Introduction

Birmingham is the second biggest city of the UK. It is a vibrant city with business opportunities and we want to pursue them.

Business Problem

Let us assume we are interested in retail, in particular, in opening an Italian restaurant in the area of Birmingham, UK. Therefore, we want to understand the area and which neighbourhoods would be the most attractive and interesting for this purpose.

Data

The data used will be:

List of Birmingham areas by postcode (https://en.wikipedia.org/wiki/B_postcode_area), that we will use to gather info on the neighbourhoods

- Geolocations(pgeocoder,geocoder), that we will use to collect the longitude and latitude of the neighbourhoods

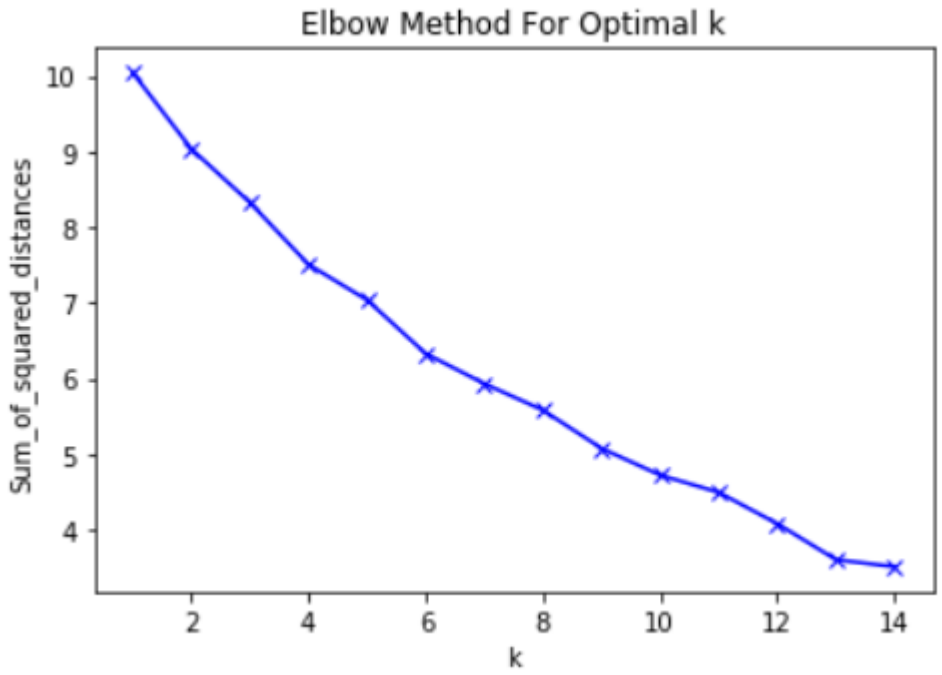
- Venues info obtained through Foursquare API, that we will use to understand the neighbourhoods and turn them into clusters, according to the types of present venues, in order to see which one would the best match for our goal

Methodology

First, through a public Wikipedia page the postal codes of the city were webscraped, using BeautifulSoup.

