

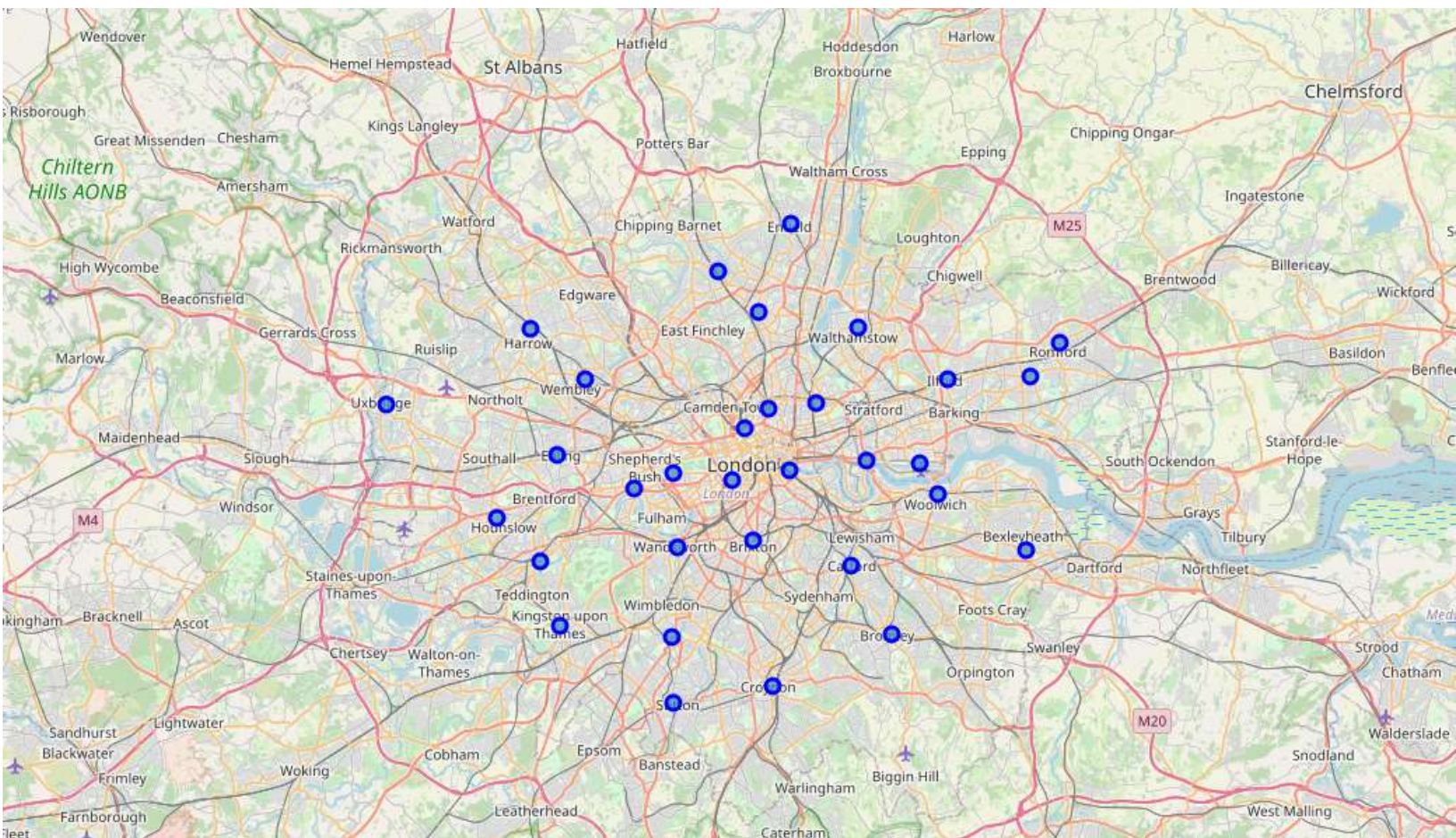
# Capstone Project – Battle of Neighborhoods

## Nighborhood recommendations for Lithuanian grocery store in London, United Kingdom

### 1. Introduction

London is a highly diverse city with many nationalities living close by. Because of high number of different nationalities London has fantastic food scene. There are many international restaurants and stores around the city. For this specific project, it would be analyzed which London neighbourhood has the best potential for the Lithuanian grocery store.

London has 32 boroughs, each having neighborhoods. Here is the map of London, with it's boroughs marked in blue.



