CAPSTONE PROJECT – THE BATTLE OF NEIGHBORHOODS REPORT

THIVIYAA KSHATRIYAR

APRIL 29 2020

CHAPTER – 1

INTRODUCTION

1.1 BACKGROUND

“Migration is an expression of the human aspiration for dignity, safety and a better future. It is part of the social fabric, part of our very make-up as a human family”

This quote of Ban Ki-Moon states the mere fact of people migrating to places. This carries us to the question: Do individuals move until they discover a spot to settle down where they really feel upbeat, or then again do our needs and needs change after some time, inciting us to in the long run leave a town we when considered home for another zone that will bring us fulfilment? Or on the other hand, do we over and over again move to another territory without knowing precisely what we’re getting into, compelling us to retreat in fear at the principal indication of uneasiness?

To limit the odds of this incident, we ought to consistently do legitimate research when arranging our best course of action throughout everyday life. Consider the accompanying elements when picking another spot to live so you don’t wind up burning through your important time and lucrative a move, you’ll end up lamenting. Wellbeing is a top concern when moving to another territory. On the off chance that you don’t have a sense of security in your own house, you’re not going to have the option to appreciate living there.

1.2 LOCATION DATA PROVIDERS

Location data is information relating to the position of mobile phones, tablets, and laptops, or structures like historical monuments, buildings, and attractions. The most common attributes of location data are latitude and longitude (lat/long), usually expressed in coordinates which correlate to a geographical position. One of the most notable firms in the area information advertise, Foursquare currently runs a free area information and innovation stage for advertisers and designers. Utilizing Four square’s area information, brands can measure the effect of their media (utilizing approved stops) and find new crowds that they can focus with pertinent promoting, Foursquare additionally offers a Places API and a Pilgrim SDK, which empower area disclosure inside applications and sites, just as continuous area identification.

1.3 NEIGHBORHOOD SEGMENTATION AND CLUSTERING

In order to explore the neighborhoods, segment them, and group them into clusters to find similar neighborhoods clustering is used. Clustering is a form of unsupervised machine learning. K-Means clustering is the most common form of clustering used. K-means can group data only unsupervised based on the similarity of objects to each other. There are various types of clustering algorithms such as partitioning, hierarchical, or density-based clustering. K-means is a type of partitioning clustering. It divides the data into k non-overlapping subsets or clusters without any cluster internal structure or labels. This means, it's an unsupervised algorithm.

1.4 PROBLEM DEFINITION

The wrongdoing measurements dataset of London found on Kaggle has violations in every ward of London from 2015 to 2016. The year 2016 we will considered and the information of that year which is really old data starting at now. The crime percentages in every precinct may have changed over time. This venture intends to choose the most secure district in London dependent on the all-out wrongdoings, investigate the areas of that ward to locate the 10 most regular settings in every area lastly group the areas utilizing k-mean clustering.

1.5 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.

CHAPTER 2

DATA COLLECTION AND SCOURING

2.1 DATA COLLECTION

The information obtained for this venture is a blend of information from three sources. The main information wellspring of the task utilizes a London wrongdoing information that shows the wrongdoing per precinct in London. The dataset contains the accompanying segments:

● lsoa_code: code for Lower Super Output Area in Greater London.

● district: Common name for London precinct.

● major_category: High level arrangement of wrongdoing

● minor_category: Low level arrangement of wrongdoing inside significant class.

● esteem: month to month detailed include of all out wrongdoing in given district

● year: Year of detailed checks, 2015-2016

● month: Month of detailed checks, 1-12

The second wellspring of information is scratched from a Wikipedia page that contains the rundown of London wards. This page contains extra data about the wards, coming up next are the sections:

● Borough: The names of the 33 London wards.

● Inner: Categorizing the ward as an Inner London district or an Outer London Ward.

● Status: Categorizing the ward as Royal, City or other district.

● Local power: The neighborhood authority allocated to the precinct.

● Political control: The ideological group that control the precinct.

