

Identifying Family- Friendly Neighbourhoods in London, UK

IBM Applied Data Science Capstone

Cristina Gabriel

09/23/2019

1 Introduction

1.1 Problem

Finding a neighbourhood that is family-friendly can be hard. Especially, in big cities like London. Big cities are considered non-kid friendly because they are full of crime, have dangerous busy streets, and no place for kids to run and play.

When looking for a family-friendly neighbourhoods there are many variables to consider. Such as, crime rates, easy access to kid-friendly places such as parks, pools, libraries, and availability of good schools. With all the different variables to consider finding a family friendly place to live in a big city can be overwhelming.

In this project we will try to solve this problem by using location data and machine learning techniques to identify family friendly neighborhoods in London, UK. To determine which neighbourhoods are family-friendly we will focus on neighborhoods with low crime rates and easy access to kid-friendly activities.

1.2 Stakeholders Interest

This project is would be of interest to families looking for new homes, stakeholders who offer housing to families or city planners who are looking to make neighborhoods more family-friendly. Additionally, Police forces and safety offices would benefit from knowing which postal codes are higher crime so they can implement more resources in these locations to lower crime.

2 Data

2.1 Data Utilized

For this project we need data about:

- Crime Rates for each neighbourhood
- Neighbourhood boundaries
- Types of kid-friendly venues (pools, parks, libraries, etc.)
- The Amount of kid-friendly venues in each neighbourhood

2.2 Data Sources

We will extract **Crime Rate** data for each neighbourhood from **Metropolitan Police Crime Dashboard**

This data includes the number of Crime incidents for a specific latitude & Longitude

Example:

	OCU Name	SNT Code	SNT Name	Latitude (generated)	Longitude (generated)	Sum Value
0	Aviation Security(SO18)	SATST5	Heathrow Terminal 5	51.471612	-0.488698	2135

We will define **neighborhood boundaries** by their **Post Code Pre-fix** which we will obtain with **GeoPy's Reverse Geocoding** from the coordinates obtained from the Crime Rate data.

Example:

	SNT Name	Latitude	Longitude	Crime_Incid	Postal_Code
310	Clerkenwell	51.526372	-0.108157	4141	EC1

Kid-friendly Venues data will be obtained by using **Foursquare API**. The total number of venues within that Category will then be calculated for each postal code.

Example:

	Postal_Code	Latitude	Longitude	Sum Value	Park	Library	Pool	Playground	Cinemas
0	BR1	51.411807	0.021526	12549	12	3	3	4	0.0

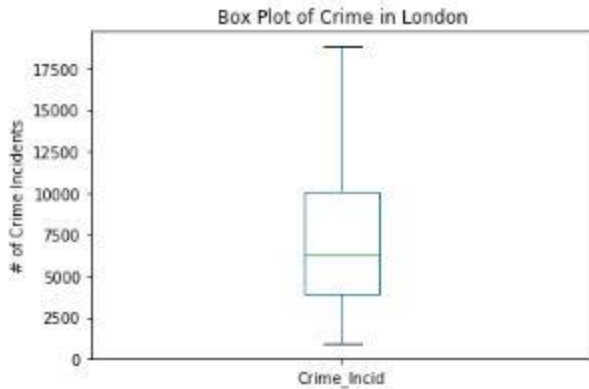
3 Methodology

To identify the most family friendly neighborhood, we used k-means clustering. K-means clustering was chosen because it meets two conditions that make k-means clustering ideal.

