Capstone Project – The Battle of Neighborhoods Report

Moses Kiplagat

January 2019

1. Introduction

1.1 Background

The average American moves about eleven times in their lifetime. This brings us to the question: Do people move until they find a place to settle down where they truly feel happy, or do our wants and needs change over time, prompting us to eventually leave a town we once called home for a new area that will bring us satisfaction in terms of career or retirement?

To minimize the chances of this happening, we should always do proper research when planning our next move in life. Consider the following factors when picking a new place to live so you don't end up wasting your valuable time and money, making a move you will end up regretting. Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.

For this problem, I assume the client is considering a move to London, and so I am choosing it as the case study for this problem.

1.2 Problem

The crime statistics dataset of London found on Kaggle has crimes in each Boroughs of London from 2008 to 2016. The year 2016 being the latest we will be considering the data of that year which is actually very old information as of now. The crime rates in each borough may have changed over time. This project aims to select the safest borough in London based on the total crimes, explore the neighborhoods of that borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering. By doing this, it gives a focus area where someone can relocate to.

1.3 Interest

Expats who are considering relocation to London will be interested to identify the safest borough in London and explore its neighborhoods and common venues around each neighborhood. This way, they can focus their research more on costs of housing in the area, schools, facilities offered etc.

2. Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project is a combination of data from three sources. The first data source of the project uses a London crime data that shows the crime per borough in London. The dataset contains the following columns:

- **lsoa_code** : code for Lower Super Output Area in Greater London.
- **borough** : Common name for London borough.
- **major_category** : High level categorization of crime
- **minor_category** : Low level categorization of crime within major category.
- **value** : monthly reported count of categorical crime in given borough
- **year** : Year of reported counts, 2008-2016
- **month** : Month of reported counts, 1-12

The second source of data is scraped from a wikipedia page that contains the list of London boroughs . This page contains additional information about the boroughs, the following are the columns:

- **Borough** : The names of the 33 London boroughs.
- **Inner** : Categorizing the borough as an Inner London borough or an Outer London Borough.
- **Status** : Categorizing the borough as Royal, City or other borough.
- **Local authority** : The local authority assigned to the borough.
- **Political control** : The political party that control the borough.
- **Headquarters:** Headquarters of the Boroughs.
- **Area (sq mi)** : Area of the borough in square miles.
- **Population (2013 est)[1]** : The population in the borough recorded during the year 2013.
- **Co-ordinates** : The latitude and longitude of the boroughs.
- **Nr. in map** : The number assigned to each borough to represent visually on a map.

The third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on a wikipedia page. This dataset is created from scratch using the list of neighborhood available on the site, the following are 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

The data preparation for each of the three sources of data is done separately. From the London crime data, the crimes during the most recent year (2016) are only selected. The major categories of crime are pivoted to get the total crimes per borough as per the category (see *fig 2.1*).

[15]:

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	1287	1949	919	378	534	5607	6067	16741
1	Barnet	3402	2183	906	499	464	9731	7499	24684
2	Bexley	1123	1673	646	294	209	4392	4503	12840
3	Brent	2631	2280	2096	536	919	9026	9205	26693
4	Bromley	2214	2202	728	417	369	7584	6650	20164

Fig 2.1 London crime data after preprocessing

The second data is scraped from a wikipedia page using the **Beautiful Soup** library in python. Using this library we can extract the data in the tabular format as shown in the website. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form (see *fig 2.2*). This is important because we will be merging the two datasets together using the Borough names.

22]:

	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

Fig 2.2 List of London Boroughs

The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset (see *fig 2.3*). The purpose of this dataset is to visualize the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016.

	Borough	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	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	1287	1949	919	378	534	5607	6067	16741
1	Barnet	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	3402	2183	906	499	464	9731	7499	24684
2	Bexley	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	1123	1673	646	294	209	4392	4503	12840
3	Brent	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	2631	2280	2096	536	919	9026	9205	26693
4	Bromley	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	2214	2202	728	417	369	7584	6650	20164

Fig 2.3 London Borough Crime

After visualizing the crime in each borough we can find the borough with the lowest crime rate and hence tag that borough as the safest borough. The third source of data is acquired from the list of neighborhoods in the safest borough on Wikipedia. This dataset is created from scratch, the pandas data frame is created with the names of the neighborhoods and the name of the borough with the latitude and longitude are left blank (see *fig 2.4*).

44]:

	Neighborhood	Borough	Latitude	Longitude
0	Berrylands	Kingston upon Thames		
1	Canbury	Kingston upon Thames		
2	Chessington	Kingston upon Thames		
3	Coombe	Kingston upon Thames		
4	Hook	Kingston upon Thames		
5	Kingston upon Thames	Kingston upon Thames		
6	Kingston Vale	Kingston upon Thames		
7	Malden Rushett	Kingston upon Thames		

Fig 2.4 Neighborhoods of the safest borough

The coordinates of the neighborhoods is obtained using Google Maps API geocoding to get the final dataset (See *Fig 2.5*).

]:

	Neighborhood	Borough	Latitude	Longitude
0	Berrylands	Kingston upon Thames	51.393781	-0.284802
1	Canbury	Kingston upon Thames	51.417499	-0.305553
2	Chessington	Kingston upon Thames	51.358336	-0.298622
3	Coombe	Kingston upon Thames	51.419450	-0.265398
4	Hook	Kingston upon Thames	51.367898	-0.307145
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898
9	New Malden	Kingston upon Thames	51.405335	-0.263407
10	Norbiton	Kingston upon Thames	51.409999	-0.287396
11	Old Malden	Kingston upon Thames	51.382484	-0.259090
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366
13	Surbiton	Kingston upon Thames	51.393756	-0.303310
14	Tolworth	Kingston upon Thames	51.378876	-0.282860

Fig 2.5 Neighborhoods of the safest borough

The new dataset is used to generate the venues for each neighborhood using the Foursquare API.