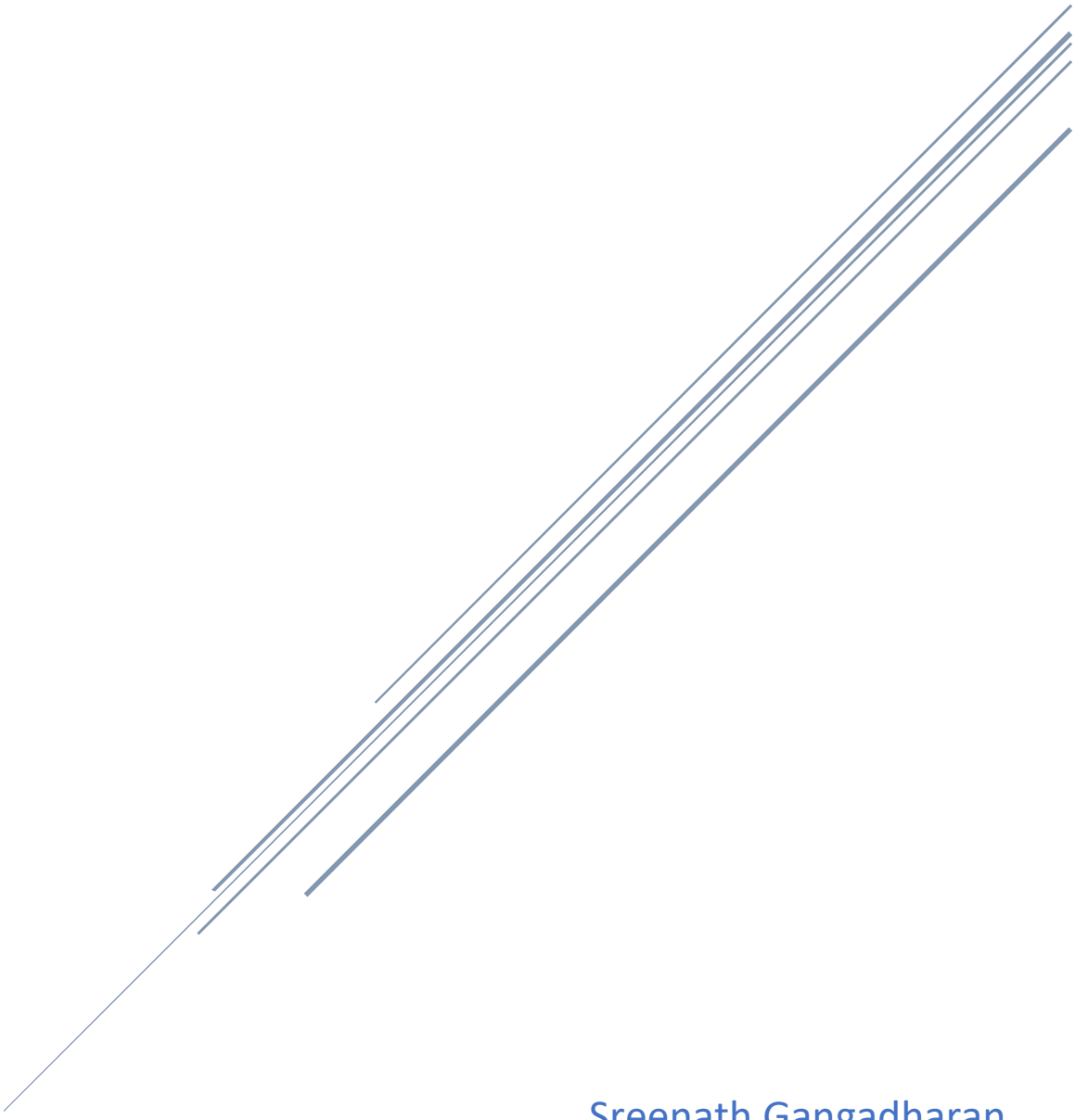
BATTLE OF NEIGHBORHOOD

Capstone Project



Sreenath Gangadharan
July 2020

Finding a Safe Home In London

1.Introduction

1.1 Problem

"*We are moving out to a new Home /Neighborhood* "this is a very common phrase among the American communities. In fact, Americans score the highest among other is the number of moves in lifetime. There may be different reasons for these movements, each stage of life has different requirements and your current community or neighborhood may not be having all those. But what if, you note down most of the requirements you can expect and use it to find it at the Time of choosing a community. It will surely reduce the tendency to move again. It will help you in conserving your time and money and also peace of mind.

