Capstone Project - The Battle of Neighborhoods (Week 2)

"Segmentation and Clustering of Boroughs in London using Metropolitan Police Service (MPS) Borough Level Crime Data & Foursquare Developer API: Feasible Recommendations for Individuals and Businesses"

Date: March 12, 2020

Author: Sandeep Aswathnarayana

Data Scientist, Graduate Research Assistant

Southern Methodist University, Dallas, TX 75205, United States

1. INTRODUCTION & BUSINESS PROBLEM:

A family planning to build new home, a potential business owner or contractor looking to setup their office, an international student looking for affordable housing facilities or a traveler interested in exploring the city - all have a pivotal factor in common to be taken into consideration (irrespective of the budget/affordability) before they make a decision: 'SAFETY'.

In this project, we will study in detail the segmenting, clustering and classification of London Boroughs using Metropolitan Police Service data and Foursquare Developer API. As the city grows and develops, it becomes increasingly important to examine and understand it quantitatively. The MPS provides open data for Developers, Investors, Policy Makers, City Planners who possess an interest in answering the following questions for development and safety of residents:

- What neighborhoods have the highest crime?
- Is population density correlated to crime level?
- Using Foursquare data, what venues are most common in different locations within the city?
- Does London Datastore provide with specific enough or thick enough data to empower decisions to be made? Or is it too aggregate to provide value in its current detail? Let's find out.

2. TARGET AUDIENCE:

1. **Professionals relocating to London for work** - The number of workers in London is projected to increase by 582,000 (up 10%) in the next 10 years. This is the equivalent to about 58,200 more jobs each year. Apprenticeship starts in London more than doubled between 2009/10 and 2010/11 and since then have maintained a fairly steady level.
2. **International Students & Overseas Nationals** - There were over 5 million international visitors to London in 2019 Q2, up 2.3% from the same quarter a year before. In 2018/19, there were

over 240,000 new National Insurance Number (NINo) registrations from overseas nationals in London, which was 2% higher than the year before.

3. **Small Businesses** - With a potential opportunity to establish their footprint, small businesses benefit from the spending by local communities, travelers and visitors. Total spend by international visitors alone in London was at £12.1bn in 2018 down 10.2% on 2017. Spending in the most recent quarter (2019 Q1) was 1.8%, 6.0% higher compared with the same quarter a year previous.
4. **Policy Makers** - By grouping the neighborhoods into most similar groups, the GLA and Mayors Office for Policing and Crime (MOPAC) have enabled both the police and the public to understand performance compared to similar neighborhoods in London.
5. **Law Enforcement** - Police and partners now share best practice with like-for-like neighborhoods. The Metropolitan Police Service actively utilizes the Neighbourhood Confidence and Crime Comparator at neighborhood level to identify similar areas where public confidence metrics vary the most, and respond to the challenge of MOPAC to reduce these gaps.
6. **Families and Communities** - The population of London in 2017 was 8,904,000 up 7% from five years ago. The population is projected to increase to 9.7 million by 2025 (an increase of 17% from the 2011 Census) and reach 10 million by 2030.
7. **Housing** - In 2019/20 the total number of GLA funded affordable housing starts in London was 12,546 accounting for 90.6% of total London housing starts. In 2018/19 London’s dwelling stock saw a net increase of 35,959 dwellings compared to the year before.
8. **Voluntary Crime Prevention Movements** - Ex: Neighborhood Watch

3. DATA & DATA SOURCES:

3.1. MPS Borough Level Crime Data:

MPS Borough Level Crime Data counts the number of crimes in London at the borough-level per month, based on the crime type.

The data is available in two files for each level of geography - the most up to date data covering the last available 24 months only and one covering all historic full calendar years. To analyze the most recent patterns, I opted to explore the one with the last available 24 months.

In March 2019, the Metropolitan Police Service started to provide offences grouped by the updated Home Office crime classifications. This currently only covers the most recent 24 months of data.