● Headquarters: Headquarters of the Boroughs.

● Area (sq mi): Area of the ward in square miles.

● Population (2016 est)[1]: The populace in the ward recorded during the year 2016.

● Co-ordinates: The scope and longitude of the wards.

● Nr. in map: The number allotted to every precinct to speak to outwardly on a guide.

The third information source is the rundown of Neighborhoods in the Royal Borough of Kingston upon Thames as found on a Wikipedia page. This dataset is made with the help of Foursquare location data and preparatory utilizing the rundown of neighborhood accessible on the site, coming up next are segments:

● Neighborhood: Name of the area in the Borough.

● Borough: Name of the Borough.

● Latitude: Latitude of the Borough.

● Longitude: Longitude of the Borough.

2.2 DATA CLEANING

The information groundwork for every one of the three wellsprings of information is done independently. From the London wrongdoing information, the violations during the recent years are just chosen. The significant classes of wrongdoing are turned to get the complete violations per the districts for each significant class.

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	0	5	0	0	0	3	2	20
1	Barnet	3	0	0	0	0	1	4	16
2	Bexley	0	1	0	0	0	1	1	6
3	Brent	0	2	1	0	1	1	0	10
4	Bromley	0	2	1	0	3	6	8	40

The subsequent information is scratched from a Wikipedia page utilizing the Beautiful Soup library in python. Utilizing this library, we can separate the information in the forbidden organization as appeared in the site. After the web scratching, string control is required to get the names of the precincts in the right structure. This is significant on the grounds that we will be combining the two datasets together utilizing the Borough names.

	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	Barnet House, 2 Bristol Avenue, Colindale	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

The two datasets are converged on the Borough names to shape another dataset that consolidates the vital data in one dataset. The reason for this dataset is to envision the crime percentages in every district and recognize the precinct with the least wrongdoings recorded during the year.

	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	0	5	0	0	0	3	2	20
1	Barnet	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369088	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W	3	0	0	0	0	1	4	16
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	0	1	0	0	0	1	1	6
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	0	2	1	0	1	1	0	10
4	Bromley	Bromley London Borough	Conservative	Civic Centre, Stockwell Close	57.97	317899	51°24′14″N 0°01′11″E / 51.4039°N	0	2	1	0	3	6	8	40

In the wake of envisioning the wrongdoing in every precinct we can discover the ward with the most reduced wrongdoing rate and thus label that ward as the most secure precinct. The third wellspring of information is obtained from the rundown of neighborhoods in the most secure district on wikipedia. This dataset is made without any preparation, the pandas information outline is made with the names of the areas and the name of the district with the scope and longitude left clear.

	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		
8	Motspur Park	Kingston upon Thames		
9	New Malden	Kingston upon Thames		
10	Norbiton	Kingston upon Thames		
11	Old Malden	Kingston upon Thames		
12	Seething Wells	Kingston upon Thames		
13	Surbiton	Kingston upon Thames		
14	Tolworth	Kingston upon Thames		

The coordinates of the areas is be gotten utilizing Google Maps API geocoding to get the last dataset.

	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

The new dataset is utilized to produce the 10 most basic settings for every area utilizing the Foursquare API, at long last utilizing k-means clustering calculation to group comparable neighborhoods together.

