Capstone Project Report: The Battle of Neighbourhoods

London's Crime Rate Analysis and Clustering of the Safest Neighbourhoods of London

Vincent Stanislaus Pereira

August 2021

Table of Contents

1. Introduction [Index]

London is one of the most multicultural cities in the world. It is a melting pot of cultures, where one can taste the best of the world cuisine. It is a major centre for banking and finance, insurance, world trade, media, advertising, tourism, theatre, fashion, arts and more. Fusing gritty, historic pomp with shimmering modernity, world-class culture and fashion-forward shopping, the UK's capital has it all and there's something for everyone. The vibrancy of the city extends across all 32 of its boroughs, all of which are home to a plethora of unique neighbourhoods.

2. Business Problem [Index]

The decision to move to a new a city or a new country altogether, is a harrowing one. But after having decided to move to London, the next challenge one faces is to decide where to live in London. If one looks at the map of London, they will find a haphazard cluster of neighbourhoods and villages, each with their own distinct features and identity. Some of London's best neighbourhoods are usually established on the typical tourist trail, while others are constantly evolving, taking turns to emerge as the new cool hotspot. The following questions then arise in our mind,

- Which neighbourhood is right for us?
- Which part of the city has the best parks and playgrounds?
- Which schools fall in the neighbourhood?
- What area has the best craft beer scene or all-night eateries?
- Where can one find the hippest bookstores or outdoor yoga?

And at the top of all these doubts, the most intriguing questions anyone would face are,

- What is the crime rate in the area?
- Is it a secure neighbourhood?
- Is it safe to venture out in the night?

All these questions and more plague our mind and then the quest to find the answers begins.

3. Objective of the Capstone Project [Index]

The objective of this assignment is to give an insight into what some of the safest London neighbourhoods can offer its residents and tourists.

To help uncover the best that London has to offer, this project aims to do the following,

- Identify the safest boroughs and wards in London based on the latest crime data
- Find the Latitude & the Longitude coordinates of the preferred neighbourhoods by using their Postcodes
- Plot the safest neighbourhoods on the Map of London using the geographical coordinates obtained
- Locate the most common venues in the vicinity of 500 metres from these neighbourhoods
- Cluster these neighbourhoods based on the common venues using a Machine Learning algorithm (K-Means Clustering)

4. Interested Parties [Index]

The objective of this project is to identify and recommend the best & safest neighbourhoods in London to anyone who wants to visit or relocate to London. The interested parties could be anyone from the below mentioned list,

- Young couples
- Families with children
- Executives
- Tourists, etc.

5. Description of Data [Index]

1. MPS Ward Level Crime Data for London

- This dataset has been extracted from the Metropolitan Police Service's "Recorded Crime: Geographic Breakdown" Data available on the London Datastore, https://data.london.gov.uk/dataset/recorded_crime_summary
- This data provides the number of crimes recorded per month according to crime type at the geographic level of London's Wards for the period July 2019 to June 2021

2. List of London Boroughs

- This dataset has been extracted from the Wikipedia.org page: https://en.wikipedia.org/wiki/List_of_London_boroughs
- It has been used to fetch more information on the different Boroughs of London, like the local authority of the borough, the political party controlling the local authority, the Head Quarters of the local authority, the area of the Borough, its population, its coordinates, and its designated number on the map of London
- With this information we can get more insight in to the various Boroughs of London

3. London Postcodes

- This dataset has been extracted from Doogal.co.uk: https://www.doogal.co.uk/london_postcodes.php

- The dataset has a complete list of London postcode districts
- Even though this dataset already had the Latitude and the Longitude data available, I have used the ArcGIS API to re-fetch the coordinates of the preferred locations

4. ArcGIS API Data

- ArcGIS is an online API that enables us to connect people, locations, and data using interactive maps
- We use the ArcGIS API to get the geographical coordinates (Latitude and Longitude) of the neighbourhoods of London by providing the Postcodes of the desired locations
- The following information is obtained for each Postcode,
    - **Latitude:** Latitude of the Postcode
    - **Longitude:** Longitude of the Postcode