Below is a list of the crime types covered under the new HO categories:

Major Category - Minor Category:

- Arson and Criminal Damage - Arson / Criminal Damage
- Burglary: Burglary - Business and Community / Burglary - Residential**
- Drug Offences: Drug Trafficking / Possession of Drugs
- Miscellaneous Crimes Against Society: Absconding from Lawful Custody / Bail Offences / Bigamy / Concealing an Infant Death Close to Birth / Dangerous Driving / Disclosure, Obstruction, False or Misleading State / Exploitation of Prostitution / Forgery or Use of Drug Prescription / Fraud or Forgery Associated with Driver Records / Going Equipped for Stealing / Handling Stolen Goods /

Making, Supplying or Possessing Articles for use i / Obscene Publications / Offender Management Act / Other Forgery / Other Notifiable Offences / Perjury / Perverting Course of Justice / Possession of False Documents / Profiting From or Concealing Proceeds of Crime / Soliciting for Prostitution / Threat or Possession With Intent to Commit Criminal / Wildlife Crime

- Possession of Weapons: Other Firearm Offences / Possession of Firearm with Intent / Possession of Firearms Offences / Possession of Other Weapon / Possession of Article with Blade or Point
- Public Order Offences: Other Offences Against the State, or Public Order / Public Fear Alarm or Distress / Racially or Religiously Aggravated Public Fear / Violent Disorder
- Robbery: Robbery of Business Property / Robbery of Personal Property
- Sexual Offences*: Other Sexual Offences / Rape
- Theft: Bicycle Theft / Other Theft / Shoplifting / Theft from Person
- Vehicle Offences: Aggravated Vehicle Taking / Interfering with a Motor Vehicle / Theft from a Motor Vehicle / Theft or Taking of a Motor Vehicle
- Violence Against the Person: Homicide / Violence with Injury / Violence without Injury

To note:

Fraud data was transferred from individual police forces to National Action Fraud in March 2013

**Prior to April 2017, police recorded burglary offence categories were split such that dwellings (domestic burglary) and buildings other than dwellings (non-domestic burglary) were separately identifiable, where:

- domestic burglary covers residential premises, including attached buildings such as garages
- non-domestic burglary covers non-residential premises, including businesses and public buildings, as well as non-attached buildings within the grounds of a dwelling, such as sheds and detached garages From April 2017 onwards a new classification of police recorded burglary was introduced, dividing offences into two categories of “residential” and “business and community”

“Residential” burglary includes all buildings or parts of buildings that are within the boundary of, or form a part of, a dwelling and includes the dwelling itself, vacant dwellings, sheds, garages, outhouses, summer houses and any other structure that meets the definition of a building. It also includes other premises used for residential purposes such as houseboats, residential care homes and hostels.

“Business and community” burglary includes all buildings or parts of buildings that are used solely and exclusively for business purposes or are otherwise entirely outside the classification of residential burglary.

3.2. List of London boroughs:
The motive behind using this dataset is to fill the gap that our actual crime dataset lacks to address i.e., extract the key attributes or columns from the List of London boroughs dataset will help us extract and analyze the attributes including Population Density for each of the 32 boroughs and their respective Co-ordinates.

3.3. Foursquare Location Data:
The Foursquare Venues & Places Database gives the full details about a venue including location, tips, and categories. We can access precise, up-to-date community-sourced venue data. Its large selection of rich

and firmographic location data unlocks the potential to enhance our app or website with the ability to describe locations, analyze trends, and improve user experience.

If the venue ID given is one that has been merged into another venue, the response will show data about the other venue instead of giving you an error. User authenticated calls will also receive information about who is here now. This is a Premium endpoint with access to venue's photos, tips, hours, menu, categories, recommendations, events, stats, etc.

Using these 3 major datasets as the basis for our project, let's start leveraging its features and attributes to address our business problem.

4. METHODOLOGY:

The brief overview of the methodology:

- Importing necessary libraries and loading the data sets of interest
- Data Cleaning, Data Wrangling, Feature Selection
- Examine the crime frequency by neighborhood
- Study the crime types and then pivot analysis of crime type frequency by boroughs
- Understand correlation between crimes and population density
- Perform K-Means Clustering Analysis on venues by locations of interest based on findings from crimes and boroughs
- Determine the venues which are in the proximity of relatively high crime count and choose the locations of interest accordingly

4.1. Getting started

- Importing the necessary libraries
- Load the dataset into the pandas data frame
- Develop Visualizations that help uncover insights about the crime frequency based on boroughs and crime type

Data Cleaning & Data Wrangling: Create a column for incident count as “Number of Incidents” on the original data frame.

Word Cloud Generation: Import the necessary libraries, read the mask image (London Borough Map, in this case), create a word cloud object ‘wc’ with “generate(str(london_crime['MinorText']))” comprising of the column of interest we are trying to visualize and explore. This column includes the “MajorText” I.e., Major Type of Crime Incidents occurring in the 32 boroughs of London in the past 24 months.