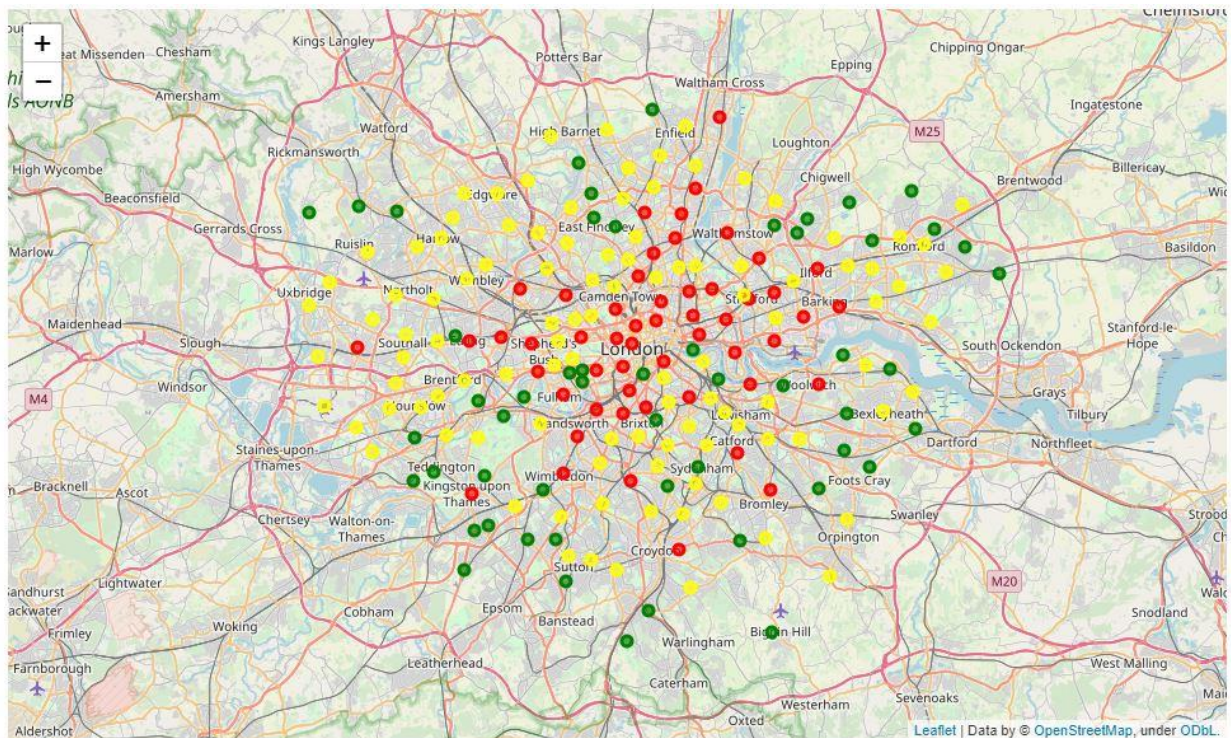
1. We do not have labeled data
2. We know how many clusters we would like (3: unfamily-friendly, family- friend and cautious)

3.1 Crime Data

Before using k-means to identify our neighborhoods we first looked at the distribution of crime incidents. We plotted a box plot to determine what was a “normal” amount of crime in London. We classified normal as the data points that are within the box of the box plot. Anything below the box is considered very safe and anything above the box is considered unsafe.



Once the box plot was generated, we plotted the data points on a map of London to determine if there was any pattern of crime. Such as, the south of London being more dangerous than the north.



From the map it appears there isn't any obvious pattern of crime incidents.

3.2 K-Means

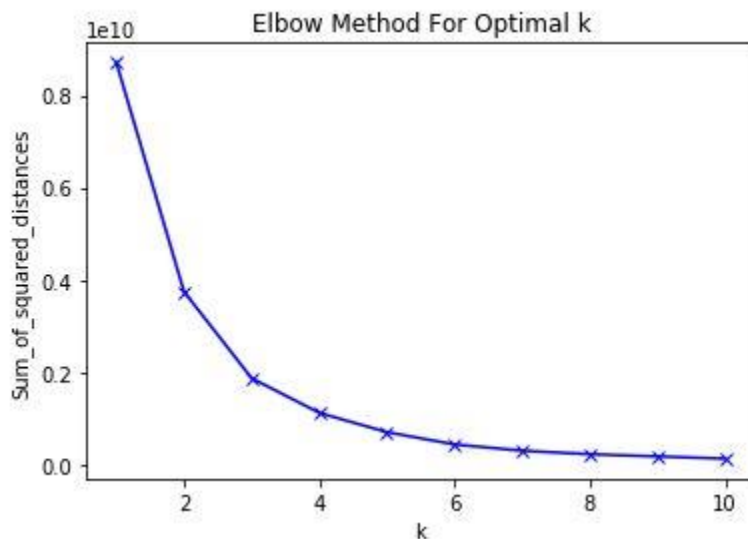
After obtaining the venue data from Foursquare API and summed the total number of venues for each category we applied K-means to the data set.