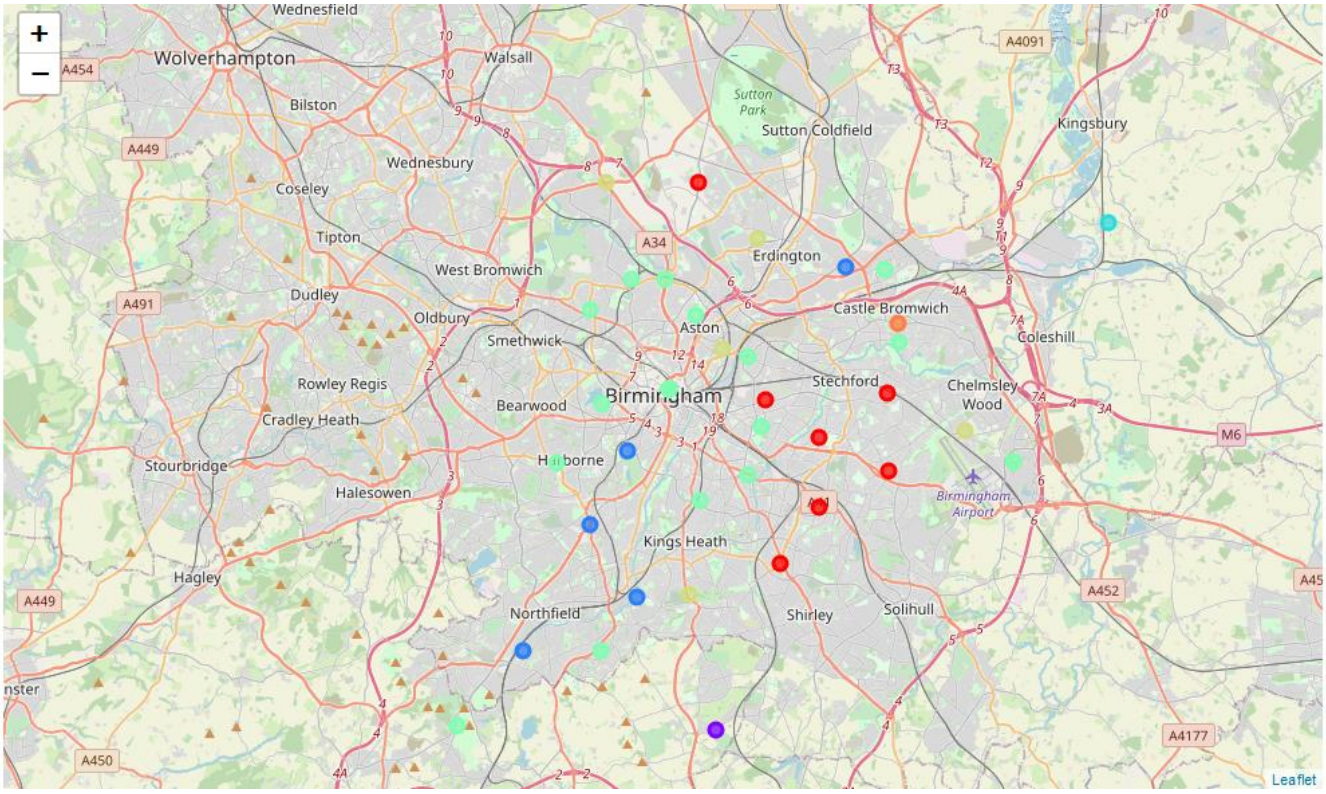
For those postal codes, thanks to pgeocode, latitude and longitude were collected.

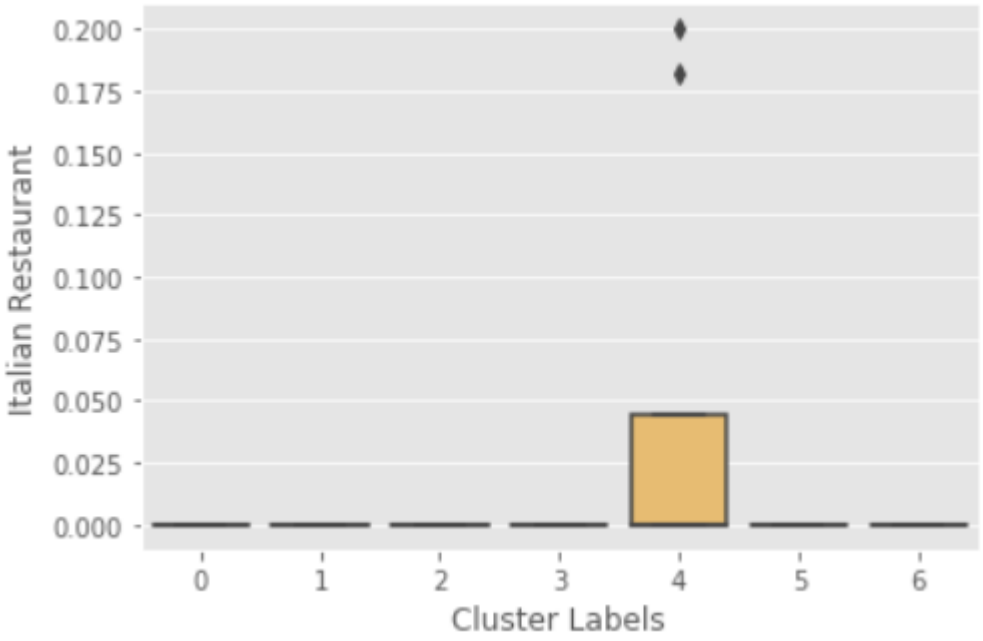
Through Foursquare API, for each neighbourhood, info about the venues were gathered. The info about the venues, were turned into quantitative variables (one hot encoding), in order to make the info usable for the KMeans algorithm. Seven clusters were formed (using the "elbow method" as an indicator).



