

# Capstone Project—The Battle of Neighborhoods in Birmingham: Restaurants

## 1. Introduction

### a. Discussion of the background

It's essential to keep the most of the costs to a minimum when you starting a small business in a promising location. Starting business in a London is not surprise but prices of the London too high, we should consider another place and that is Birmingham. With nearly 2.6 million people in urban area and 1.1 million inhabitants in the city area, Birmingham is the second largest city in England. Because of Birmingham's unique geography, (located middle of the country, major gateway between North and the South) in 2018 19 thousand business were create there and it's the most popular destination for those moving out of the capital last year.

The primary reason for choosing Birmingham is nearly 40% of its population are under 25's and its diverse population instead solely with one ethnicity. There are several research about young people spend more than any other age group on meals—much of it on fast, thanks to a lack of cooking skills. Therefore it's an ideal place to start a restaurant business in Birmingham.

### b. Problem Description

Despite the fact its great place to start a restaurant in Birmingham, there are already many successful cuisines active so in order to survive in such a competitive market it is very important to strategically plan and consider vital factors.

## 2. Data requirements

For this project we need following data:

Birmingham data that contains neighborhoods and its longitude and latitude.

Data source: [https://en.wikipedia.org/wiki/List\\_of\\_neighbourhoods\\_of\\_Birmingham](https://en.wikipedia.org/wiki/List_of_neighbourhoods_of_Birmingham)

Description: we will Scrap Birmingham districts table from Wikipedia and get coordinates using geocoder class of geopy client.

To find Restaurants in Birmingham we will use Foursquare APIs

Data source: Foursquare APIs

Description: By using this API we will get all the venues in each neighborhood. We can filter venues to get only restaurants

## 3. Methodology

Business understanding:

My main part of this business is to get the optimal location for a restaurant in Birmingham.

#### a. Data preparation and exploratory data analysis:

First we use BeautifulSoup library to scrap from Wiki which contains list of Birmingham neighborhoods. Then I use panda to transform the data to create table.

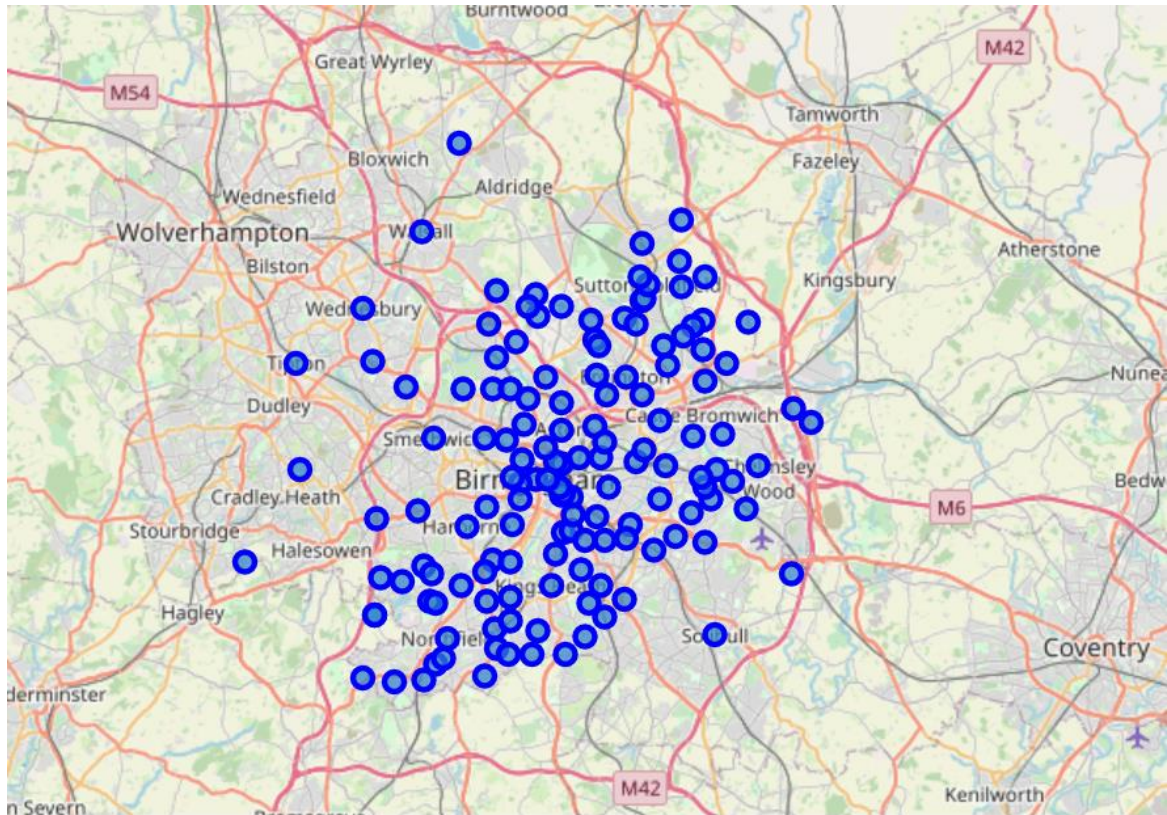
#### Getting coordinates of Neighborhood

Using geopy.geocoders we will extract Neighborhood's latitude and longitude. Some of the neighborhood's coordinates not found so we do some cleansing and we have our dataframe table.