CHAPTER 3

METHODOLOGY

3.1 EXPLORATORY DATA ANALYSIS

3.1.1 STATISTICAL SUMMARY OF CRIMES

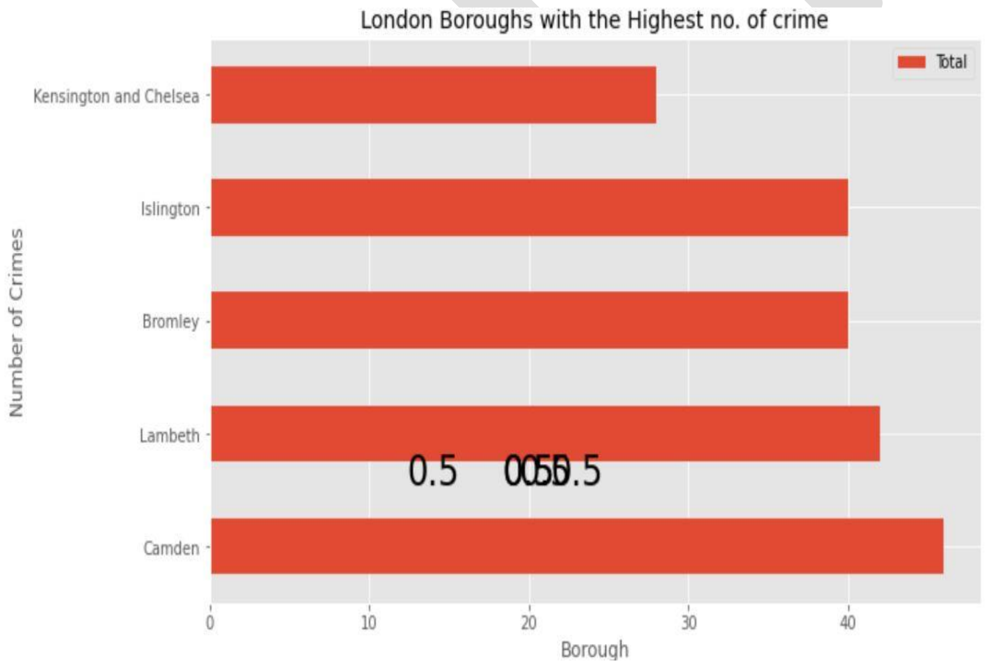
The depict work in python is utilized to get insights of the London wrongdoing information, this restores the mean, standard deviation, least, greatest, first quartile (25%), second quartile (half), and the third quartile (75%) for every one of the significant classifications of wrongdoing.

	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000
mean	0.812500	0.968750	0.531250	0.281250	0.281250	3.687500	2.406250	17.937500
std	1.148281	1.447676	0.983226	0.812578	0.683179	4.895274	2.757504	11.387422
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	9.500000
50%	0.000000	0.500000	0.000000	0.000000	0.000000	1.500000	2.000000	16.000000
75%	1.000000	1.000000	1.000000	0.000000	0.000000	4.250000	3.000000	20.500000
max	4.000000	6.000000	4.000000	3.000000	3.000000	21.000000	12.000000	46.000000

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.

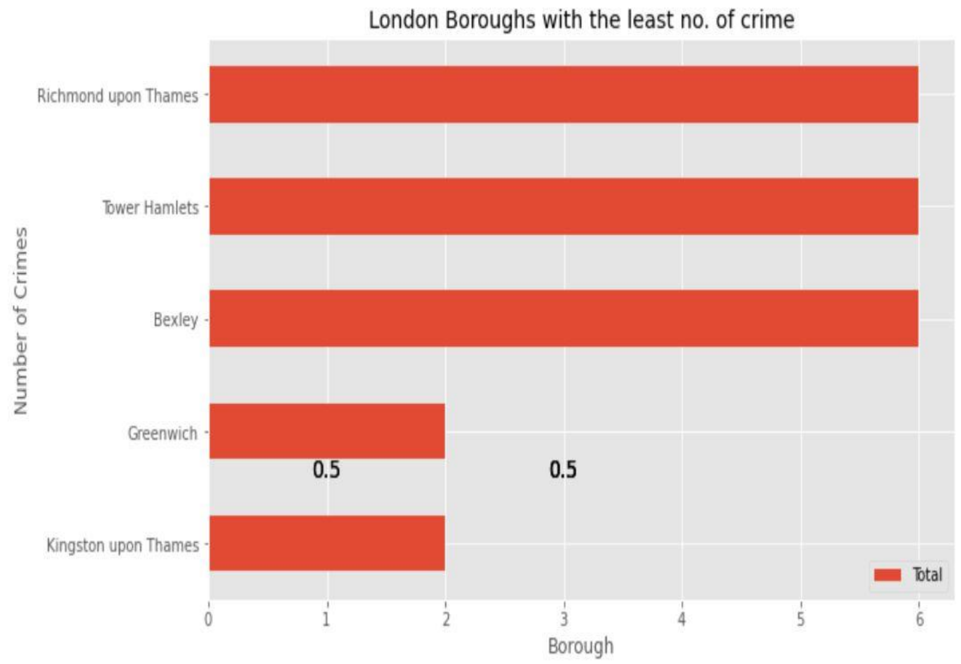
3.1.2 BOROUGHS WITH THE HIGHEST CRIME RATES

