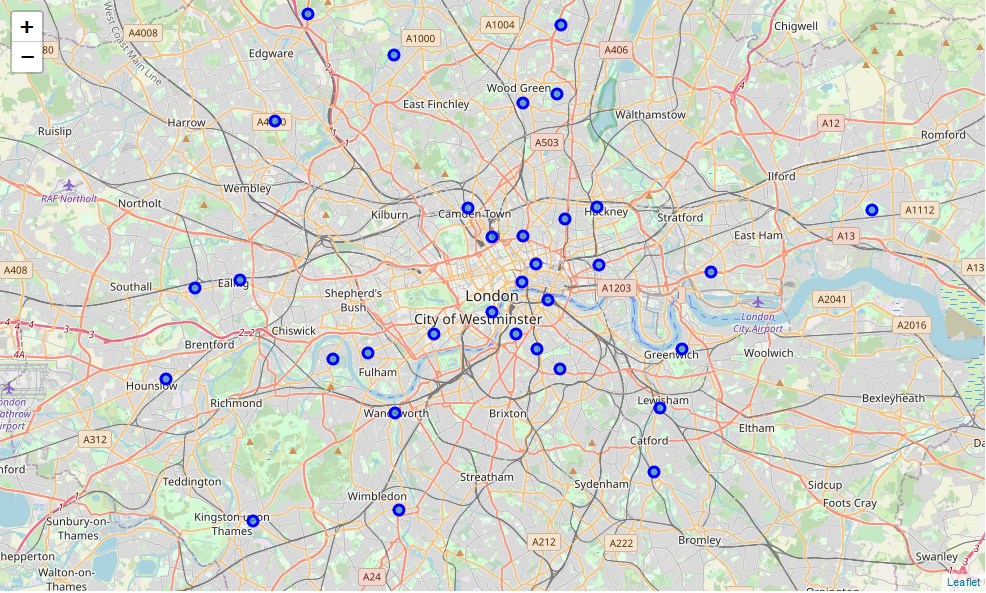
*Fig. 1. A part of clear borough list from The London Datastore.*

3. Use python geocoder library, find Latitude and Longitude, and add data to list of boroughs.



*Fig. 2. A borough table with Latitude and Longitude.*

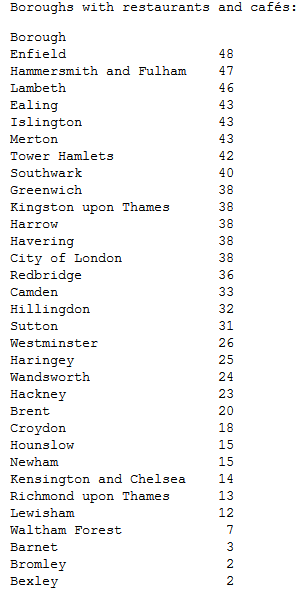
4. Create a map of all boroughs by folium.



*Fig. 3. A map of boroughs of Greater London*

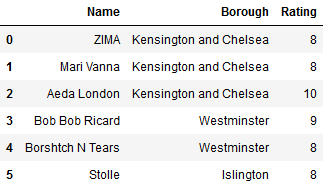
We can see from Fig. 3 some boroughs really are far from center. It’s obvious some parts of Greater London are not the best place for any restaurant.

5. Find near by venues for every borough by Foursquare API (radius 1000 m) and create a sorted list of boroughs with places similar to restaurants, cafes, etc.



As we supposed earlier boroughs at the end of the list are bad places to open restaurant.

6. Find Russian restaurants. Unfortunately, Foursquare can’t find all information about Russian restaurants therefore we collect information from TripAdvisor:



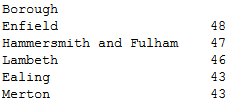
*Fig. 4. A table of Russian restaurants in Greater London*

We can see:

* Only 6 well-known Russian restaurants in Greater London;
* 3 of them are located in Kensington and Chelsea borough, 2 in Westminster and 1 in Islington.
* All restaurants have good rating (equals or more than 8) and reviews.

7. One of our goals as we described above to find boroughs where there are a lot of restaurants except Russian restaurants.

Use data from 5 and 6 steps and create a short list of good boroughs for us:

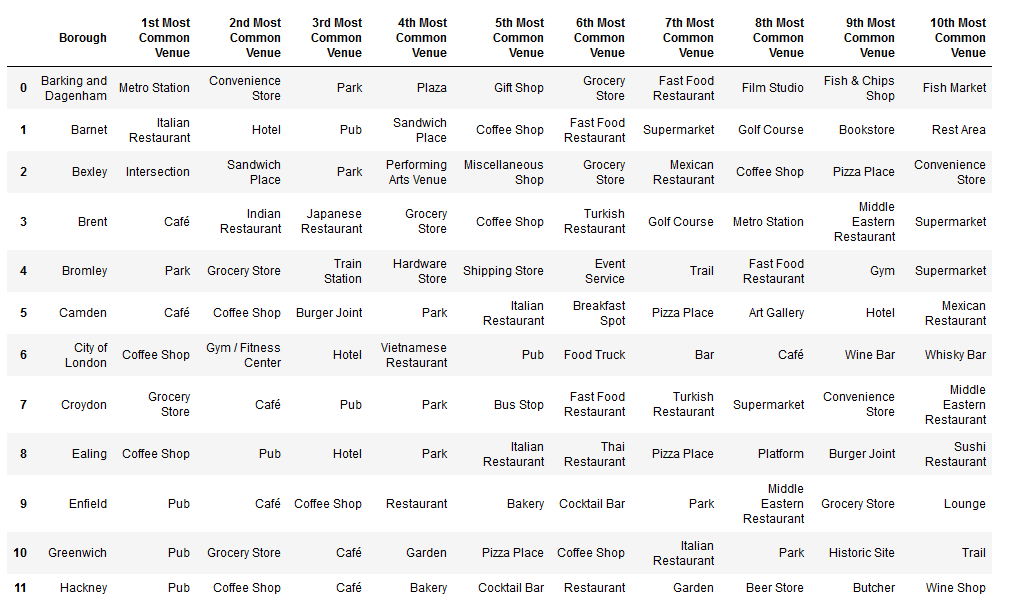


8. Find each borough from short list along with the top 5 most common venues:

|  |  |
| --- | --- |
| ----Enfield----  venue freq  0 Café 0.10  1 Pub 0.10  2 Coffee Shop 0.06  3 Restaurant 0.05  4 Cocktail Bar 0.03 | ----Hammersmith and Fulham----  venue freq  0 Café 0.11  1 Italian Restaurant 0.09  2 Coffee Shop 0.07  3 Pub 0.07  4 Grocery Store 0.06 |
| ----Lambeth----  venue freq  0 Pub 0.16  1 Café 0.08  2 Hotel 0.07  3 Coffee Shop 0.06  4 Park 0.05 | ----Ealing----  venue freq  0 Coffee Shop 0.11  1 Pub 0.11  2 Hotel 0.05  3 Park 0.04  4 Italian Restaurant 0.04 |
| ----Merton----  venue freq  0 Grocery Store 0.06  1 Coffee Shop 0.06  2 Italian Restaurant 0.05  3 Pub 0.05  4 Thai Restaurant 0.04 |  |

We can see again these boroughs have good infrastructure if we want to use them as “a center of attraction”.

9. Create a dataframe and display the top 10 venues for each borough:

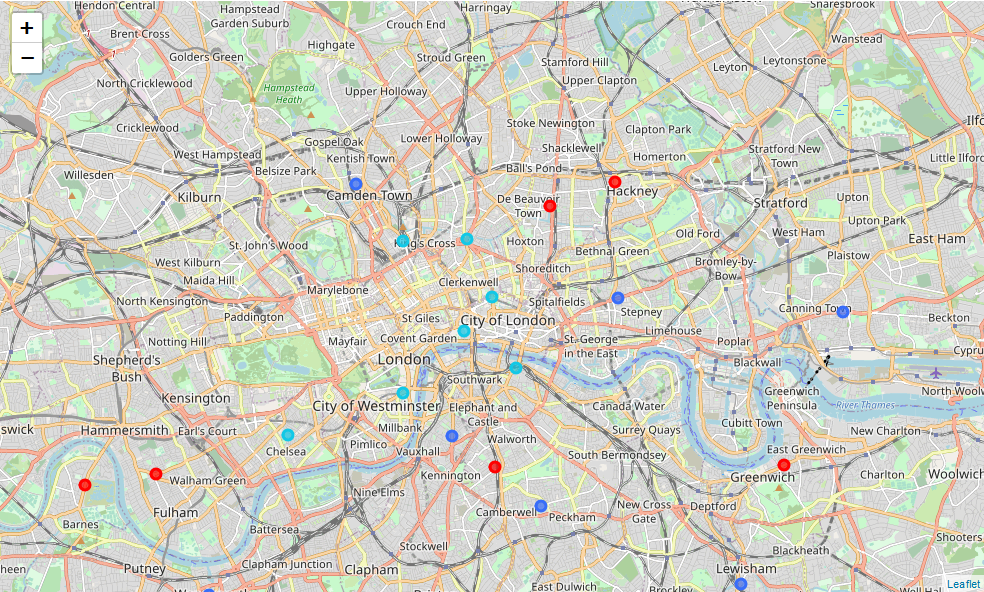


*Fig. 5. A fragment of table the top 10 venues for each borough.*

10. Run k-means clustering, find 4 similar clusters of boroughs and create a map.

Light blue circles are the best places for Russian restaurant. However, we have to remember we want to exclude borough from Fig. 4. (Kensington and Chelsea borough, Westminster and Islington) because we don’t want compete with existent Russian restaurant.

If we compare a list from step 5 and Fig. 6 we can see Lambeth and Tower Hamlets are the best boroughs for a Russian restaurant.



*Fig. 6. A map of similar borough clusters in Greater London*

**4. Results section**

Main research results:

* The best boroughs to open a Russian restaurant are Lambeth and Tower Hamlets;
* Greater London don’t have a lot of Russian restaurant;
* Existent Russian restaurants have relatively compact location: Kensington and Chelsea, Westminster and Islington and they have high rating;
* Lambeth and Tower Hamlets boroughs are similar to boroughs where Existent Russian restaurants are located.

**5. Discussion section**

Main results are described at the previous section. We can recommend Lambeth and Tower Hamlets as the best boroughs to open a Russian restaurant.

These boroughs have a lot restaurants, cafes and entertainments what can be used as ‘a center of attraction’. At the same time these boroughs don’t have direct competitors (Russian restaurant). It’s important for business development.

We should avoid such boroughs as Kensington and Chelsea, Westminster and Islington to run a Russian restaurant because they have Russian restaurants and these restaurants have high rating and good reviews.

**6. Conclusion section**

Our goal was to find the best borough in Greater London to run a Russian restaurant.

In order to find the best borough, we use such criteria’s as

* “a center of attraction”;
* minimum Russian restaurants criteria;
* k-means clustering to find similar boroughs to boroughs where Russian restaurants are located now.

As a result, we can recommend are Lambeth and Tower Hamlets boroughs.

In the future, it’s a good idea to collect more information about revenue, population, rent payments, crime level, future projects, etc. It’ll help to choose some good places inside boroughs.