

THE BATTLE OF NEIGHBORHOODS

To find the best place to live in London, considering the crime occurrence and venues density

CAPSTONE PROJECT

BARBORA HUSÁKOVÁ

20.04.2020



Table of Contents

- Introduction: Business Problem
- Data Acquisition and Cleaning
- Methodology
- Results and Discussion
- Conclusion

Introduction: Business Problem

Business Problem

- Each **London** borough has its own characteristics, depending on what one's preferences are, one needs to decide where to settle down when relocating to London. Some districts offer higher safety, others more local amenities and nightlife activities while somebody may seek for quite neighborhood with park areas and local markets
- As there are many factors to be taken into account when looking for a borough where one is going to spend several months or even years of life, we have decided to **analyze London's neighborhoods** from **safety and venues perspective** in order to help potential future Londoners to find a proper and suitable borough. Moreover, these analytics can aid real estate agencies as well, when categorizing the boroughs and giving recommendations to the people searching for a real estate.

Main Goal

- The main objective of this project is to **find the best place to live in London, taking into account the crime occurrence and venues density** meaning to classify boroughs in London based on the safety indicator (reflecting the current status of the crime incidents) and topmost common venues in each of them.

Data Acquisition and Cleaning

In order to proceed with the analytics part and to achieve the set objectives, the below data has been collected and processed in order to give us better insight of the safety and venues availability:

- **Total number of crimes** in the last two years per borough and crime category
- **The most common venues** in each borough that are reported at the time of analysis

Data Acquisition and Cleaning

London Crime Dataset

- London crime records per borough and crime category in the last two years
- Source: London Datastore
- https://data.london.gov.uk/dataset/recorded_crime_summary

| | MajorText | MinorText | BoroughName | 201803 | 201804 | 201805 | 201806 | 201807 | 201808 | 201809 | ... | 201905 | 201906 | 201907 | 201908 | 201909 | 201910 | 201911 |
|---|---------------------------|-----------------------------------|----------------------|--------|--------|--------|--------|--------|--------|--------|-----|--------|--------|--------|--------|--------|--------|--------|
| 0 | Arson and Criminal Damage | Arson | Barking and Dagenham | 6 | 3 | 4 | 12 | 6 | 5 | 3 | ... | 11 | 3 | 5 | 3 | 6 | 9 | 8 |
| 1 | Arson and Criminal Damage | Criminal Damage | Barking and Dagenham | 115 | 122 | 126 | 123 | 127 | 101 | 107 | ... | 138 | 113 | 134 | 118 | 109 | 109 | 99 |
| 2 | Burglary | Burglary - Business and Community | Barking and Dagenham | 38 | 36 | 24 | 33 | 30 | 18 | 33 | ... | 22 | 27 | 31 | 35 | 37 | 30 | 30 |
| 3 | Burglary | Burglary - Residential | Barking and Dagenham | 122 | 75 | 93 | 77 | 94 | 84 | 99 | ... | 114 | 96 | 71 | 67 | 80 | 97 | 114 |
| 4 | Drug Offences | Drug Trafficking | Barking and Dagenham | 7 | 3 | 8 | 6 | 9 | 7 | 10 | ... | 8 | 6 | 8 | 6 | 6 | 9 | 10 |

Data Acquisition and Cleaning

London Boroughs Dataset

- List of London boroughs and the population in the respective borough that is to be used to calculate the ratio of the crimes per 1000 people
- Source: Wikipedia
- https://en.wikipedia.org/wiki/List_of_London_boroughs