5. Foursquare API Data

- Foursquare is a location data provider with information about different venues and events within an area of interest
- The information obtained from the Foursquare API includes venue names, locations, menus, reviews, photos, etc.
- The Foursquare location platform is, thus, used by us as a data source since all the required information about the different venues in various neighbourhoods of the desired Borough or Ward can be obtained through their API

6. Methodology [Index]

a. Importing Libraries [Index]

- Libraries used in this Project are,
  - **Pandas:** For creating and manipulating dataframes
  - **Numpy:** For scientific computation
  - **JSON:** To handle JSON files
  - **Requests:** To handle http requests
  - **Matplotlib:** It is a data visualisation and graphical plotting library
  - **Plotly:** It is also a visualisation library for creating interactive and publication-quality charts / graphs
  - **Folium:** It is used for visualising geospatial data and plotting interactive maps
  - **Geocoder:** To retrieve Location Data
  - **Scikit Learn:** To use K-Means Clustering, a Machine Learning Algorithm

b. Extracting, Scraping, Exploring, Cleaning and Processing the Datasets [Index]

- After importing all the required libraries, we will extract the data from different sources and clean it so that it is ready for processing and analysing

Dataset 1: Metropolitan Police Service Ward Level Crime Data for London [Index]

- The extracted data is the most recent data available updated till June 2021
- This data counts the number of crimes per month according to crime type at the geographic level of London's Wards for the period July 2019 to June 2021
- In March 2019, the Metropolitan Police Service started to provide offences grouped as per the updated Home Office crime classifications
- Below is a list of the crime types covered under the new Home Office categories:

No.	Major Category (Labeled as "Crime Head")	Minor Category (Labeled as "Crime Sub-Head")
1.	• Arson and Criminal Damage	• Arson • Criminal Damage
2.	• Burglary	• Burglary - Business and Community • Burglary - Residential
3.	• Drug Offences	• Drug Trafficking • Possession of Drugs
4.	• 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 • 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 Crime • Wildlife Crime
5.	• 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
6.	• Public Order Offences	• Other Offences Against the State, or Public Order • Public Fear Alarm or Distress • Racially or Religiously Aggravated Public Fear • Violent Disorder

7.	• Robbery	• Robbery of Business Property • Robbery of Personal Property
8.	• Sexual Offences	• Rape • Other Sexual Offences
9.	• Theft	• Bicycle Theft • Other Theft • Shoplifting • Theft from Person
10.	• Vehicle Offences	• Aggravated Vehicle Taking • Interfering with a Motor Vehicle • Theft from a Motor Vehicle • Theft or Taking of a Motor Vehicle
11.	• Violence Against the Person	• Homicide • Violence with Injury • Violence without Injury

- Before cleaning the data, the dataset contained 29 columns
- Post cleaning and processing the data, 5 columns have been renamed and 1 has been added
- The dataset now contains the following 30 columns:
  - **Ward Code:** Code of the Ward in the London Borough
  - **Ward:** Name of the Ward in the London Borough
  - **Borough:** Name of the London Borough
  - **Crime Head:** High level categorisation of crime
  - **Crime Sub-Head:** Low level categorisation of crime within Crime Head

- **201907 … 202106:** 24 separate columns for the Year and Month of the Reported Crime starting from 201907 to 202106. These columns show the number of reported crimes in the month for a particular Ward.
- **Total:** Total Crimes Reported for a particular Ward from July 2019 to June 2021

- While the exploring the dataset, it was found that there were two Wards by the name of "Belmont" in Harrow as well as in Sutton. Hence, in order to segregate them so as not to cause any confusion during analysis, their names were changed to "Belmont Harrow" and "Belmont Sutton".
- Further, in order to maintain consistency, the names of these two Wards were also changed in the fourth dataset, which had the London Postcodes
- The original dataset contained a total of 22,403 records
- Once the dataset was processed to include only the Top 5 safest Boroughs of London, the number of records reduced to 3,007 from 22,403 records
- After the dataset was processed further, to include only the Top 50 safest Wards of London, the number of records reduced to 1,549 from 3,007 records