Contrasting five districts and the most elevated crime percentage during the years it is clear that Westminster has the most elevated wrongdoings recorded followed by Lambeth, Southwark, New ham and Tower Hamlets. Westminster has a fundamentally higher crime percentage than the other 4 precincts.

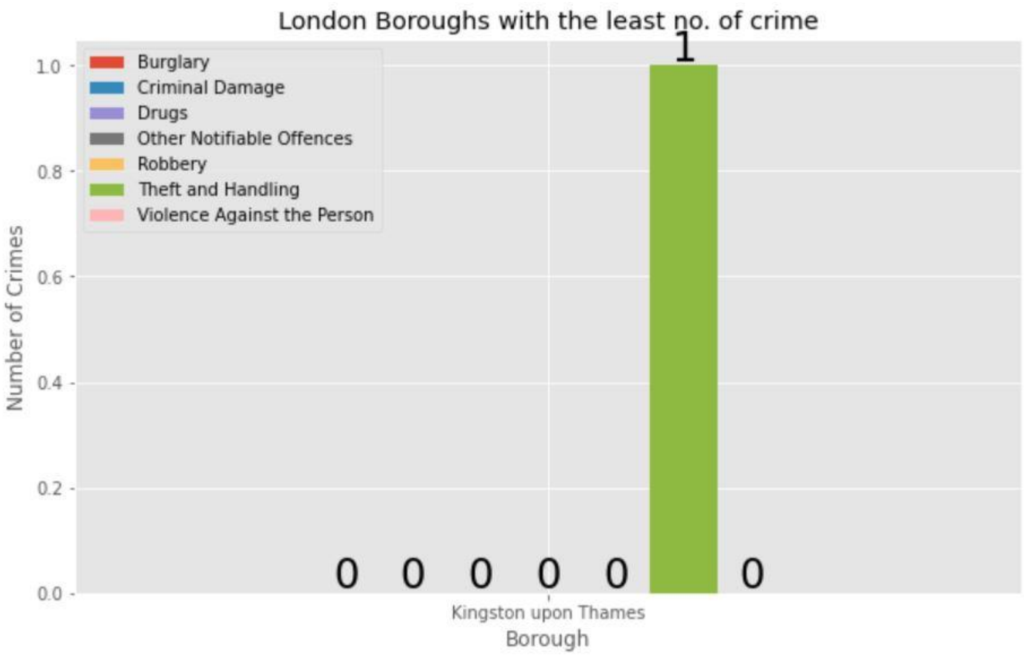


3.1.3 BOROUGHS WITH THE LOWEST CRIME RATES

Contrasting five precincts and the most reduced crime percentage during the year, City of London has the least recorded violations followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton

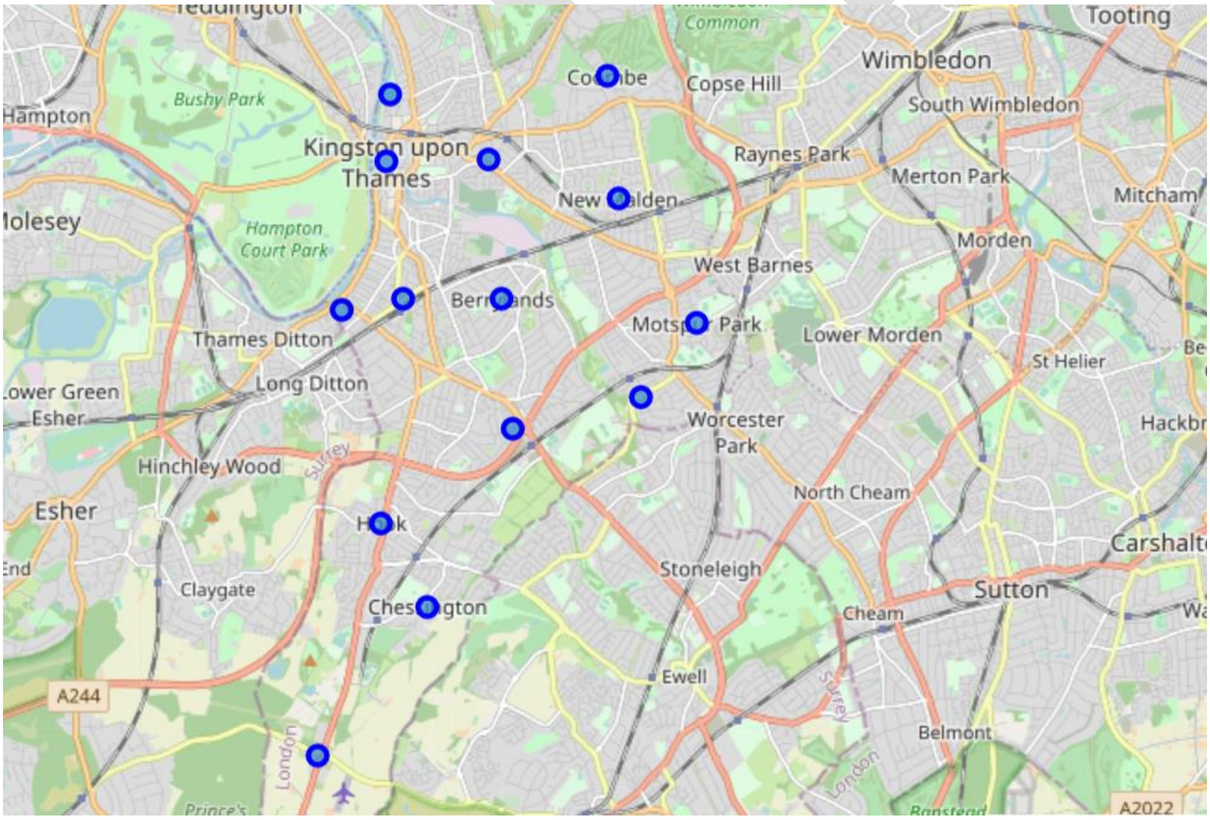


City of London has a fundamentally lower crime percentage since it is the 33rd head division of Greater London yet it's anything but a London precinct. It has a region of 1.12 square miles and a populace of 7000 starting at 2013 which proposes that it is a little. Henceforth we will consider the following district with the most reduced crime percentage as the most secure ward in London which is Kingston upon Thames.



3.1.4 NEIGHBORHOODS IN KINGSTON UPON THAMES

There are 15 neighborhoods in the illustrious ward of Kingston upon Thames, they are envisioned on a guide utilizing folium on python.



3.2 MODELLING

Utilizing the last dataset containing the areas in Kingston upon Thames alongside the scope and longitude, we can discover all the scenes inside a 500 meter sweep of each neighborhood by interfacing with the Foursquare API. This profits a json document containing all the scenes in every local which is changed over to a pandas data frame. This information outline contains all the scenes alongside their directions and classification.

One hot encoding is done on the settings information. (One hot encoding is a procedure by which straight out factors are changed over into a structure that could be given to ML calculations to make a superior showing in expectation). The Venues information is then gathered by the Neighborhood and the mean of the settings are determined, at last the 10 regular scenes are determined for every one of the areas.