| | Borough | Inner | Status | Local authority | Political control | Headquarters | Area (sq mi) | Population (2013 est)[1] | Co-ordinates | Nr. in map |
|---|-------------------------------|-------|--------|---|-------------------|---|--------------|--------------------------|---|------------|
| 0 | Barking and Dagenham [note 1] | NaN | NaN | Barking and Dagenham London Borough Council | Labour | Town Hall, 1 Town Square | 13.93 | 194352 | 51°33'39"N 0°09'21"E / 51.5607°N 0.1557°E | 25 |
| 1 | Barnet | NaN | NaN | Barnet London Borough Council | Conservative | North London Business Park, Oakleigh Road South | 33.49 | 369088 | 51°37'31"N 0°09'06"W / 51.6252°N 0.1517°W | 31 |
| 2 | Bexley | NaN | NaN | Bexley London Borough Council | Conservative | Civic Offices, 2 Watling Street | 23.38 | 236687 | 51°27'18"N 0°09'02"E / 51.4549°N 0.1505°E | 23 |
| 3 | Brent | NaN | NaN | Brent London Borough Council | Labour | Brent Civic Centre, Engineers Way | 16.70 | 317264 | 51°33'32"N 0°16'54"W / 51.5588°N 0.2817°W | 12 |
| 4 | Bromley | NaN | NaN | Bromley London Borough Council | Conservative | Civic Centre, Stockwell Close | 57.97 | 317899 | 51°24'14"N 0°01'11"E / 51.4039°N 0.0198°E | 20 |

Data Acquisition and Cleaning

London Boroughs Coordinates

- Get the coordinates of the respective boroughs
- Source: Geopy API
- <https://pypi.org/project/geopy/>

| | Borough | Population | Latitude | Longitude |
|---|----------------------|------------|-----------|-----------|
| 0 | Barking and Dagenham | 194352 | 51.554117 | 0.150504 |
| 1 | Barnet | 369088 | 51.653090 | -0.200226 |
| 2 | Bexley | 236687 | 51.441679 | 0.150488 |
| 3 | Brent | 317264 | 51.563826 | -0.275760 |
| 4 | Bromley | 317899 | 51.402805 | 0.014814 |

Data Acquisition and Cleaning

London Venues in the Respective Boroughs

- List of top 50 venues in the neighborhood
- Source: Foursquare API
- <https://api.foursquare.com>

| | BoroughName | Borough Latitude | Borough Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|----------------------|------------------|-------------------|--------------|----------------|-----------------|------------------------|
| 0 | Barking and Dagenham | 51.554117 | 0.150504 | Central Park | 51.559560 | 0.161981 | Park |
| 1 | Barking and Dagenham | 51.554117 | 0.150504 | Lara Grill | 51.562445 | 0.147178 | Turkish Restaurant |
| 2 | Barking and Dagenham | 51.554117 | 0.150504 | Iceland | 51.560578 | 0.147685 | Grocery Store |
| 3 | Barking and Dagenham | 51.554117 | 0.150504 | Wilko | 51.541002 | 0.148898 | Furniture / Home Store |
| 4 | Barking and Dagenham | 51.554117 | 0.150504 | Shell | 51.560415 | 0.148364 | Gas Station |

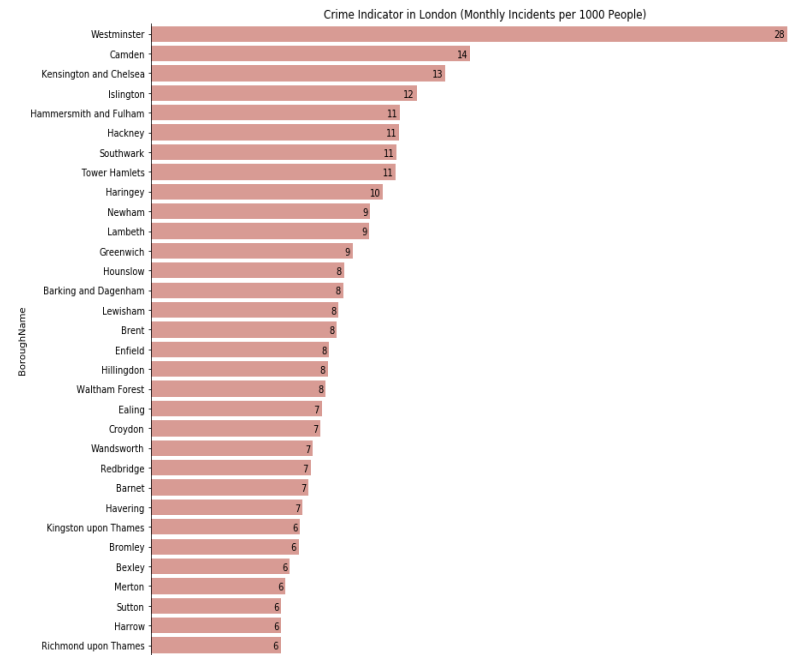
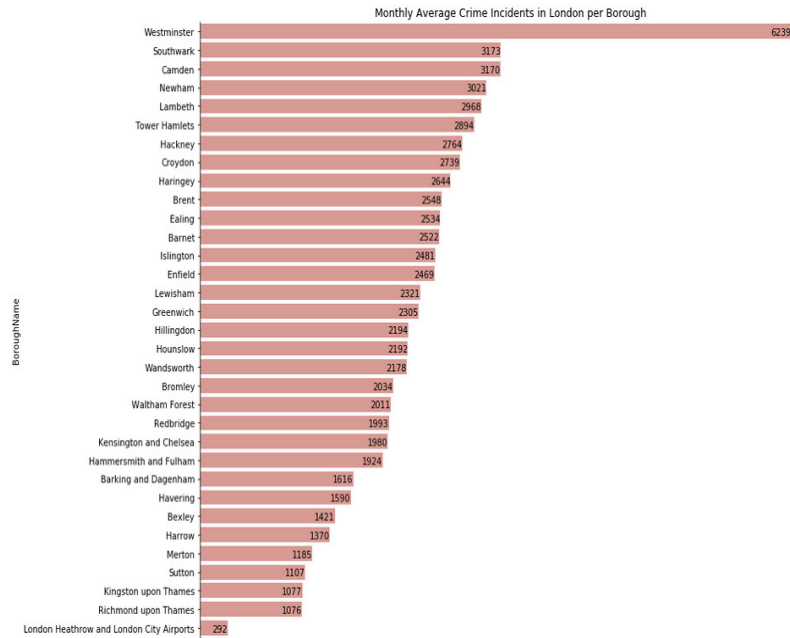
Methodology