Results

The result of the operation can be better seen plotted on a map of the area

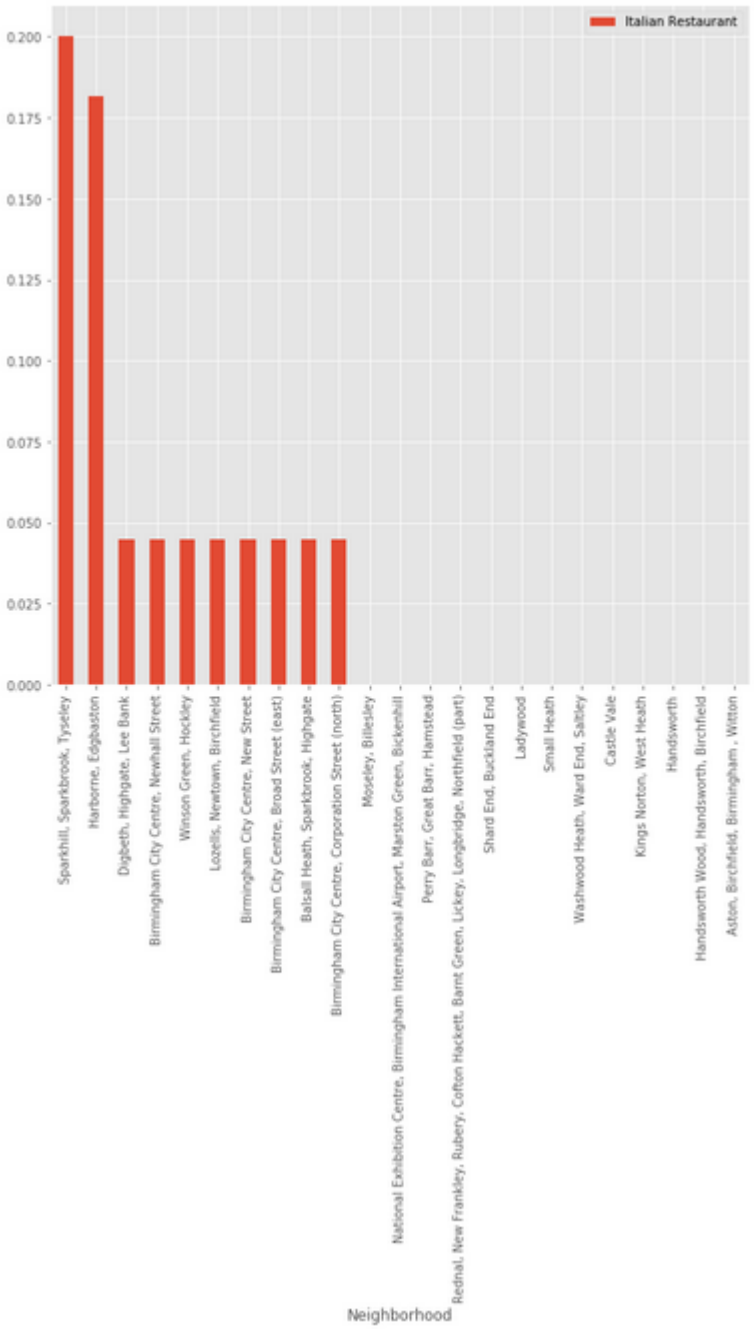




Italian restaurants were find only in cluster n.4 (the one in light green on the map).

As the map shows at glance, the neighbourhoods of this cluster are mostly located near the city center.

Looking inside cluster 4, we can see that the distribution is quite unbalanced.



Discussion

From what discussed so far, there is a presence of Italian restaurants only in one cluster, which most likely, suggests that the only certain areas have the charateristics to make the business work. If this is correct, we see, at the same time, that within the cluster, the distribution is extremely uneven.

From this, two recommendations could spring:

- Placing the business in the city-center would seem like a “safe choice”
- Placing the business in the neighbourhoods of the cluster 4, where there are no Italian restaurants would be a riskier one. I say riskier because, even if the neighbourhoods belong to the same cluster, the fact there are no Italian restaurants could have a specific reason.

Conclusion

In this report we tried to answer the question regarding where to open an Italian restaurant in Birmingham, UK. To do that we use data about neighbourhoods’ venues in order to cluster them and find the best fitting areas. Even though using only the mentioned data probably simplifies the issue too much, the research recommends the areas within custer n.4.