To assist individuals with finding comparative neighborhoods in the most secure precinct we will group comparable neighborhoods utilizing K - implies bunching which is a type of unaided machine learning calculation that bunches information dependent on predefined group size. We will utilize a bunch size of 5 for this task will bunch the 15 neighborhoods into 5 groups.

The motivation to direct a K-implies bunching is to group neighborhoods with comparative settings together so that individuals can waitlist the zone of their inclination's dependent on the settings/pleasantries around every area.

CHAPTER 4

RESULTS

In the wake of running the K-implies grouping we can get to each bunch made to see which neighborhoods were appointed to every one of the five groups. Investigating the areas in the main bunch.

The cluster one is the greatest bunch with 9 of the 15 neighborhoods in the precinct Kingston upon Thames. Upon intently looking at these areas we can see that the most basic settings in these areas are Restaurants, Pubs, Cafe, Supermarkets, furthermore, stores.

	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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Shop & Service	Indian Restaurant	Hotel	Café	Plaza	Gym / Fitness Center	Park
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Bakery	Supermarket	Indian Restaurant	Fish & Chips Shop	Wine Shop	Fast Food Restaurant	Deli / Bodega	Department Store
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Coffee Shop	Café	Pub	Sushi Restaurant	Burger Joint	Asian Restaurant	French Restaurant	German Restaurant
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Gastropub	Office	Gym	Sushi Restaurant	Supermarket	Bar	Chinese Restaurant	Korean Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Pub	Food	Indian Restaurant	Italian Restaurant	Platform	Wine Shop	Dry Cleaner	Hardware Store
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Coffee Shop	Pub	Café	Italian Restaurant	Fast Food Restaurant	Chinese Restaurant	Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Breakfast Spot	Gastropub	Train Station

Investigating the areas in the second, third and fifth bunches, we can see these groups have just a single neighborhood in each. This is a direct result of the novel settings in each of the areas, thus they couldn't be bunched into comparative neighborhoods.

	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 M Comm Ver
2	Chessington	Kingston upon Thames	51.358336	-0.298622	0	Construction & Landscaping	Convenience Store	Deli / Bodega	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast F Restaur

The subsequent bunch has one neighborhood which comprises of Venues, for example, Restaurants,Golf courses, and wine shops.

	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	Fast Food Restaurant	Department Store	Discount Store	Dry Cleaner	Electronics Store
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	2	Gym	Park	Bus Stop	Soccer Field	Restaurant	Wine Shop	Electronics Store	Deli / Bodega	Department Store

The third bunch has one neighborhood which comprises of Venues, for example, Train stations, Restaurants, and Furniture shops.