To start this kind of business in London one should be aware of the borough with highest Lithuanian populations, as well as choose the right neighbourhood. In this report it would be discussed where and why it is recommended to open a grocery store. We would rely on the following criteria:

- High Lithuanian population
- Popularity of the convenience stores

## **2. Data and it's collection**

- Wikipedia page was used to get the data about London boroughs.
- Foursquare API was used as a source to gain information about nearby places of each neighborhood within a radius of 100 meters. The data extracted from Foursquare contains information of venues within a specific distance of the lat and long. There are several details extracted from Foursquare:
  - Neighborhood Longitude
  - Neighborhood Latitude
  - Neighborhood
  - Name of the venue (restaurant, shop, hotel, etc.)
  - Venue Longitude
  - Venue Latitude
  - Venue type (category)

## **3. Methodology**

To start with, the borough with the highest Lithuanian population was identified. Using some external materials, it was confirmed that highest Lithuanian population lives in Newham borough. So this borough was analyzed in more details and it's neighborhoods were clustered. Steps included data acquisition, scraping, cleaning, data preparation, clustering

### **3.1 Data collection and cleaning**

Data acquisition and cleaning includes following steps:

1. Identify the boroughs of London, scrape the page to get the information and coordinates of the boroughs.
2. Select the borough, which has the highest Lithuanian population and get it's coordinates.
3. Using Foursquare API analyze the venues in the selected borough.

### **3.2 Clustering**

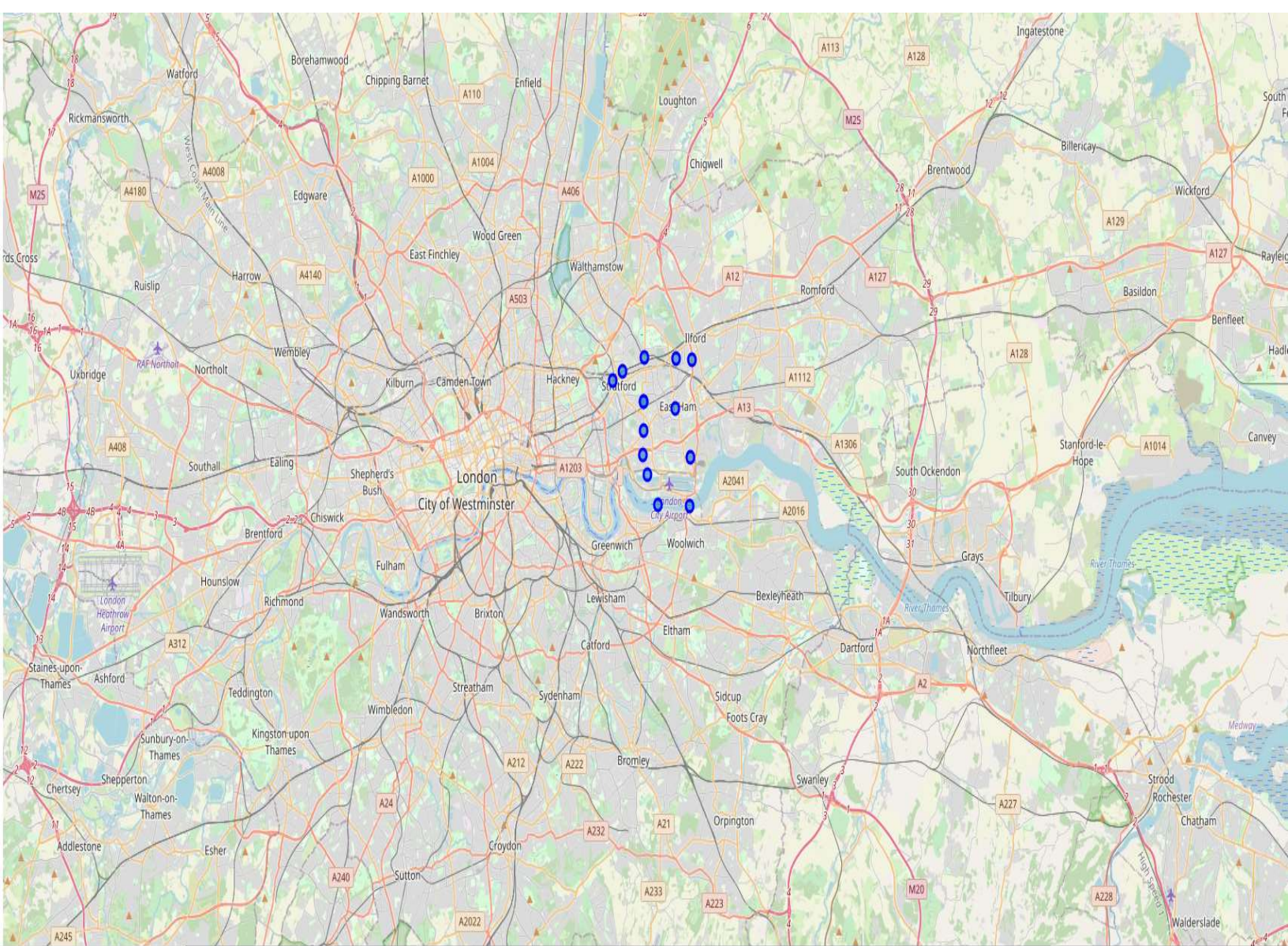
In order to understand which neighborhood is the best to open the Lithuanian shop, the neighborhoods of Newham was explored, segmented and grouped into the clusters. An unsupervised machine learning approach – k-means clustering algorithm was used.