As it is evident from the word cloud, the top crime type incidents mostly include Motor Vehicle Theft, followed by Possession of Firearm, Residential Burglary, Business Burglary. These are one set of features of interest which we would be basing our further analysis on.

The “BoroughCount” column is computed and added to the data frame by sorting and grouping values based on the “Borough”. Similarly, “IncidentCount” is computed based on the “MajorText” column values.

Our current set of features or columns of interest include: IncidentCount, MajorText, Borough, BoroughCount, Number of Incidents. Using these, let us leverage the potential of pandas libraries, matplotlib, seaborn and numpy to plot some meaningful graphs that help us uncover valuable insights into the London Borough Crime dataset.

“Borough” vs. “Number of Incidents”:

Historically, Westminster has been the borough with record high crime incidents. Over the last 24 months, the number of Incidents recorded by MPS has reached 3000. And, remember this number only includes the major crime type i.e., the minor crime type incidents haven’t been computed explicitly. Following Westminster are Southwark, Newham, Camden, Lambeth which fall in the range of 1500 to 1700.

!!!!!!! Note to the Readers: In case you find it difficult to view the Visualizations attached in this report, please do visit my GitHub for the same.

URL: https://github.com/SandeepAswathnarayana/mooc/tree/master/coursera/IBM%20Data%20Science%20Professional%20Certificate/Applied%20Data%20Science%20Capstone/Final%20Capstone%20Project/visualizations !!!!!!!

“Borough” vs. “Type of Crime”:

Leaving the “Miscellaneous Crimes Against Society” crimes aside, the top 3 crimes in most of the boroughs are: Possession of Weapons, Robbery, Vehicle Offences followed by the rest.

Create a new column ”MonthlyAverage” that computes the average number of crimes per month.

4.2. List of London Boroughs:

Information on boroughs and their population & coordinates

- Population can be used to calculate the ratio of reported crime to population for better comparison
- Coordinates can be used to get neighborhood data from Foursquare
  Source: Wikipedia
  URL: https://en.wikipedia.org/wiki/List_of_London_boroughs

Web Scraping:

The values from the 'List of London boroughs' are parsed using the HTML parser, Beautiful Soup. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.

Before diving further, make sure to use Lambda functions at times when we require a nameless function for a short period of time and reduce the stress of creating large number of functions.

Now, create a data frame with columns of interest (Borough, Population, Coordinates) using the parsed tree obtained in the previous step. Strip the unwanted strings in the "Coordinates" column using rstrip() method using required regular expressions accordingly.

Split the "Coordinates" column into three different columns using split() method thereby splitting the strings into lists. The 'separator' arguments take the necessary values based on our requirement. Use the DataFrame.drop() method to drop specified labels from rows or columns.

4.3. Foursquare Developer API

List of top 50 popular venues in the neighborhood source. First, make sure you have setup your Foursquare Developer account with necessary API credentials at https://developer.foursquare.com/