Out[13]:

	Neighborhood	Latitude	Longitude
0	Acocks Green	52.4495	-1.81924
1	Alum Rock	52.4871	-1.83153
2	Ashted	52.4861	-1.88381
3	Aston	52.5007	-1.88419
4	Aston Cross	52.001	-2.08718
5	Austin Village	52.4009	-1.97269
6	Balsall Heath	52.4572	-1.88321
7	Balti Triangle	52.4576	-1.87664
8	Bartley Green	52.4359	-1.99489
9	Beech Lanes	52.4659	-1.98457
10	Bickenhill	52.439	-1.72328

Then we're able to create map using folium since we have got coordinates.



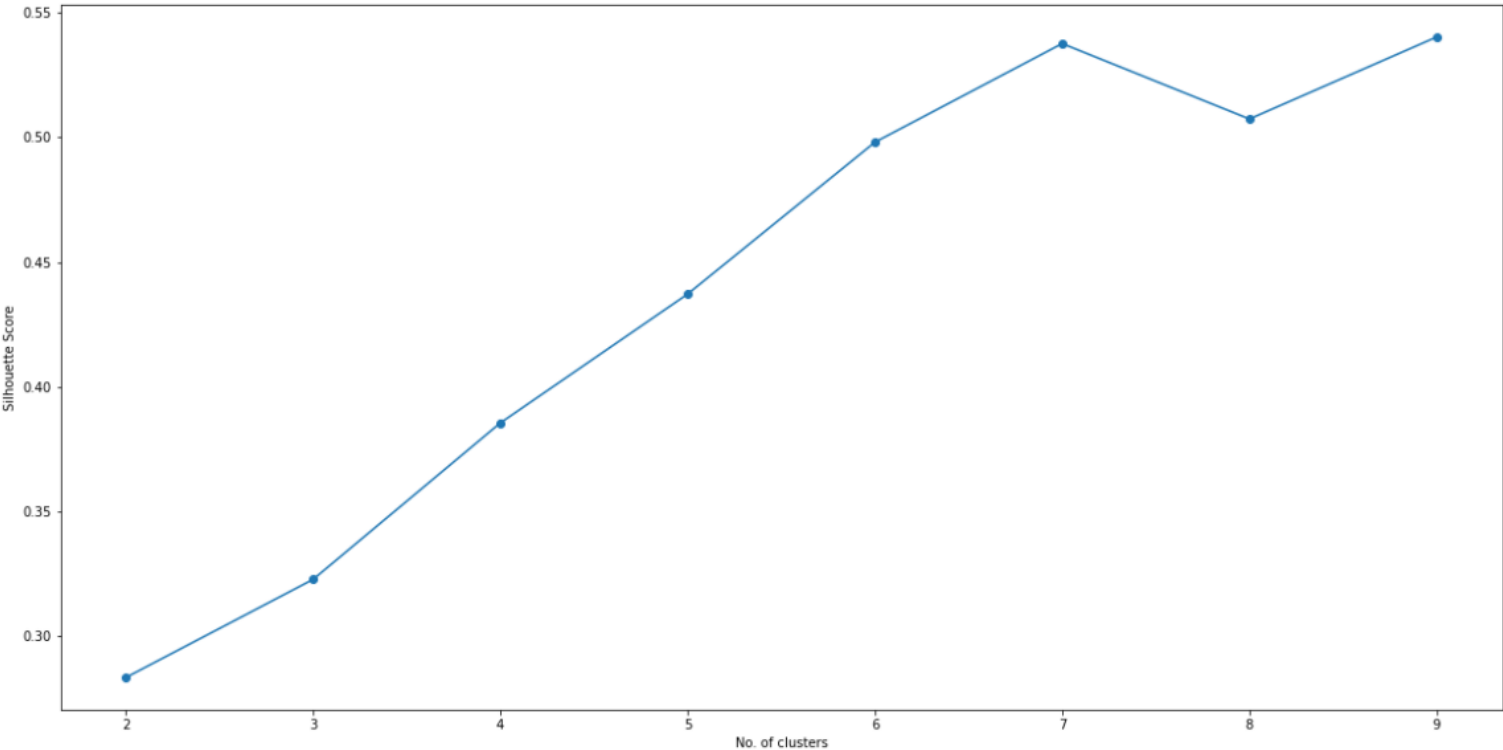
All the blue markers on the map above are Neighborhoods. Most of the blue dots are centered and few of the neighborhood located far away from center.