We did not normalize the data as we wanted the crime data to be the main driving force when clustering the data. We did this because when thinking about a family-friendly neighborhood

the most important factor is safety. Regardless, of the amount of kid-friendly venues are available families would not want to live in high crime areas.

We used 3 as our number clusters to represent the following categories: Family-Friendly, Cautious, and safe.

To ensure we had the optimal number of categories and not missing any hidden observations we plotted the inertia (sum of squared distances) for each data point



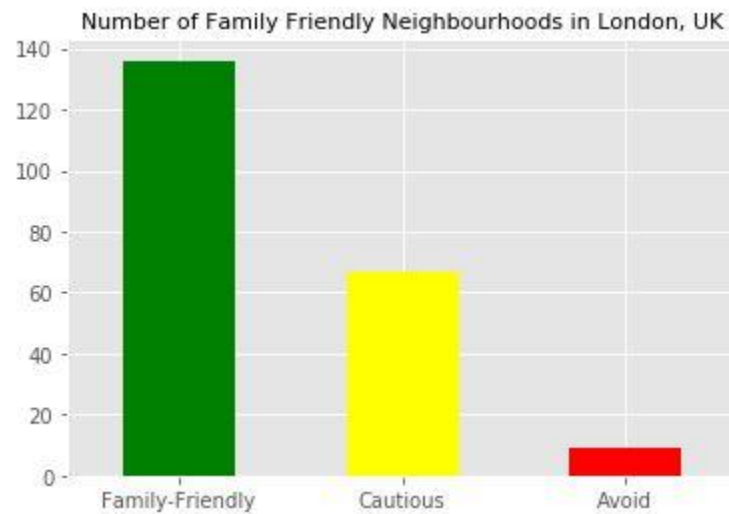
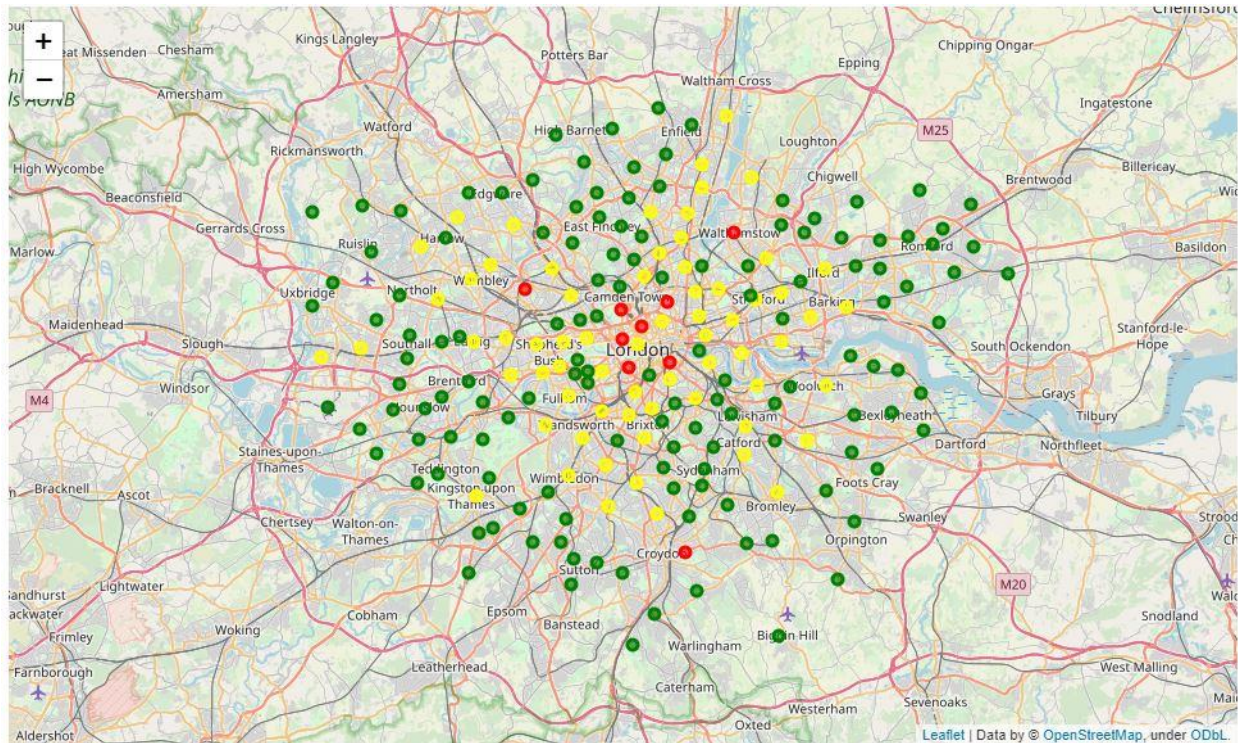
This is also known as the elbow method. We find the elbow point for the data points which is 3.