3 main parts:

- **Exploratory data analysis** will serve us as the source of better understanding the distribution of the crimes across the boroughs as well as map's visualizations will help us fully grasp the geographic layout of the boroughs
- **Modelling** part will focus on the unsupervised learning algorithm, namely k-means through which we will cluster the similar boroughs into 5 clusters in order to be able to analyze these from venues perspective/density/type
- **Cluster analysis** will concentrate on the overview of each group of boroughs with the venues in order to better understand which venues are most typical for the respective clusters and perform the classification of these into more informative categories

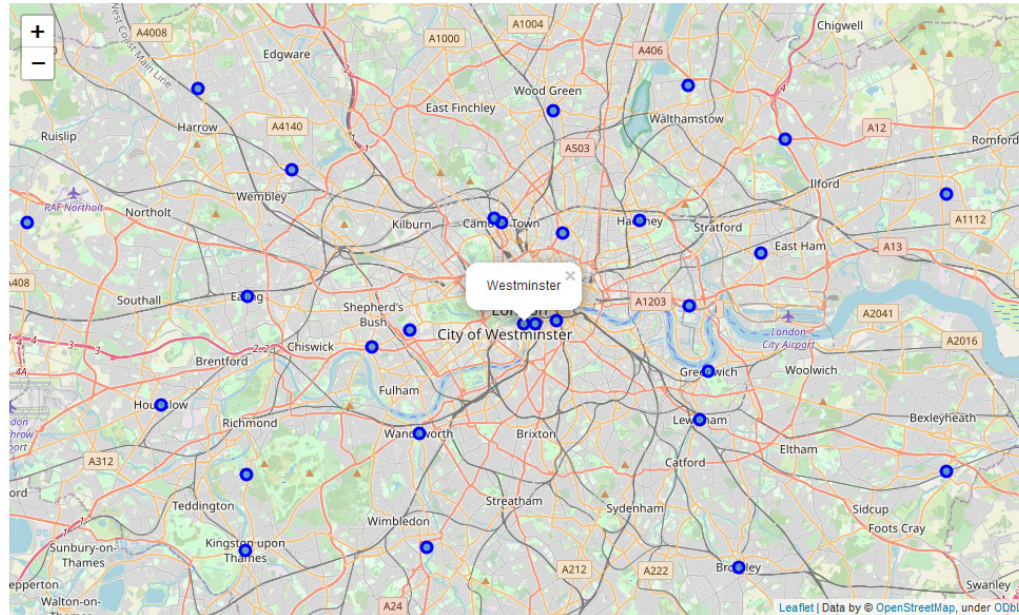
Methodology

Exploratory Data Analysis



Methodology

Exploratory Data Analysis



Methodology

Modelling

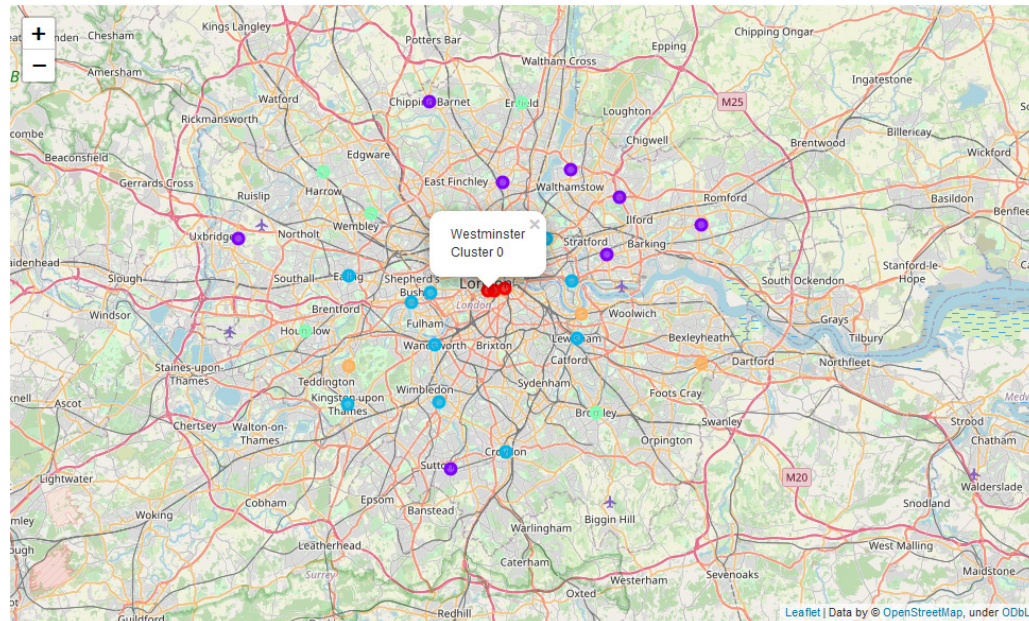
- We will perform **k-means** clustering to group the boroughs according to what amenities they have using Foursquare data. In order to utilize the k-means algorithm, **one hot encoding** has been used to transform the categorical data into a numerical data. Afterwards, the data is grouped by borough and the **mean of frequency of each venue category** is considered in the next steps when modelling.

| BoroughName | Afghan Restaurant | African Restaurant | American Restaurant | Argentinian Restaurant | Art Gallery | Art Museum | Arts & Crafts Store | Asian Restaurant | Athletics & Sports | ... | Vegetarian / Vegan Restaurant | Veterinarian | Video Game Store | Video Store |
