Capstone Project:The battle of Neighborhoods - Sanjana Dasari

1. Introduction

• Background: 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.

• Problem: 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.

· Interest: Expats who are considering to relocate to London will be interested to identify the safest borough in London and explore its neighborhoods and common venues around each neighborhood.

Boroughs with the lowest crime rates
Neighborhoods in Kingston upon Thames
Modelling
London Boroughs with the least no. of crime
12:55
· Using th
12:01

2. Data Acquisition and Cleaning

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 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 third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found

on the wikipedia page.

Data Cleaning: The data clearing press for each of the three sources of data are done separately

. From the London cime dala, the crimes during the most recent year (2016) are only
Selecind. The mo
r ies of crime are valed to othe total crimes per the boroughs lar ech jur category

· The second data is scraped from a Wikipedia are using the Beautiful Soubibrary in python Using this library we can extract the cria in the tabular format as shown in the website.

3. Methodology

· The two dala seis are merged on the Borough mes lo on ahdala sel. The purpose of
#Hà | sim vi la la la la la. Thi Hi Full li a = b El Minh và liEl H. Bình I = n Hàn khiữ

Data Cleaning: The data cleaning process for each of the three sources of data are 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 the boroughs for each major category.

- 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.

- The two data sets are merged on the Borough names to form a new data set. The purpose of this data set is to visualize the crime rates in each borough and identify the

borough with the least crimes recorded during the year 2016.

• After visualizing the crime in each borough we can find the borough with the lowest crime rate. The third data set is created, with the names of the neighborhoods and the name of the borough with the latitude and longitude obtained using Google Maps API geocoding.

• The new data set is used to generate the 10 most common venues for each neighborhood using the Foursquare API, finally using k means clustering algorithm to cluster similar neighborhoods together.

3. Methodology

Exploratory Data Analysis

Statistical summary of crimes

	Burglary	Criminal Damage	Drugs	other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000					

33.000000
33.000000
33.000000
33.000000
33.000000

mean 2069.242424
479.060606
682.666667
8913.121212
7041.848485 22306.696970
1941.545455 1179.212121 625.207070 586.406416

std
737.448644
223.298698
441.425366
4620.565054
2513.601551
8828.228749

min
2.000000
2.000000
10.000000
6.000000
4.000000
129.000000
25.000000
178.000000

1650.000000
743.000000
378.000000
377.000000
5919.000000
5936.000000 16903.000000 7409.000000 22730.000000
1989.000000 1063.000000
490.000000
599.000000
8925.000000
25% 1531.000000 **50%** 2071.000000 75% 2631.000000

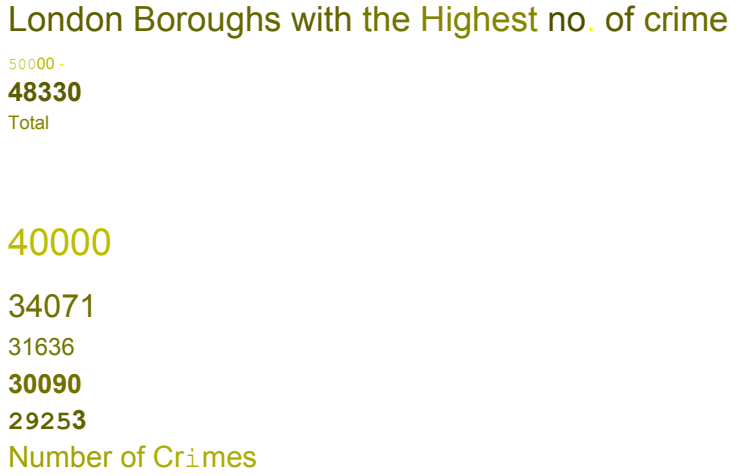
max 3402.000000
2351.000000 1617.000000
551.000000
936.000000

10789.000000
8832.000000 27174.000000

3219.000000 2738.000000
1305.000000 1822.000000
27520.000000
10834.000000 48330.000000

The count for each of the major categories of crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by Violence against the person', 'Criminal damage. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'

Boroughs with the highest crime rates



10000 -

Westminster
Lambeth
Newham
Tower Hamlets
Southwark Borough

Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs.

Boroughs with the lowest crime rates

London Boroughs with the least no. of crime

12755
Total

12000

11074
10658
9983

10000

Number of Crimes

178

City of London

Kingston upon Thames
Richmond upon Thames
Merton

Sutton Borough

Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton.

• City of London has a significantly lower crime rate because it i is the 33rd principal division of Greater London but it is not a London borough. It has an area of 1.12 square miles and a population of 7000 as of 2013 which suggests that it is a small area.