## **4. Results**

### **4.1 Neighborhoods selected for clustering**

There are 13 neighborhoods in Newham borough. Since information is available, all of the neighborhoods would be clustered and analyzed. Some of these neighborhoods would potentially be recommended for the future business development. Below there is a map with all the neighborhoods marked.





## 4.2 Clustering of neighborhood

The venues information, which was obtained from the Foursquare API contains the information about all the neighborhoods in Newham borough. They were clustered, segmented and grouped by the categories. Five clusters were created. Cluster 1 consists of 8 neighborhoods, Cluster 2 consists of 2 and other 3 clusters has just 1 neighborhood featured. In most cases the same venues are among the top visited, such as: hotel, pub, grocery store and gym.



	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Beckton	Hotel	Furniture / Home Store	Gym / Fitness Center	Discount Store	Clothing Store	Pub	Shopping Plaza	Light Rail Station	Grocery Store	Frozen Yogurt Shop
2	Custom House	Hotel	Chinese Restaurant	Coffee Shop	Gym	Pub	Restaurant	Bakery	Bridge	Café	Convenience Store
3	East Ham	Pub	Clothing Store	Bookstore	Electronics Store	Sandwich Place	Coffee Shop	Supermarket	Fast Food Restaurant	Grocery Store	Park
4	Forest Gate	Grocery Store	Fast Food Restaurant	Train Station	Bakery	Pub	Café	Fish & Chips Shop	Wine Bar	Dessert Shop	Discount Store
7	Maryland	Hotel	Pub	Café	Indian Restaurant	Sculpture Garden	Moving Target	Liquor Store	General Entertainment	Chinese Restaurant	Eastern European Restaurant
8	North Woolwich	Pier	Hotel	Clothing Store	Sculpture Garden	Pharmacy	Plaza	Italian Restaurant	Pub	River	Fountain
9	Plaistow	Grocery Store	Bus Stop	Café	African Restaurant	Park	Indian Restaurant	Gym / Fitness Center	Pub	Business Service	Doner Restaurant
11	Stratford	Pub	Clothing Store	Café	Ice Cream Shop	Pizza Place	Juice Bar	Department Store	Cosmetics Shop	Sandwich Place	Coffee Shop

Newham\_merged.loc[Newham\_merged['Cluster Labels'] == 1, Newham\_merged.columns[[1] + list(range(5, Newham\_merged.shape[1]))]]

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
12	Upton Park	Convenience Store	Bus Stop	Bookstore	Boutique	Wine Bar	Doner Restaurant	Donut Shop	Eastern European Restaurant	Electronics Store	English Restaurant
13	West Ham	Convenience Store	Bus Stop	Bookstore	Boutique	Wine Bar	Doner Restaurant	Donut Shop	Eastern European Restaurant	Electronics Store	English Restaurant

Newham\_merged.loc[Newham\_merged['Cluster Labels'] == 2, Newham\_merged.columns[[1] + list(range(5, Newham\_merged.shape[1]))]]

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Little Ilford	Grocery Store	Ice Cream Shop	Indian Restaurant	Fried Chicken Joint	Wine Bar	Fish & Chips Shop	Dessert Shop	Discount Store	Doner Restaurant	Donut Shop

Newham\_merged.loc[Newham\_merged['Cluster Labels'] == 3, Newham\_merged.columns[[1] + list(range(5, Newham\_merged.shape[1]))]]