|------------------------|-------------------|--------------------|---------------------|------------------------|-------------|------------|---------------------|------------------|--------------------|-----|-------------------------------|--------------|------------------|-------------|
| 0 Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 |
| 1 Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 |
| 2 Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 |
| 3 Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 |
| 4 Barking and Dagenham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 |

| BoroughName | 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 Barking and Dagenham | Supermarket | Grocery Store | Bus Station | Gas Station | Martial Arts Dojo | Bus Stop | Convenience Store | Restaurant | Park | Golf Course |
| 1 Barnet | Coffee Shop | Pub | Park | Grocery Store | Italian Restaurant | Golf Course | Bookstore | Café | Restaurant | Pizza Place |
| 2 Bexley | Pub | Greek Restaurant | Playground | Fast Food Restaurant | Chinese Restaurant | Fruit & Vegetable Store | Steakhouse | Garden | Gastropub | Museum |
| 3 Brent | Coffee Shop | Hotel | Grocery Store | Clothing Store | Indian Restaurant | Sporting Goods Shop | Sandwich Place | Burger Joint | Stadium | Pedestrian Plaza |
| 4 Bromley | Pub | Clothing Store | Gym / Fitness Center | Coffee Shop | Burger Joint | Pizza Place | Indian Restaurant | Supermarket | Stationery Store | Burrito Place |