• We will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

Neighborhoods in Kingston upon Thames

St Margarets
VydriuSWUTLET
BRIX
+ EDFONT
Whitton
Roehampton
Twickenham
Feltham

Balham
Richmond
Park
Southfields

Strawberry
Hill
Streathar
Hill
Putney Vale
Wimbledon
Park
Hanworth
Ham
Upper Tooting
Hampton Hill
KingsOn Vale W*imbledo*n
C*omm*on
Streatham
Teddington

Kempton Park
Tooting Wimbledon

South Wimbledon
CoObe
Streatham
Hampton
*Bu*shy Park
Cottenham
Park
Charlton
Kingston upon o
Thames
Sunbury-on
Thames
Lonesom*e No*rbury

New Olden
"Raynes Park Merton Park

Mitcham
Molesey
Hampton Court Park
Morden
1 West Barnes
pperton
O
Berronds
Mots o Park:
Lower Morden
Walton-on
Thames
Thames Ditton
Long Ditton
St Helier
Beddington
Comer

Worcester
Park
Hackbridge
Hinchley Wood
ridge
North Cheam
Beddington
Esher
Hersham
Carshalton
-Wadd
West End
Stoneleigh
Claygate
Sutton
ChesOgton
B365
Cheam
--Wallington

Ewell
A244

Belmont
A307

Pur
- *Banst*ead
*Dow*ns
A2022
*Pri*nce's *COVETTs*
Oxshott
Epsom
Leaflet

There are 15 neighborhoods in the royal borough of Kingston upon Thames, they are visualised on a map using folium on python.

Modelling

• Using the final data set containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API.

Neighborhood Neighborhood Latitude Neighborhood Longitude
Venue venue Latitude venue Longitude
Venue Category

Berrylands
51.393781
-**0.284802** Surbiton Racket & Fitness Club
51.392676
-0.290224 Gym / Fitness Center

Berrylands
51.393781
-0.284802
Alexandra Park
51.394230
-0.281206
Park

Berrylands
51.393781
-0.284802
51.392302
-0.281534
Bus Stop
K2 Bus Stop

Cafe Rosa

3
Berrylands
51.393781
-0.284802
51.390175
-0.282490
Café

Canbury

51.417499
-0.305553
The Boater's Inn
51.418546
-0.305915
Pub

• One hot encoding is done on the venues data. The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.

To help people find similar neighborhoods in the safest borough 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.

- *W*e will use a cluster size of 5 for this project that will cluster the 15 neighborhoods into 5 clusters. The reason to conduct a K- means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

4. Results

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Visualizing the clustered neighborhoods on a map using the folium library.

Ham
Upper Tooting
Hampton Hill
Kingston Vale Wi*mbledo*n
Common
Streatham
Teddington
Tooting Wimbledon
South Wimbledon
Kempton Park
Coombe
Streatham
Hampton
Bushy Pa
Cottenham
Park
Fiton
Kingston upon
Thares
Sunbury-on
Thames
Lonesome Norbury
Newlden
Raynes Park Merton Park
Mitcham
Molesey
*H*ampton *Court P*ark
Morden
West Barnes
rton
O
Berylinds
Mots pu Park
Lower Morden
Walton-on
Thames
Thames Ditton
Long Ditton
St Helier
Beddington
Corner
Broad

Worcester
Park
Hackbridge

Hinchley Wood
North Cheam
Beddington
Esher

Hersham
Carshalton
--Waddon
West End
Stoneleigh
Claygate
Sutton
Chegton
B365
-Wallington

Ewell
A244

Belmont
A307

Each cluster is color coded for the ease of presentation, we can see that majority of the neighborhood falls in the red cluster which is the first cluster. Three neighborhoods have their own cluster (Blue, Purple and Yellow), these are clusters two three and five. The green cluster consists of two

neighborhoods which is the 4th cluster.

Cluster 1: Looking into the neighborhoods in the first cluster

Neighborhood Borough
Latitude Longitude
Cluster
Labels
1st Most Common
Venue
2nd Most Common
Venue
3rd Most Common
Venue
4th Most Common
Venue
5th Most 6th Most 7th Most 8th Most Common Common Common Common
Venue Venue Venue Venue

Canbury
Kingston
upon 51.417499 -0.305553 **Thames**
Pub
Café
Plaza
Fish & Chips
Shop
Supermarket

Spa
Shop & Service
Park

Hook
Kingston
upon 51.367898 **Thames**
-0.307145
Bakery
Convenience
Store
Indian **Restaurant**
Fish & Chips
Shop
Wine Shop "
Food Electronics
Store
Farmers Market

Kingston upon
Thames
Kingston
upon 51.409627 **Thames**
Cosmetics
-0.306262
0 Coffee Shop
Café
Sushi **Restaurant**
Burger Joint
Record
Shop
Pub
Market
Shop

Malden Rushett

Kingston

upon 51.341052 **Thames**
-0.319076

o Convenience

Pub
Garden

Center
Restaurant

Fast Food **Restaurant**

Discount

Store

Dry **Cleaner**

Electronics

Store
Store

Sushi
New Malden

Kingston

upon 51.405335 −0.263407 **Thames**

0

Gastropub

Gym

Korean Indian **Restaurant Restaurant**

Fish & Chips Shop

Dry Cleaner

Restaurant S*u*permarket

Norbiton

Kingston

upon 51.409999 Thames

−0.287396

Indian **Restaurant**

Pub
Food

Italian **Restaurant**

Platform
Grocery
Store
Farmers
Market
Dry Cleaner

Seething Wells
Kingston
upon 51.392642 **Thames**
-0.314366
Indian **Restaurant**
Coffee Shop
Italian **Restaurant**
Pub
Café
Wine Fast Food Chinese Shop Restaurant Restaurant