4 Results

After running K-means we obtained our three clusters corresponding to their categories as follows:

Cluster Labels	Category
0	Family-Friendly
1	Not Family-Friendly
2	Cautious

We again plotted the data points on a map of London to visualize the data



We can see that there are a lot more green dots on the map indicating that there are many family friendly neighborhoods to choose from. In fact, there are 136 family-friendly neighborhoods to choose from and just 9 to be avoided.

5 Discussion

With more family-friendly neighborhoods to choose from then there are to avoid and be cautious of combined. It should be relatively easy for stakeholders to find a neighborhood based on their preferences.

5.1 Stakeholders: Businesses, City-Planners & Developers

It should be noted that some of the neighborhoods ranked as the safest have a lack of kid-friendly venues within a mile radius. This is great news for developers and businesses who wish to find areas to build kid-friendly venues.

Top 10 Safest Postal Codes	
1	UB9
2	TN16
3	RM11
4	SM3
5	SW14
6	BR4
7	RM4
8	IG4
9	KT5
10	E18

5.2 Stakeholders: Families

For families looking for a good balance of low crime and convenient access to kid-friendly venues. The following Post Codes would be good locations to look for a home.

Top 10 Family-Friendly Neighborhoods with the Most Kid-Friendly Venues						
Rank	Post Code	Parks	Libraries	Pools	Playground	Cinemas
1	SE11	50	17	4	30	1
2	E1W	50	12	3	12	2
3	N5	40	11	4	20	1
4	NW8	32	10	2	26	2
5	W10	34	11	6	16	1
6	W9	29	12	2	24	0
7	NW5	29	11	3	22	1
8	N19	31	8	2	22	1
9	SW5	29	13	3	10	2
10	SE5	29	8	4	13	1

5.3 Stakeholders: Police & Safety Offices

Crime rate in these postcodes are very high and hence should be avoided by families. But those in charge of police forces (Police Commissioners) and safety would benefit from implementing more resources into these areas to bring the crime down.

Postal Codes to Avoid
CR0
E17
N1
NW1
NW10
SE1
SW1
W1
WC1

5.4 Further Analysis and Improvement Ideas

- 1. Add more kid-friendly venues for analysis.**

Parks, Playgrounds, Libraries, Pools, and Cinemas are not the only kid friendly activities that could be available.

- 2. Add Boundaries for each Postal Code**

Currently only using a 1-mile radius to define the boundaries of the Postal Codes assigned latitudes and longitudes. This could cause some unwanted overlapping of postal code zones while searching venues. This would also allow for choropleth maps to be made.

- 3. Standardize Crime Incidents based on population within the Postal Code.**

. It is noted that many of the postal codes that are to be avoided are centrally located where the population density is most likely larger than the postal codes further from the central. This could cause the crime rate to look superficially high.

- 4. Filter the results further by budget**

This will help better cater to families of all budgets. London is an expensive city to live in, especially when it comes to housing. The average cost for renting 2-bedroom apartment in London is \$2338 (£1875) making it the 6th most expensive rental market in the world according to Business Insider. London ranks 8th most expensive place to buy in the world based on average property price.

6 Conclusion

After evaluating the crime rates of different postal codes, we then obtained kid friendly activities within a mile of the post codes. Further clustering the results in order to get the most family-friendly neighborhoods based on all our criteria.

The results suggest there are many family-friendly neighborhoods to choose from. Which is great news as many people believe it is impossible to find family-friendly neighborhoods in Big Cities. There are 136 family-friendly Postal Codes located across London. So, stakeholders would not have a problem finding the right neighborhood based on their own preferences, needs & budget.

It should be noted that some of the neighborhoods ranked as the safest do not have many identified kid-friendly activities. But some stakeholders may find this as helpful information because it would be a prime location to open/build kid friendly venues to attract new families to the neighborhood and solve a problem for the families already living there.