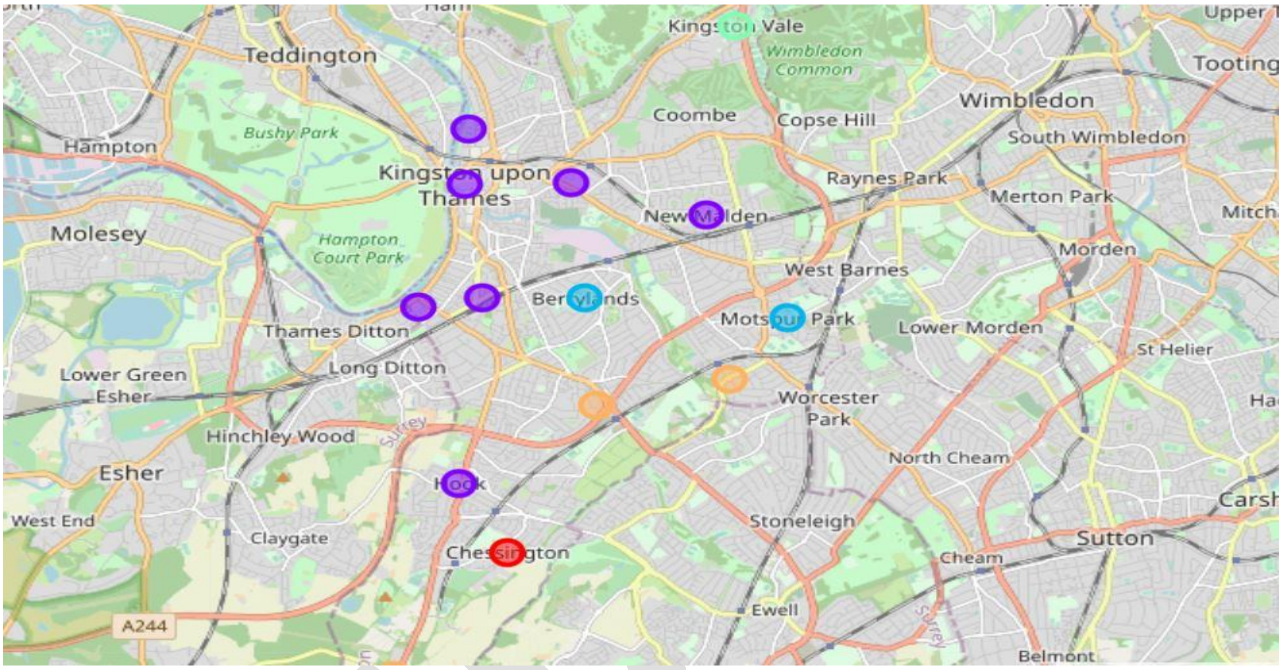
	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10 C
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	3	Grocery Store	Bar	Soccer Field	Sandwich Place	Fast Food Restaurant	Deli / Bodega	Department Store	Discount Store	Dry Cleaner	Ele

The fourth bunch has two neighborhoods in it, these areas have basic scenes, for example, Parks, Gym/Fitness focuses, Bus Stops, Restaurants, Electronics Stores and Soccer fields and so forth.

	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 Mo Comm Ven
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	4	Grocery Store	Garden Center	Pub	Restaurant	Electronics Store	Cosmetics Shop	Deli / Bodega	Department Store	Disco St
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	4	Grocery Store	Train Station	Food	Construction & Landscaping	Bagel Shop	Deli / Bodega	German Restaurant	Gastropub	Gard Cen
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	4	Grocery Store	Pharmacy	Sandwich Place	Coffee Shop	Pizza Place	Furniture / Home Store	Discount Store	Italian Restaurant	C

The fifth group has one neighborhood which comprises of Venues, for example, Grocery shops, Bars, Restaurants, Furniture shops, and Department stores. We will investigate the areas in the fourth group.

Visualizing the clustered neighborhoods on a map using the folium library



Each bunch is shading coded for the simplicity of introduction, we can see that greater part of the neighborhood falls in the red bunch which is the main group. Three neighborhoods have their own group (Blue, Purple and Yellow), these are bunches two three and five. The green group comprises of two neighborhoods which is the fourth bunch.

CHAPTER 5

DISCUSSION

The point of this venture is to help individuals who need to move to the most secure ward in London, expats can pick the areas to which they need to migrate dependent on the most basic scenes in it. For instance, if an individual is searching for an area with great availability and open transportation we can see that Clusters 3 and 4 have Train stations also, Bus stops as the most widely recognized scenes. On the off chance that an individual is searching for an area with stores and eateries in a closeness then the areas in the primary bunch is reasonable. For a family I feel that the areas in Cluster 4 are increasingly appropriate levy to the regular settings in that group, these areas have basic scenes, for example, Parks, Gym/Fitness focuses, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is perfect for a family. The selections of neighborhoods may fluctuate from individual to individual.

CHAPTER 6

CONCLUSION

This venture enables an individual to show signs of improvement comprehension of the areas as for the most widely recognized scenes in that area. It is constantly useful to utilize innovation to remain one stride ahead for example discovering progressively about spots before moving into a neighborhood. We have quite recently accepted wellbeing as an essential worry to waitlist the most secure precinct of London. The eventual fate of this venture incorporates taking different factors, for example, cost of living in the zones into thought to waitlist the precinct, for example, sifting territories based on a predefined spending plan.