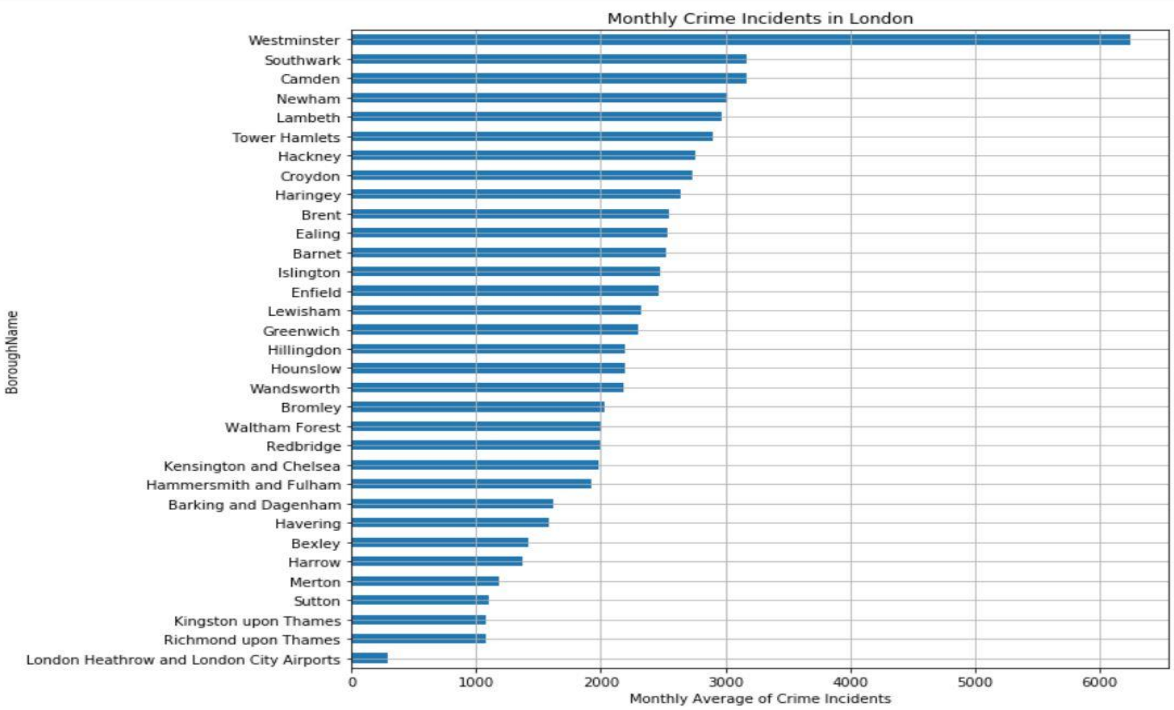
Make sure you have the CLIENT_ID, CLIENT_SECRET and API Version by printing them. It is highly recommended to keep these credentials confidential to avoid fraudulent use your Foursquare account, if accessed by others.

Now, let's create a function to explore the boroughs based on the popular venues that returns relevant information for each nearby venue. Next, GET the top 50 venues in 500m radius of the center of each borough.

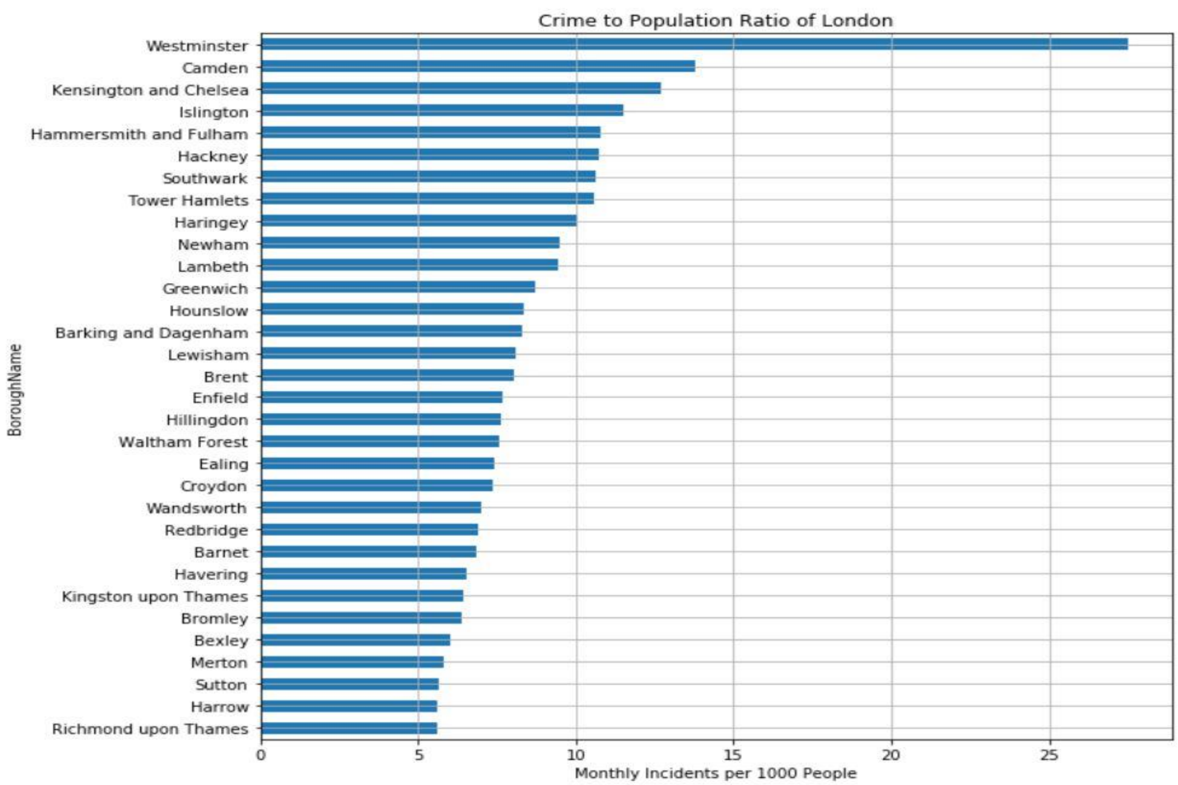
4.4. Exploratory Data Analysis

In this section, let's clean the data and explore the same more thoroughly. Then, conduct cluster analysis to classify the boroughs into different levels of preference by ranking them.