In today's world **Safety** is the biggest concern when choosing a community or place to live. Almost all others are secondary and may be little bit adjustable. And we will be trying to find out the Safest Area or Borough in London

1.2 Background

One of my elderly friends approached me to help him in suggesting a safe area for his family to settle down. He had done basic research and feels that safety must be given the most priority.

I choose it as an opportunity to do some research on crime records of London and analyze the data available to find out the safest borough so that it can help him with a good suggestion and to use the model again for future reference

1.3 Interest

At the end of this analysis this will help with anyone who tries to identify the safest borough in London and will help to explore a little bit on the communities and the amenities nearby.

2. Data Acquisition and Cleaning

2.1 Data Used and Source

- **London Crime Data**
- **List of London Boroughs**
- **List of districts in the Royal Borough of Kingston upon Thames**

London Crime Data

London crime data shows the crime per borough in London is taken from Kaggle . The dataset contains the following columns:

- **lsoa_code: Area code**
- **borough : Common name of Borough**
- **major_category: High level crime**
- **minor_category : Low level crime**
- **Value : count of crime**
- **Year : year of record**
- **Month: Month of record**

List of London Boroughs

List of London boroughs is scrapped from the wiki page, this page contains additional information about the boroughs

- **Borough: The names London boroughs.**
- **Inner: Categorizing the borough as an Inner or Outer London Borough.**
- **Status: Categorizing the borough as Royal, City or other.**
- **Local authority: The local authority of borough.**
- **Political control: The political party ruling**
- **Headquarters: HQ of the Boroughs.**
- **Area : Area of the borough .**
- **Population : The population in the borough recorded**
- **Co-ordinates: The latitude and longitude of the boroughs.**
- **Nr. in map: The number assigned to each borough to represent visually on a map.**

List of districts in the Royal Borough of Kingston upon Thames

List of districts in the Royal Borough of Kingston upon Thames is wiki page , we tweaked it to find the below columns

- Neighborhood: Name of the neighborhood in the Borough.
- Borough: Name of the Borough.
- Latitude: Latitude of the Borough.
- Longitude: Longitude of the Borough.

2.2 Data Cleaning

From the London crime data, the crimes during the year 2016 is only selected. The major categories of crime are pivoted to get the total crimes per borough as per the category.

Beautiful Soup Library is used to scrap List of London boroughs of wiki to obtain borough details. Many string manipulation and removal is required to obtain data in desired format.

These two datasets are merged to on Borough Names to get all the required information inside a single dataset , which can help us visualize crimes count against each borough so that we can find which borough has highest and which have lowest crimes against them

Then we will create the third dataset from List of districts in the Royal Borough of Kingston upon Thames using pandas and set the Neighborhood and will fill the latitude and longitude from **geocoding API**

Then we will use this dataset to generate venues of each neighborhood using **FourSquare API**