In [202]:
```
crime_df.head()
```
Out[202]:

	Ward Code	Ward	Borough	Crime Head	Crime Sub-Head	201907	201908	201909	201910	201911	...	202010	202011	202012	202101	202102	202103	20210
0	E05000026	Abbey	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	2	2	...	0	0	0	1	0	0	[illegible]
1	E05000026	Abbey	Barking and Dagenham	Arson and Criminal Damage	Criminal Damage	11	13	10	12	9	...	15	17	12	14	4	10	[illegible]
2	E05000026	Abbey	Barking and Dagenham	Burglary	Burglary - Business and Community	7	10	2	4	10	...	2	1	6	3	1	3	[illegible]
3	E05000026	Abbey	Barking and Dagenham	Burglary	Burglary - Residential	0	5	6	7	13	...	4	7	9	3	4	3	[illegible]
4	E05000026	Abbey	Barking and Dagenham	Drug Offences	Drug Trafficking	3	0	0	1	1	...	0	0	0	0	2	1	[illegible]

5 rows × 30 columns

```
In [15]: crime_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22403 entries, 0 to 22402
Data columns (total 30 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   Ward Code       22403 non-null  object
 1   Ward            22403 non-null  object
 2   Borough         22403 non-null  object
 3   Crime Head      22403 non-null  object
 4   Crime Sub-Head  22403 non-null  object
 5   201907          22403 non-null  int64 
 6   201908          22403 non-null  int64 
 7   201909          22403 non-null  int64 
 8   201910          22403 non-null  int64 
 9   201911          22403 non-null  int64 
 10  201912          22403 non-null  int64 
 11  202001          22403 non-null  int64 
 12  202002          22403 non-null  int64 
 13  202003          22403 non-null  int64 
 14  202004          22403 non-null  int64 
 15  202005          22403 non-null  int64 
 16  202006          22403 non-null  int64 
 17  202007          22403 non-null  int64 
 18  202008          22403 non-null  int64 
 19  202009          22403 non-null  int64 
 20  202010          22403 non-null  int64 
 21  202011          22403 non-null  int64 
 22  202012          22403 non-null  int64 
 23  202101          22403 non-null  int64 
 24  202102          22403 non-null  int64 
 25  202103          22403 non-null  int64 
 26  202104          22403 non-null  int64 
 27  202105          22403 non-null  int64 
 28  202106          22403 non-null  int64 
 29  Total           22403 non-null  int64 
dtypes: int64(25), object(5)
memory usage: 5.1+ MB
```

Dataset 2: List of London Boroughs [Index]

- The dataset, “List of London Boroughs”, has been extracted from Wikipedia.org
- It has been used to fetch more information on the different Boroughs of London, like the local authority of the borough, the political party controlling the local authority, the Head Quarters of the local authority, the area of the

Borough, its population, its coordinates, and its designated number on the map of London

- With this information we can get more insight in to the various Boroughs of London
- Post cleaning and processing the data, 2 columns have been dropped and 5 columns have been renamed
- The dataset now contains the following 8 columns:
    - **Borough:** Name of the London Borough
    - **Local Authority:** Name of the Local Authority
    - **Political Control:** Name of the Political Party controlling the Local Authority
    - **Head Quarters:** Address of the Local Authority
    - **Area (sq mi):** Area of the Borough in square miles
    - **Population (2019 estimate):** 2019 estimate of the Population of the Borough
    - **Co-ordinates:** Latitude and Longitude of the Borough
    - **Borough No. on Map:** Designated Number of the Borough on the Map of London
- The dataset contains a total of 32 records, which is the total number of London Boroughs, excluding the City of London