Methodology

Cluster Analysis



Methodology

Cluster Analysis

Cluster 0: Cultural Area

| | BoroughName | CrimeIndicator | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|-------------|----------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 0 | Westminster | 27.503479 | 0 | Hotel | Coffee Shop | Café | Theater | Park |
| 1 | Southwark | 10.630679 | 0 | Theater | Scenic Lookout | Coffee Shop | Hotel | Grocery Store |
| 4 | Lambeth | 9.446011 | 0 | Theater | Hotel | Scenic Lookout | Park | Art Gallery |

Cluster 1: Vibrant Area

| | BoroughName | CrimeIndicator | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----|----------------------|----------------|----------------|-----------------------|-----------------------|-----------------------|--------------------------|-----------------------|
| 8 | Haringey | 10.036600 | 1 | Turkish Restaurant | Café | Pub | Mediterranean Restaurant | Bakery |
| 3 | Newham | 9.493485 | 1 | Grocery Store | Pub | Indian Restaurant | Park | Café |
| 24 | Barking and Dagenham | 8.317383 | 1 | Supermarket | Grocery Store | Bus Station | Gas Station | Martial Arts Dojo |

Cluster 2: Cosmopolitan Area

| | BoroughName | CrimeIndicator | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----|----------------------|----------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-------------------------------|
| 28 | Merton | 5.829393 | 2 | Park | Pub | Italian Restaurant | Café | Coffee Shop |
| 30 | Kingston upon Thames | 6.459604 | 2 | Café | Pub | Coffee Shop | Burger Joint | Thai Restaurant |
| 25 | Havering | 6.566183 | 2 | Coffee Shop | Pub | Market | Yoga Studio | Vegetarian / Vegan Restaurant |

Cluster 3: Busy Area

| | BoroughName | CrimeIndicator | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----|-------------|----------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 17 | Hounslow | 8.352959 | 3 | Clothing Store | Coffee Shop | Indian Restaurant | Pub | Sandwich Place |
| 9 | Brent | 8.031429 | 3 | Coffee Shop | Hotel | Grocery Store | Clothing Store | Indian Restaurant |
| 13 | Enfield | 7.703921 | 3 | Coffee Shop | Pub | Clothing Store | Sandwich Place | Turkish Restaurant |

Cluster 4: Quiet Area

| | BoroughName | CrimeIndicator | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----|----------------------|----------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 15 | Greenwich | 8.732374 | 4 | Pub | Garden | Park | Café | Market |
| 26 | Bexley | 6.002829 | 4 | Pub | Greek Restaurant | Playground | Fast Food Restaurant | Chinese Restaurant |
| 31 | Richmond upon Thames | 5.621674 | 4 | Pub | Café | Park | Garden | Grocery Store |

Results and Discussion

- The aim of this project is to help people select the **safest area**, assembled by the **venue's similarity**. Now, per one's preferences, newcomers to London can shortlist the areas of their interest.
- It highly depends what people are looking for so if somebody prefers vibrant or busy areas so these can be found in the boroughs clustered as No 1 and 3. On the other side, if somebody is seeking a quiet area with many parks and playgrounds, then cluster No 4 is ideal for him/her. Cluster 0 is cultural one, however, as we can see it has high criminal indicator. Thus, the factor of safety plays an important role when deciding which borough to select.
- In the **top 5 safest boroughs**, 4 types of clusters (Quiet, Busy, Cosmopolitan and Vibrant) are listed, so everybody can find a borough differentiating itself based on the venues' types while the crime indicator remains low.

Conclusion

- From our data-driven analysis, one can find useful information on the **overview of the current safety status** along with **the most common venues**. As one can see, all the safest areas are suburban ones, in south-west, east and north London.
- The analytics comprises of two perspectives of the respective boroughs, so it would be helpful to **incorporate also other factors** as the next steps that might be performed in the future, such as the real estate prices, time to commute to work/school, cost of living, budget limitations, etc. The crime dataset provides the details on the type of the crime so this could be included as well into analyzing safety in order to distinguish the seriousness of the committed crime and this way create weights that could increase the accuracy of the safety indicator.
- The current research successfully performed the high-level analysis and provides people searching for a place to live in London a **recommendation based on the most popular venues types and safety**. This project enables/helps future Londoners to make more informed decision when looking for a safe borough in London.