"Borough" vs. "Monthly Average of Crime Incidents":



“Borough” vs. “Monthly incidents per thousand people”:



From the above bar plots, it is evident that Westminster has the highest number of reported crime incidents, followed by Southwark, Camden, Newham.

Merge the borough crime and borough information data into a single data frame. Due to its original nature, the three columns “Population”, “Latitude” and “Longitude” have object as their default datatype. So, convert them into int64, float64 and float64 respectively.

Import geopy.geocoders, which is a python client that makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources. Using this, get the geographical coordinates of London. Now, create a map of London using Folium, a Python library that helps us create several types of Leaflet and Choropleth maps. For Folium styling, markers and more, please visit https://python-visualization.github.io/folium/

Now, we have successfully created a folium map with crime incidents covering all the 32 boroughs of London.

4.5. Cluster Analysis of Boroughs using K-Means

In this part, we will conduct K-Means Clustering to group the boroughs based on the recreational, convenience facilities they possess as provided by the Foursquare data.

One-hot Encoding:

Before this, make sure you retrieve the number of unique “Venue Category” column using unique() method.

K-Means, like most of the ML algorithms, cannot work with categorical data. They require all input variables and output variables to be numeric. This can be attained either by Integer Encoding or One-hot Encoding. For categorical variables where no ordinal relationship exists (just like the ones in our case), the integer encoding is not enough. So, let’s proceed with One-hot encoding using binary vector or “dummies” to represent each integer value.

After the completion of one-hot encoding, the “BoroughName” column is added back to the data frame. Now group rows by boroughs and compute the mean of frequency of each venue category.

Let’s get the Top 5 venues for each neighborhood. Let’s put these values into a new pandas data frame with columns created based on the number of top venues.

K-Means Clustering:

Perform K-Means Clustering on the new data frame that we obtained in the previous step. Make sure you set the number of clusters, add the “Cluster Labels” column. Let’s merge this dataset with “london_merged” data frame to add Latitude, Longitude values for each of the boroughs.

Now that we have built clusters, let’s reproduce these clusters onto a Folium map using markers and color scheme to make sure we distinguish the 5 cluster groups by its colors.

4.6. Segmenting & Classification of Top 5 Clusters

Now, let's observe each cluster and name them based on their characteristics accordingly. Cluster 0, Cluster 1, Cluster 2, Cluster 3, and Cluster 4.

Now, let's review all the analysis and key observations before coming up with recommendations and conclusion on narrowing down our research to provide safer boroughs. As mentioned in the earlier sections, our key criteria of location decision will be based on safety and atmosphere.

5. RESULTS

5.1 Safety and Safety Score:

Safety can be determined by the crime rate we have calculated. We can use the 'CrimeToPop' (Recorded crime per 1000 people) as our safety score. Let's sort the values in descending order in the london_merged data frame and display them to create a Safety Score scale.

The Safety Score is calculated by Z-Score Normalization of "CrimeToPop" values. These final "Score" values are sorted in descending order to obtain the most preferred neighborhoods at the top of the data frame.

5.2. Atmosphere & Cluster Classification

Based on the cluster analysis results, let us give each cluster a name depending on the characteristics inferable from the popular venues around.

The "Atmosphere" column is initially set to zeros. Later, they are computed based on the characteristics, vide or atmosphere of the 5 Clusters we obtained using K-Means.

Finally, the "Score" column is calculated by simply adding the values in "Safety" and "Atmosphere" columns.