In [26]:
```
london_bor_df = london_bor_df.drop(["Inner", "Status"], axis = 1)
london_bor_df
```

Out[26]:

	Borough	Local Authority	Political Control	Head Quarters	Area (sq mi)	Population (2013 estimate)	Co-ordinates	Borough No. on Map
0	Barking and Dagenham	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	212906	.mw-parser-output .geo-default,.mw-parser-outp...	25
1	Barnet	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	395896	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W	31
2	Bexley	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	248287	51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E	23
3	Brent	Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	329771	51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W	12
4	Bromley	Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	332336	51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E	20
5	Camden	Camden London Borough Council	Labour	Camden Town Hall, Judd Street	8.40	270029	51°31′44″N 0°07′32″W / 51.5290°N 0.1255°W	11
6	Croydon	Croydon London Borough Council	Labour	Bernard Weatherill House, Mint Walk	33.41	386710	51°22′17″N 0°05′52″W / 51.3714°N 0.0977°W	19
7	Ealing	Ealing London Borough Council	Labour	Perceval House, 14-16 Uxbridge Road	21.44	341806	51°30′47″N 0°18′32″W / 51.5130°N 0.3089°W	13
8	Enfield	Enfield London Borough Council	Labour	Civic Centre, Silver Street	31.74	333794	51°39′14″N 0°04′48″W / 51.6538°N 0.0799°W	30
9	Greenwich	Greenwich London Borough Council	Labour	Woolwich Town Hall, Wellington Street	18.28	287942	51°29′21″N 0°03′53″E / 51.4892°N 0.0648°E	22

In [221]:
```
london_bor_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
Data columns (total 8 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   Borough                     32 non-null     object 
 1   Local Authority             32 non-null     object 
 2   Political Control           32 non-null     object 
 3   Head Quarters               32 non-null     object 
 4   Area (sq mi)                32 non-null     float64
 5   Population (2013 estimate)  32 non-null     int64  
 6   Co-ordinates                32 non-null     object 
 7   Borough No. on Map          32 non-null     int64  
dtypes: float64(1), int64(2), object(5)
memory usage: 2.1+ KB
```

Dataset 3: <u>Merged Dataset of Dataset 1 and Dataset 2</u> [Index]

- The third dataset has been created by merging the first two datasets, i.e., by merging the datasets, "MPS Ward Level Crime (most recent 24 months)" and "List of London Boroughs"
- The two datasets have been merged on the common column present in both the datasets, i.e., the "Borough" column

- After merging and reindexing the columns, the dataset contains the following 37 columns:
    - **Ward Code:** Code of the Ward in the London Borough
    - **Ward:** Name of the Ward in the London Borough
    - **Borough:** Name of the London Borough
    - **Local Authority:** Name of the Local Authority
    - **Political Control:** Name of the Political Party controlling the Local Authority
    - **Head Quarters:** Address of the Local Authority
    - **Area (sq mi):** Area of the Borough in square miles
    - **Population (2019 estimate):** 2019 estimate of the Population of the Borough
    - **Co-ordinates:** Latitude and Longitude of the Borough
    - **Borough No. on Map:** Designated Number of the Borough on the Map of London
    - **Crime Head:** High level categorisation of crime
    - **Crime Sub-Head:** Low level categorisation of crime within Crime Head
    - **201907 … 202106:** 24 separate columns for the Year and Month of the Reported Crime starting from 201907 to 202106. These columns show the number of reported crimes in the month for a particular Ward.
    - **Total:** Total Crimes Reported for a particular Ward for the period July 2019 to June 2021
- The dataset contains a total of 22,403 records

In [28]:

```
london_crime_df = london_crime_df.reindex(["Ward Code", "Ward", "Borough", "Local Authority", "Political Control", "Head Qua
london_crime_df
```

Out[28]:

	Ward Code	Ward	Borough	Local Authority	Political Control	Head Quarters	Area (sq mi)	Population (2013 estimate)	Co-ordinates	Borough No. on Map	...	202010	202011	202012	2021
0	E05000026	Abbey	Barking and Dagenham	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	212906	.mw-parser-output .geo-default,.mw-parser-outp...	25	...	0	0	0	
1	E05000026	Abbey	Barking and Dagenham	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	212906	.mw-parser-output .geo-default,.mw-parser-outp...	25	...	15	17	12	
2	E05000026	Abbey	Barking and Dagenham	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	212906	.mw-parser-output .geo-default,.mw-parser-outp...	25	...	2	1	6	

- The merged dataset provides more information on the different Boroughs of London, like the local authority of the borough, the political party controlling the local authority, the address of the local authority, the area of the Borough, its population, its coordinates, and its designated number on the map of London
- Thus, with this information we can get more insight in to the various Boroughs of London

Dataset 4: London Postcodes [Index]

- As the name suggests, this dataset has been used to fetch the Postcodes of the different neighbourhoods in London
- This dataset has been created to find Venues in the Neighbourhood of London using the Foursquare API
- It has been extracted from Doogal.co.uk and has a complete list of London postcode districts
- The original dataset had 49 columns, but after the cleaning process, 28 columns were dropped as the same were not required for analysis

- Post cleaning and processing the data, the dataset contains the following 21 columns:
  - **Postcode Data**
  - **Latitude Data**
  - **Longitude Data**
  - **Nearest Station**
  - **Distance to Station**
  - **Ward Code**
  - **Ward**
  - **District Code**
  - **District**
  - **Constituency Code**
  - **Constituency**
  - **LSOA Code**
  - **Lower Layer Super Output Area**
  - **MSOA Code**
  - **Middle Layer Super Output Area**
  - **London Zone**
  - **Postcode Area**
  - **Postcode District**
  - **Easting**
  - **Northing**
  - **Grid Ref**
- While the exploring the 1st Dataset, i.e., "London Crime", it was found that there were two Wards by the name of "Belmont" in Harrow as well as in

Sutton. Hence, in order to segregate them so as not to cause any confusion during analysis, their names were changed to "Belmont Harrow" and "Belmont Sutton".

- In order to maintain consistency, the names of these two Wards were also changed in this dataset
- Even though this dataset already had the Latitude and the Longitude data available, I have used the ArcGIS API to re-fetch the coordinates of the preferred locations
- Before cleaning the data, the dataset contained a total of 3,24,634 records
- The number of records were reduced to 1,79,704 from 3,24,634 after removing the Postcodes that were not in use
- Once the dataset was processed to include only the Top 5 safest Boroughs of London, the number of records reduced to 20,249 from 1,79,704
- After the dataset was processed further, to include only the Top 50 safest Wards of London, the number of records reduced to 10,083 from 20,249
- Now, we could have used these 10,083 Postcodes of the Top 50 Wards of London to find their coordinates, but the process of fetching the coordinates for so many postcodes would have taken a lot of time. Hence, it was necessary to reduce the number of records further.
- Therefore, in order to reduce the dataset further, I selected the location that was nearest to the Station
- After processing the dataset, it was found that there was a total of 81 locations in the Top 50 safest Wards of London that were nearest to the Stations
- Thus, the number of Postcodes were reduced from 10,083 to 81

In [217]:

```
1 london_postcodes_df.head()
```

Out[217]:

	Postcode Data	Latitude Data	Longitude Data	Nearest Station	Distance To Station	Ward Code	Ward	Borough Code	Borough	Constituency Code	...	LSOA Code	Lower Layer Super Output Area	MSOA Code	Middle Layer Super Output Area
0	BR1 1AA	51.401546	0.015415	Bromley South	0.218254	E05000109	Bromley Town	E09000006	Bromley	E14000604	...	E01000675	Bromley 018B	E02000144	Bromley South
1	BR1 1AB	51.406333	0.015208	Bromley North	0.253666	E05000109	Bromley Town	E09000006	Bromley	E14000604	...	E01000676	Bromley 008B	E02000134	Bromley North & Sundridge
3	BR1 1AE	51.404543	0.014195	Bromley North	0.462939	E05000109	Bromley Town	E09000006	Bromley	E14000604	...	E01000677	Bromley 018C	E02000144	Bromley South
4	BR1 1AF	51.401392	0.014948	Bromley South	0.227662	E05000109	Bromley Town	E09000006	Bromley	E14000604	...	E01000675	Bromley 018B	E02000144	Bromley South
5	BR1 1AG	51.401392	0.014948	Bromley South	0.227662	E05000109	Bromley Town	E09000006	Bromley	E14000604	...	E01000675	Bromley 018B	E02000144	Bromley South

5 rows × 21 columns

In [45]:

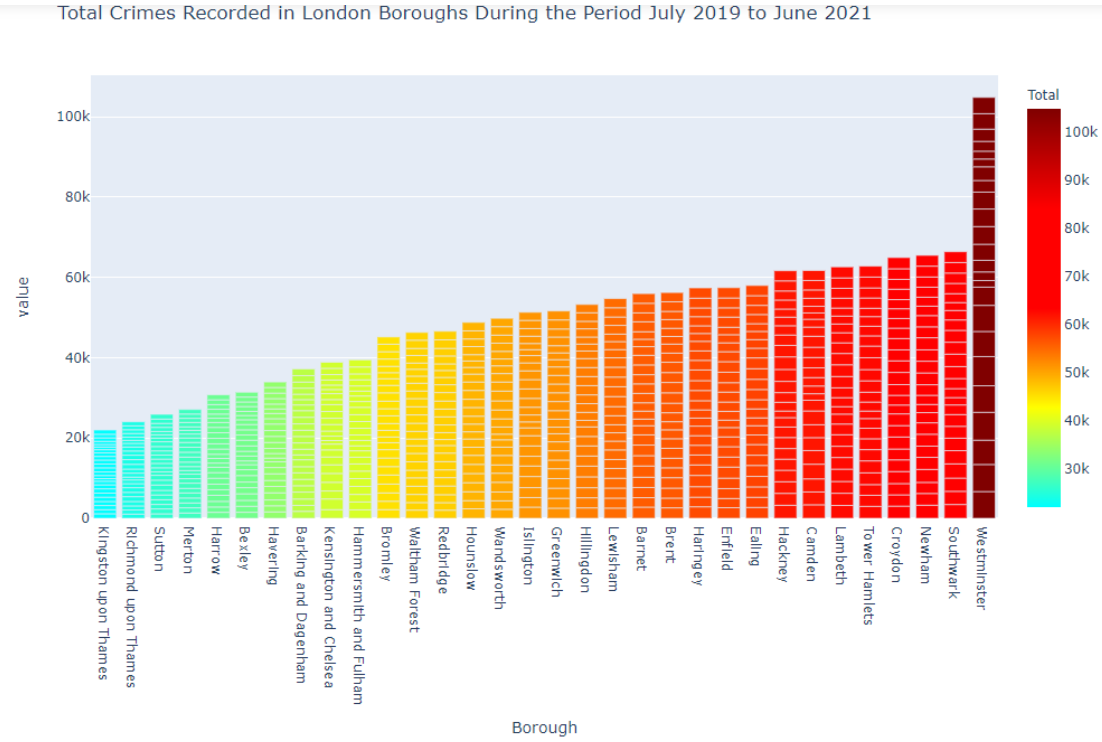
```
1 london_postcodes_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 179704 entries, 0 to 324633
Data columns (total 21 columns):
 #   Column                          Non-Null Count   Dtype  
---  ------                          --------------   -----  
 0   Postcode Data                   179704 non-null  object 
 1   Latitude Data                   179704 non-null  float64
 2   Longitude Data                  179704 non-null  float64
 3   Nearest Station                 179704 non-null  object 
 4   Distance To Station             179704 non-null  float64
 5   Ward Code                       179704 non-null  object 
 6   Ward                            179704 non-null  object 
 7   Borough Code                    179704 non-null  object 
 8   Borough                         179704 non-null  object 
 9   Constituency Code               179704 non-null  object 
 10  Constituency                    179704 non-null  object 
 11  LSOA Code                       179704 non-null  object 
 12  Lower Layer Super Output Area   179704 non-null  object 
 13  MSOA Code                       179704 non-null  object 
 14  Middle Layer Super Output Area  179704 non-null  object 
 15  London Zone                     179704 non-null  int64  
 16  Postcode Area                   179704 non-null  object 
 17  Postcode District               179704 non-null  object 
 18  Easting                         179704 non-null  int64  
 19  Northing                        179704 non-null  int64  
 20  Grid Ref                        179704 non-null  object 
dtypes: float64(3), int64(3), object(15)
memory usage: 30.2+ MB
```