Grocery
Surbiton
Kingston
upon 51.393756 **Thames**
-0.303310
Coffee Shop
Pub Supermarket
Breakfast
Spot
Gastropub
French **Restaurant**
Train Station
Store

Tolworth
Kingston
upon 51.378876 **Thames**
-0.282860

Grocery
Store
Pharmacy
Furniture / Home Store
Train Station
Pizza Place
Discount
Store
Coffee
Shop
Bus Stop Du

The cluster one is the biggest cluster with 9 of the 15 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Restaurants, Pubs, Cafe, Supermarkets, and stores

Cluster 2: Looking into the neighborhoods in the second cluster.

Cluster
1st Most Common
Neighborhood Borough
Latitude Longitude
2nd Most Common
vente
3rd Most Common
Venue
4th Most 5th Most 6th Most 7th Most 8th Most 9th Most Common Common
Common Common Common Common
Venue Venue Venue Venue Venue Venue
Labels
Venue
Venue

Kingston
upon 51.358336 Thames
French
Chessington
-0.298622
Fast Food Restaurant
Wine Shop
Golf German Course Restaurant
Fried Chicken
Garden
Center
Gastropub
Furniture / Home

Store
Joint Restaurant

The second cluster has one neighborhood which consists of Venues such as Restaurants, Golf courses, and wine shops.

Cluster 3: Looking into the neighborhoods in the third cluster.

Neighborhood Borough
Latitude Longitude
Cluster Labels
1st Most Common
Venue
2nd
Most Common
Venue
3rd Most 4th Most 5th Most 6th Most 7th Most Common Common
Common Common Common
Venue Venue Venue Venue Venue
8th Most 9th Most Common Common
Venue Venue

Old Malden
Kingston
upon 51.382484 **Thames**
-0.25909
2
Train Station
Pub
Pub
Food Gastropub
Food Gast
Garden
Center
Furniture / Home
Store
Fried Chicken
Joint
French **Restaurant**
Deli / **Bodega**

The third cluster has one neighborhood which consists of Venues such as Train stations, Restaurants, and Furniture shops.

Cluster 4： Looking into the neighborhoods in the fourth cluster.

Neighborhood Borough
Latitude Longitude

Cluster Labels
1st Most Common

Venue
2nd

Most Common
Venue
3rd Most 4th Most 5th Most 6th Most 7th Most 8th Most 9th Most Common
Common Common Common Common Common Common Common

Venue Venue Venue Venue Venue Venue Venue

Wine
Berrylands
Kingston
upon 51.393781 -0.284802 **Thames**

Gym/Fitness

3

Park

Café Bus Stop

Fish & Chips Shop

Electronics

Store

Farmers Fast Food Market Restaurant

Shop

Center

Motspur Park
Kingston
upon 51.390985 Thames

-0.248898

3

Park

i Restaurant

Soccer

Field

Bus Stop

Wine Fast Food Shop Restaurant

Dry Electronics Cleaner **Store**

The fourth cluster has two neighborhoods in it, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields etc.

Cluster 5: Looking into the neighborhoods in the fourth cluster.

Cluster

Neighborhood Borough

Latitude Longitude

1st Most Common

Venue
2nd
Most Common
umon Venue
3rd Most 4th Most 5th Most 6th Most 7th Most Common Common Common Common Common
Venue Venue Venue *V*enue Venue
8th Most Common
Venue
9th Most Common
Venue
Labels

Kingston Vale
Kingston
upon 51.43185 **Thames**
-0.258138
Grocery
Store
Bar
Italian **Restaurant**
Soccer
Field
Garden **Center**
Furniture / Home
Store
Fried Chicken
Joint
French Department Restaurant
Store

The fifth cluster has one neighborhood which consists of

Venues such as Grocery shops, Bars, Restaurants, Furniture shops, and Department stores.

5. Discussion

• The aim of this project is to help people who want to relocate to the safest borough in London, expats can chose the neighborhoods to which they want to relocate based on the most common venues in it.

· For example if a person is looking for a neighborhood with good
connectivity and public transportation we can see that Clusters 3 and 4 have Train stations and Bus stops as the most common venues.

• If a person is looking for a neighborhood with stores and restaurants in a
close proximity then the neighborhoods in the first cluster is suitable.

• For a family I feel that the neighborhoods in Cluster 4 are more suitable dues to the common venues in that

cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is ideal for a family.

• The preference of venues may vary from person to person, they can select a neighborhood based on ones priorities.

6.Conclusion

• This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood.

• We have just taken safety as a primary concern to shortlist the safest borough of London. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough, such as filtering areas based on a predefined budget.