3. Methodology

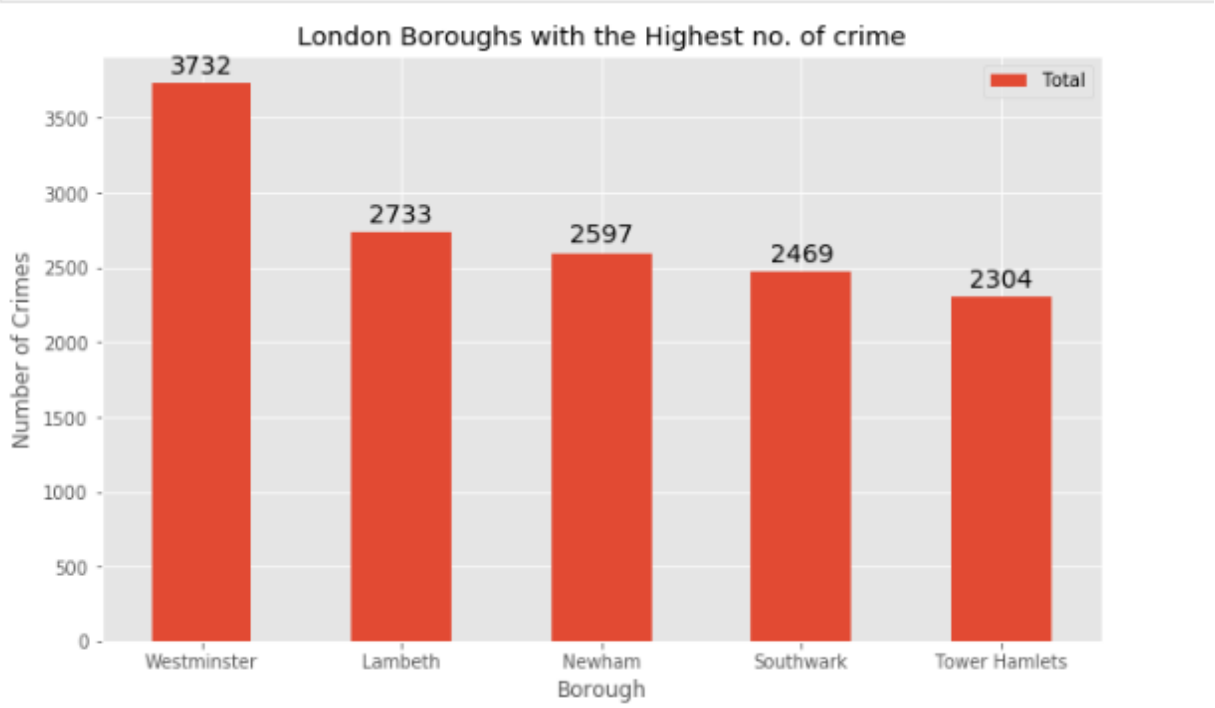
Explanatory Data Analysis

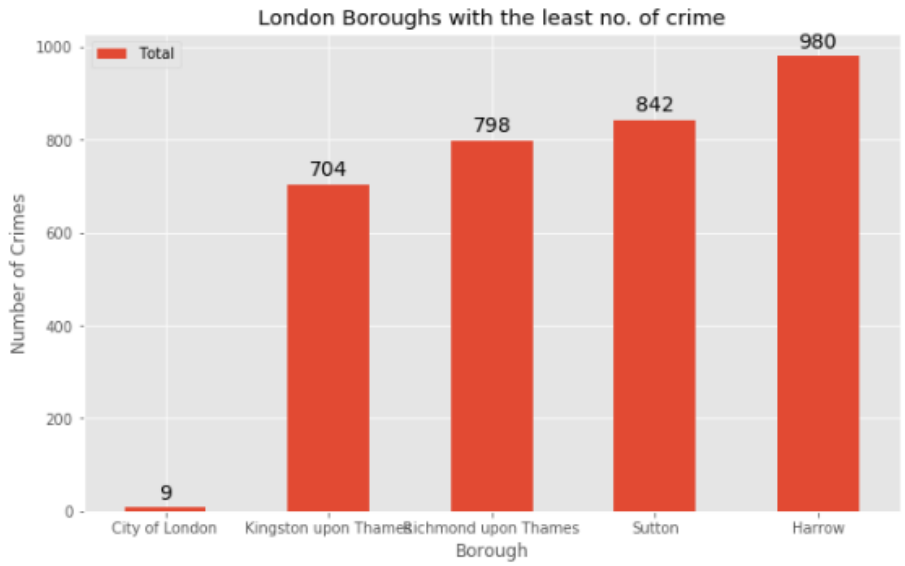
Crime data analysis

- First we created a statistical summary of crimes for all the 33 London borough , removed all data except for the year 2016 ,analyzed it and converted it to a panda data frame with name of crimes as columns and number of occurrence as value

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	112	147	65	30	35	380	460	1229
1	Barnet	264	181	62	43	29	720	583	1882
2	Bexley	77	140	62	18	10	357	354	1018
3	Brent	213	180	145	41	59	689	687	2014
4	Bromley	159	152	59	34	25	557	486	1472

- Visualized the data of top 5 borough with high crimes and least crimes and rejected the top 5 crime boroughs as we have safety as primary concern





The borough city of London and Kingston Upon Thames has the lowest no. of crimes recorded for the year 2016. Let us check the details of the borough

- We neglected city of London due to the extremely less area and the population in the data received and with the fact that it's under London municipal corporation
- We concluded that Kingston Upon Thames have the second least crime rate and decided to explore more on same

Neighborhood Analysis

- Collected the data of all 15 Neighborhoods Royal Borough of Kingston Upon Thames of from the wiki
- Added Latitude and longitude to the dataset using geocoder

Modelling

- We used foursquare API to find all the venues within 500 meter radius of each neighborhood of Royal Borough Kingston upon of Thames

- we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use a cluster size of 5 for this project that will cluster the 15 neighborhoods into 5 clusters

Visualising the clustered neighborhoods on a map using the folium library (see fig 4.6).