Analysis of Cluster 0:

- Observation: Cluster 0 has pool, coffee shop, clothing store as popular places.
- Inference: A neighborhood where consumer taste and personal income is a major driving force.
- Classification: Cluster 0 can be categorized as a "**Neighborhood that drives demand"**

Analysis of Cluster 1:

- Observation: Cluster 1 has yoga studios, café, cupcake shops as popular venues. can be categorized
- Inference: A thriving yoga community includes: Gorgeous studios, world-class yoga instruction, beautiful scenery and proximity to nature. A culture of healthful living. Like-minded people who value service, giving back, and balance. And above all, a sense that your community and surroundings are a source of support and inspiration.
- Classification: **"Major Metropolises who want to live a healthier life", "Most vegan-friendly neighborhoods"**

Analysis of Cluster 2:

- Observation: Cluster 2 has hotel, airport service, airport as popular places.
- Inference: A neighborhood where consumer commute and accommodation are the driving force. Both these places mostly attract affluent professionals, active middle-aged adults.
- Classification: **"Traveler or Tourist Destination", "Active/Resort/Luxury Neighborhoods"**

Analysis of Cluster 3:

- Observation: Cluster 3 has Indie movie theatre, Indian restaurant, convenience store, coffee shop as popular places.

- Inference: A neighborhood where consumer commute and accommodation are the driving force. If you're going the indie route, you need to know what the local audience wants to see, particularly if the established movie theaters aren't providing it. Would theater-goers like to watch more foreign films? Art films? Big-screen showings of old classics? Even if you have a personal vision about which movies you want to show, mixing in some sure-fire audience hits will be essential to your bottom line.
- Classification: **“Entertainment District”, “Neighborhood Movie House”**

Analysis of Cluster 4:

- Observation: Cluster 4 has Indian restaurant, airport lounge, café, fast food as popular places.
- Inference: These are located in the neighborhoods which are protected areas with separation from street traffic and high visibility; serving local neighborhoods and adjoining schools, libraries, or police and fire facilities.
- Classification: **“Recreational Neighborhood”, “Schools/Out-of-yard/Family Neighborhood”**

5.3. Inference based on Safety Scores

- From our analysis, we have found that the five boroughs below are the best places to move in, based on safety and atmosphere of the neighborhood. The top five boroughs all mostly belong to the Cluster 1 (**“Major Metropolises who want to live a healthier life”, “Most vegan-friendly neighborhoods”**),Cluster 2 (**“Traveler or Tourist Destination”, “Active/Resort/Luxury Neighborhoods”**) or Cluster 3 (**“Entertainment District”, “Neighborhood Movie House”**) which is busy for the most part of a day. Therefore, what differentiates them apart is the Safety Score, which was calculated from monthly recorded crimes per 1000 people. Finally, these clusters are plotted using the Folium Map.

5.4. Key Insights & Takeaways

- In May 2019 the total number of offences in London was 75717 - up 6.1% from the same month last year and up 1.2% from the previous month.
- Over the last 12 months, 25 of London's 32 Boroughs have seen an increase in the number of crimes committed compared with the previous 12 months. The biggest increase was seen in Westminster (up 23.3%) while the largest fall in crime was seen in Islington (down 6.8%).
- The number of crimes on the transport network in the quarter Jul-Sept 2019/20 (9,043) was up 15.2 per cent compared with the same quarter in 2018/19 (7,853).

Let's review the goals of this project.

The idea for the Capstone Project is to show that when driven by venue and location data from Foursquare, backed up with open source crime data, that it is possible to present the cautious and nervous traveler with a list of attractions to visit supplemented with a graphics showing the occurrence of crime in the region of the venue.

A general, broader perspective of this project:

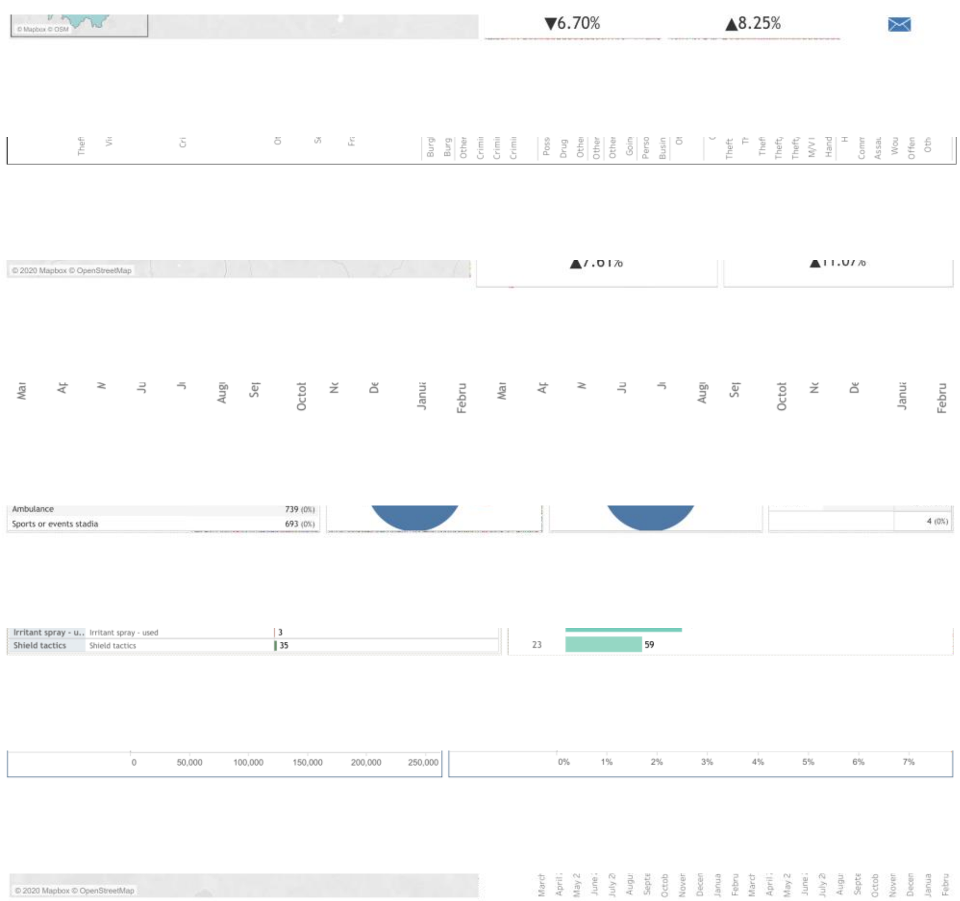
- Individuals or Businesses decide on a City of intent to move in or set up their offices (London, in this case)
- The Foursquare API is scrapped for the top venues in the city

- Based on the top venue locations, the list is augmented with additional geographical data
- Based on the additional geographical data the nearby, relatively safer locations of interest are chosen
- The historical crime within a predetermined distance of all venues are obtained
- A map is presented to the to the individual, potential business owner or traveler with the selected venues and crime statistics of the area

5.5. Tableau Dashboards & Visualizations

I developed some cool visualizations with the help of some additional data available on London Datastore (https://data.london.gov.uk/) using Tableau (https://www.tableau.com/). The images of the dashboards are included in this section which helps us uncover some great insights dwelling on the new key features which we never got to see in our original borough crime dataset.

The dashboards are self-explanatory and the numbers speak for themselves:



6. DISCUSSION:

Observations:

- The London Datastore - Greater London Authority enables us to gain an understanding of the crime volume by type by area but not specific enough to understand the distribution properties.
- Valuable questions such as, "are these crimes occurring more often in a specific area and at a certain time by a specific demographic of people?" cannot be answered nor explored due to what is reasonably assumed to be personal and private information with associated legal risks.
- There is value to the city to explore the detailed crime data using data science to predict frequency, location, timing and conditions to best allocated resources for the benefit of its citizens and its police force. However, human behavior is complex requiring thick profile data by individual and the conditions surrounding the event(s).

Programs by The Mayor's Office for Policing And Crime (MOPAC):

- The Mayor investing in services that would substantially help reduce crime by supporting young people at risk of or engaged in serious violence and help them into employment, education or training: Young Londoners Fund, The London Gang Exit (LGE) Service, Stepping Stones Programme, London Needs You Alive, Safer School Officers

Recommendations to tackle and prevent the top 3 high frequency crime types:

Recommendations to combat Residential Burglary:

- Property Marking: Make your valuables less attractive to thieves by using property or police approved forensic marking solution only visible under Ultraviolet (UV) light. This can take the form of a special marker using ink or paint. Mark or etch your property with your postcode, house or flat number or the first three letters of your address. All of these measures help police identify stolen goods, making it harder for thieves to sell
- Register your valuables using an accredited property database
- MetTrace is our project to reduce burglary across London through forensic property marking
- If you opt for a safe in your home, make sure it is fire rated. Home safes are also insurance rated according to the type and value of items designed to be kept inside, so check with your insurance company first to make sure you're adequately covered
- Install a visual burglar alarm, shed or garage burglar alarm, adequate outdoor lighting
- Leave radios and lights on a timer
- Join or form a Neighbourhood Watch scheme

Recommendations to combat Burglary at Businesses:

- Businesses need to make sure they have a monitored alarm and that it's fully operational as approved by the NSI and SSAIB
- Other Safety Devices: Anti-ram raider bollards, bolted-down safe with a time lock and anti-tamper sensors, smoke-generating devices that activate on unauthorized entry, etc.