## **b. Problem approach and using k-means clustering**

### **Using foursquare location data**

After getting neighborhood's coordinate values, we merge data values with original dataframe appropriately. Now Foursquare API is used to explore the neighborhood and segment them. First, the data frame with all the venue information by foursquare for the given latitude and longitude values. FourSquare API returns 1672 results and 251 unique categories. Later on, I will extract and concentrate in restaurant category only and explore all the neighborhoods. Next we will find frequency of occurrence of each restaurant using one hot encoding and a new data frame is created with its corresponding top 10 common restaurants based on cuisine.

In the end we check the pattern for each neighborhood and get the information about the top ten common restaurants for each neighborhood using k-means clustering. To do this we have first find best value for k so we will use elbow method. Elbow method uses silhouette\_score from sklearn,metrics. From the graph below we choose k=9 .



Finally, k-means clustering is performed on the following data frame with k=X, to determine the pattern of the top 10 restaurants based on cuisine in every neighborhood.

Out[43]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
7	Alum Rock	52.487071	-1.831530	Desi Karahi	52.488817	-1.830819	Asian Restaurant	6	Asian Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Italian Restaurant
38	Balsall Heath	52.457225	-1.883207	Diwan Balti	52.453767	-1.887040	Indian Restaurant	0	Indian Restaurant	Fast Food Restaurant	Middle Eastern Restaurant	Moroccan Restaurant
39	Balsall Heath	52.457225	-1.883207	Lahore	52.456860	-1.876458	Indian Restaurant	0	Indian Restaurant	Fast Food Restaurant	Middle Eastern Restaurant	Moroccan Restaurant
40	Balsall Heath	52.457225	-1.883207	Bader Restaurant	52.457440	-1.876448	Moroccan Restaurant	0	Indian Restaurant	Fast Food Restaurant	Middle Eastern Restaurant	Moroccan Restaurant
43	Balsall Heath	52.457225	-1.883207	Al-Frash Balti	52.457261	-1.876494	Indian Restaurant	0	Indian Restaurant	Fast Food Restaurant	Middle Eastern Restaurant	Moroccan Restaurant

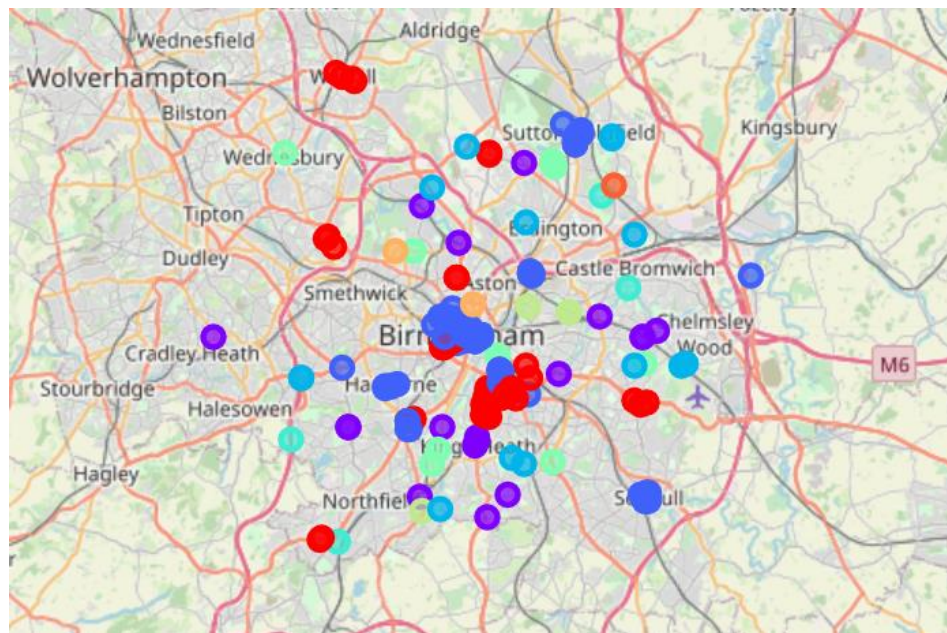
## 4. Results

We used kmeans cluster on mean occurrence of venue category

All X clusters follow unique patterns for the top 10 common restaurants for a particular neighborhood. Below graph shows the number of neighborhoods assigned to each cluster.

2	177
0	67
1	19
5	16
3	15
6	4
4	4
8	2
7	2

Cluster map shown below



## 5. Discussion

We can discard cuisine and choose preferred location with the not common restaurant or we can open a restaurant and discard location with the respective cuisine.

## **6. Conclusion**

This project done by limited data. If we have other factors such as venue cost, neighborhood population, restaurant profit (etc) we can choose location more precisely.