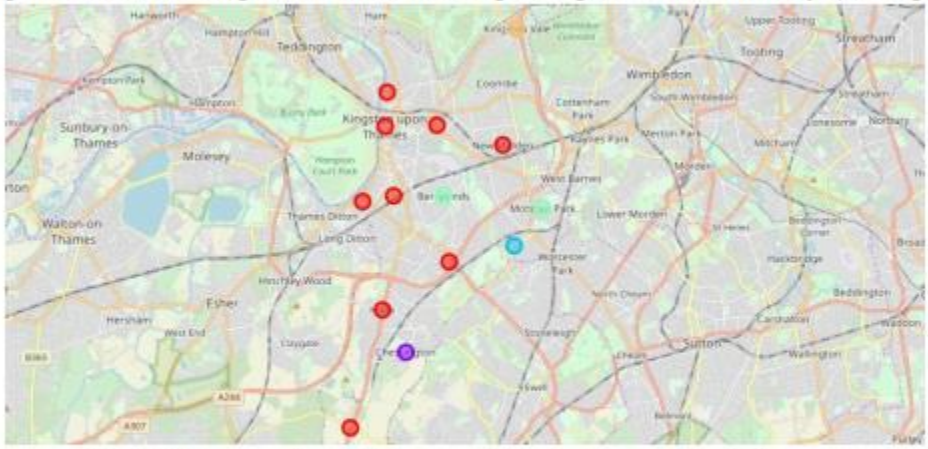


Fig 4.6 Clustered neighborhoods in the Borough of Kingston upon Thames

Cluster 1

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	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
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	1	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Discount Store	Electronics Store	Farmers Market	Fast Food Restaurant
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Grocery Store	Pub	Garden Center	Restaurant	Wine Shop	Fast Food Restaurant	Department Store	Discount Store	Electronics Store
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1	Grocery Store	Pharmacy	Restaurant	Discount Store	Hotel	Bus Stop	Italian Restaurant	Sandwich Place	Bowling Alley

Cluster 2

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	2	Gym / Fitness Center	Park	Bus Stop	Wine Shop	Food	Discount Store	Electronics Store	Farmers Market	Fast Food Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	2	Gym	Park	Bus Stop	Restaurant	Soccer Field	Fast Food Restaurant	Department Store	Discount Store	Electronics Store

Cluster 3

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	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	9 C V
1	Canbury	Kingston upon Thames	51.417499	-0.305553	3	Pub	River	Café	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Fish & Chips Shop	P
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	3	Café	Pub	Sushi Restaurant	Coffee Shop	Burger Joint	Asian Restaurant	German Restaurant	Gift Shop	F R
9	New Malden	Kingston upon Thames	51.405335	-0.263407	3	Gym	Sushi Restaurant	Korean Restaurant	Bar	Supermarket	Indian Restaurant	Gastropub	Garden Center	F H S
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	3	Food	Indian Restaurant	Italian Restaurant	Pub	Wine Shop	Rental Car Location	Hardware Store	Hotel	J R
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	3	Indian Restaurant	Coffee Shop	Pub	Gym	Pet Café	Playground	Chinese Restaurant	Café	R
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	3	Coffee Shop	Pub	Pharmacy	Italian Restaurant	Grocery Store	Gym / Fitness Center	Deli / Bodega	Farmers Market	G

Cluster 4

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	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
4	Hook	Kingston upon Thames	51.367898	-0.307145	4	Bakery	Supermarket	Indian Restaurant	Fish & Chips Shop	Wine Shop	French Restaurant	Discount Store	Electronics Store	Farmers Market

4.Results and Recommendations

This aim of the project was to find the safest borough and to determine the best neighborhood in that Borough, We had done the research with the available information and provided 4 clusters . the user can easily select from the above clusters

Recommendations are just based on the data and doesn't consider the personal preferences and sentiments

- If you are considering public transport and entertainment, then cluster 2 is recommended
- If you are considering an area with good restaurants, pubs and bar cluster 3 is recommended

4. Conclusion

We had made a recommendation based on the venue data available, We had taken safety as primary concern and then most common venue to create cluster among that place. Its always better to physically visit the place before taking decision. Also coast of living and real estate cost is not taken into consideration, which can be later added as an enhancement to this project