Recommendations to combat Motor Vehicle Theft:

- Park Mark, a Safer Parking Scheme signage, helps drivers find car parks where they can confidently leave their vehicle, knowing the environment is safer

- Watch for illegal, suspicious tow trucks. Thatcham Security category 1 or 2 alarm system with tracking, immobilization, anti-grab and movement sensors can help protect and trace your vehicle
- Before you buy a used vehicle, check for cloning: When buying a vehicle, always check the DVLA V5 document and make sure the Vehicle Identification Number (VIN) on the vehicle is the same as on the document
- Secure your port: Many modern vehicles are fitted with engine management diagnostic ports, which can be accessed without the thief needing to open the vehicle doors, boot or bonnet, but which can unlock and start your vehicle
- Secure bikes by locking them to an immovable object within a locked shed or garage

For Commuters & Visitors:

- Be assertive: From the moment you step out onto the street in the morning, look assertive and act and walk with confidence. This will always make you appear in control and much less vulnerable
- Always try to avoid sitting in an empty carriage of a double-decker night bus, train where you are more vulnerable

7. CONCLUSION:

Although all of the goals of this project were met there is definitely room for further improvement and development as noted below. With some more work, could easily be developed into a fully-fledged application using Plot.ly (https://plot.ly/) interactive, Time Series Visualizations and host it on Heroku (https://www.heroku.com/). that could support the cautious traveler in an unknown location.

A better Police Service for London
Wherever you live in the capital, and whatever your background, you should expect the same service from the Metropolitan Police Service (MPS).

Providing the best service to all Londoners is at the heart of the Police and Crime Plan and means getting the important things right: making communities safer, responding to and preventing crime, building trust and confidence, and bringing criminals to justice.

What MPS is doing?
To make the police service in London better we are:

- Making record investment from City Hall into the Metropolitan Police Service.
- Funding the recruitment of more police officers for London - bringing numbers back over 30,000.
- Providing the latest equipment and technology for officers, such as body-worn video cameras and mobile data tablets.
- Investing in a new counter-terror hub for London.
- Overseeing the work of the Metropolitan Police on behalf of Londoners, scrutinizing key issues such as the MPS Gangs Matrix and the use of Facial Recognition technology.

Get involved
Find out about community projects and schemes you can get involved with in your area:

Safer Neighbourhood Boards
Independent Custody Visitors

More information
Counter-terrorism
The London Crime Prevention Fund

Limitations:

Trade-off:

- One has to put up with no public transportation for commute in order to have a much safer neighborhood where one resides. As a matter of fact, proximity to public transportation make the neighborhoods more prone to crimes. As a result of this, most of the neighborhoods vote themselves out of the access to public transportation.
- Facts to back this up based on my experiences, anecdotes, and personal interaction with the City of Dallas: I lived in Dallas/Fort Worth Area for 4 years during which I learned that some cities including Allen, Irving, McKinney, etc. in the DFW Area voted themselves out of DART Bus/Train facilities (public transportation) to make sure their neighborhoods are less prone to crimes as opposed to the ones that voted for public transportation.

However, it obviously seems strange that the best places to live in London are all far out suburbs. This is due to limitations this research holds. Among the numerous factors that determine a good neighborhood, I narrowed down my research considering only the popular venues and crime incidents provided the time and scope of the project. This means that serious crimes like homicide was treated the same as a comparatively petty crime like shoplifting. Moreover, the number of stores in the neighborhood may be as important as what type of stores there are.

To overcome the limitations of this study, we will need further data such as housing prices, proximity to the most basic services such as hospitals, grocery stores.

References:

MPS Borough Level Crime Data

List of London boroughs

Foursquare Venues & Places Database

The Mayor's Office for Policing And Crime (MOPAC)

Alex Aklson: https://www.coursera.org/instructor/alexaklson, Data Scientist at IBM

Applied Data Science Capstone: https://www.coursera.org/learn/applied-data-science-capstone