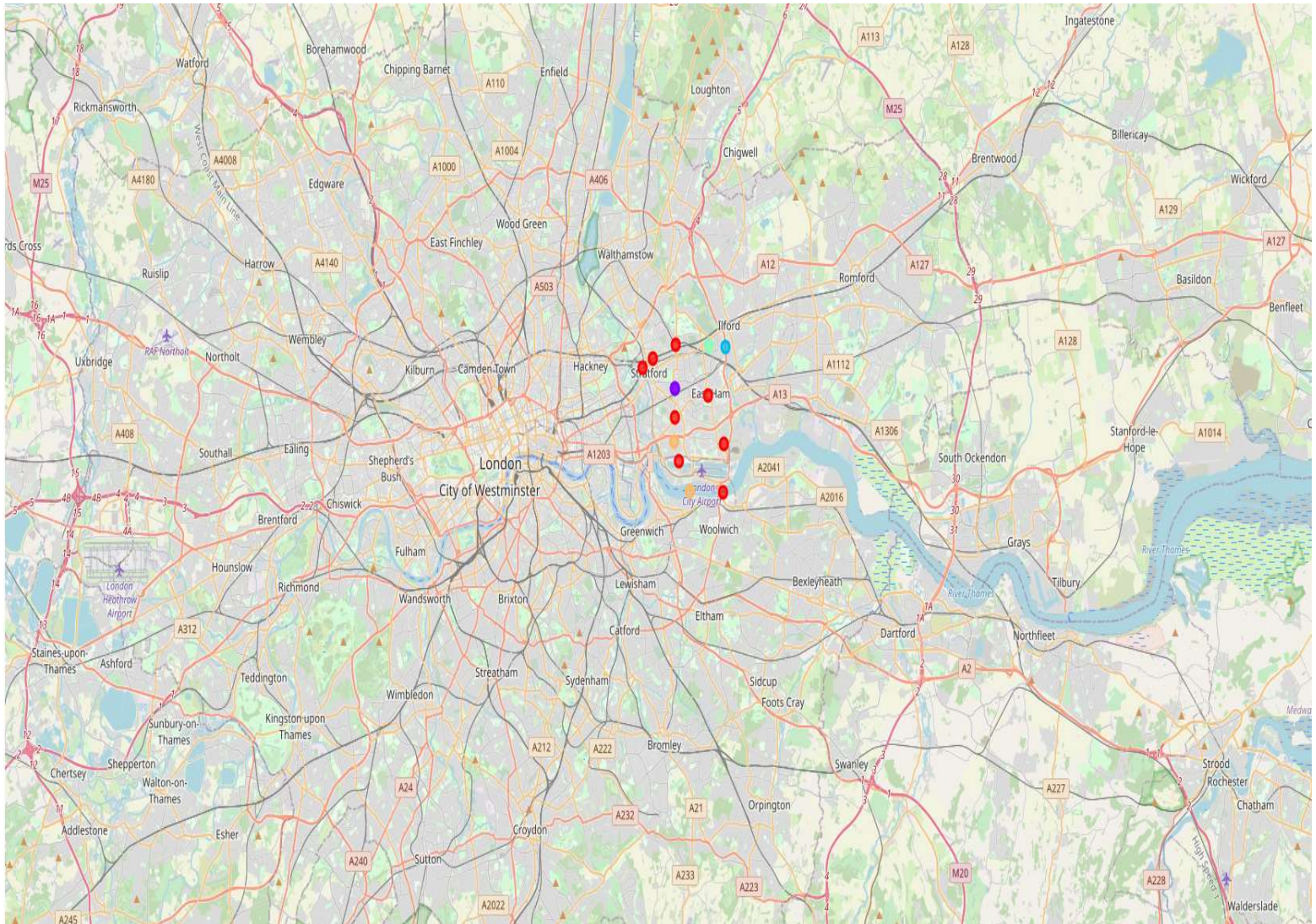
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Manor Park	Gym / Fitness Center	Turkish Restaurant	Asian Restaurant	Gas Station	Restaurant	Wine Bar	Fast Food Restaurant	Dessert Shop	Discount Store	Doner Restaurant

Newham\_merged.loc[Newham\_merged['Cluster Labels'] == 4, Newham\_merged.columns[[1] + list(range(5, Newham\_merged.shape[1]))]]

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Canning Town	Park	Convenience Store	Gas Station	Café	Fish & Chips Shop	Dessert Shop	Discount Store	Doner Restaurant	Donut Shop	Eastern European Restaurant
10	Silvertown	Gym / Fitness Center	Theater	Museum	Construction & Landscaping	Café	Paintball Field	Park	Go Kart Track	Gas Station	Furniture / Home Store

Locations of these clustered neighborhoods are also available on the map:





## 5. Discussion

In this analysis the future business owner could identify, which part of the borough is the most attractive for the Lithuanian store. Looking at the results Forest Gate, Plaistow, Upton Park, West Ham and Little Ilford has grocery store as a top visited location. According to these results, these locations would be ideal for the new grocery store. Considering the fact, that it is highly populated with Lithuanians, the specialty shop might be a great success.

## 6. Conclusion

In this project after the analysis of the neighborhoods, using the Foursquare API and other data, as well as k-means algorithm the best neighborhood for opening Lithuanian grocery store were identified. This analysis could be useful for anyone who is about to open a Lithuanian store in London.