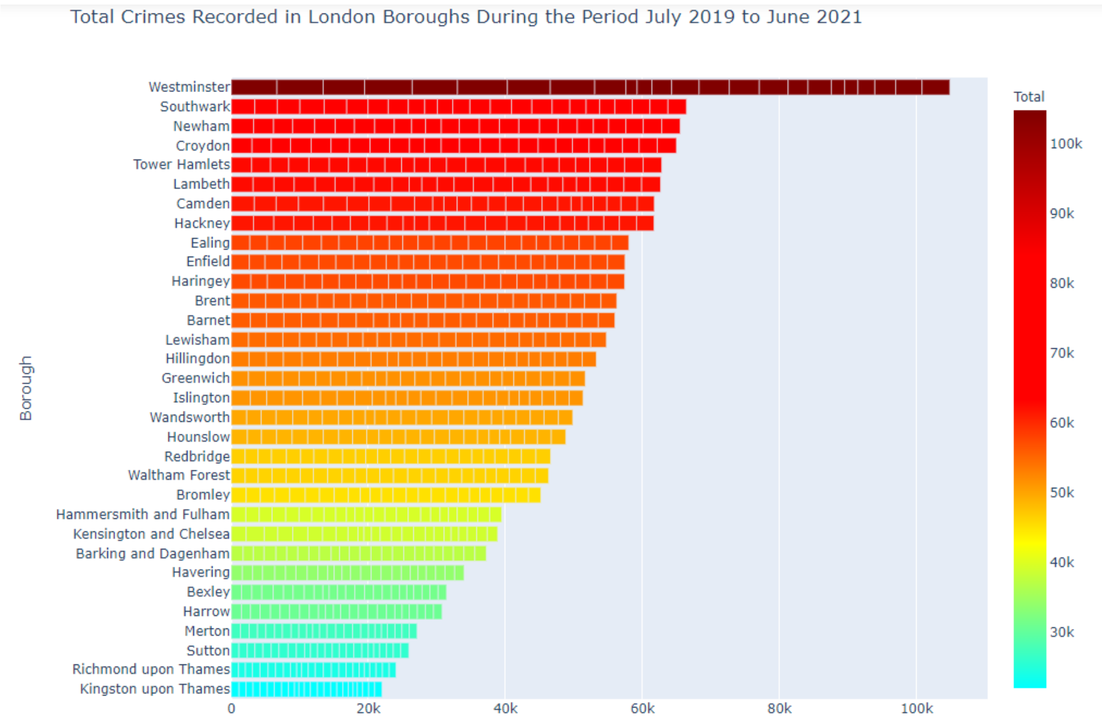
c. Understanding the Dataset Using Groupby Function and Charts [Index]

- We will then use the Groupby Function and Charts to understand the data better
  - During this process, the dataset will be used to find Boroughs that have the highest and the lowest crime rate
  - After having found the boroughs with the lowest crime rate, the data will be sorted, and the 5 safest Boroughs in London will be identified
  - Though the 5 Boroughs identified can easily serve our purpose, as these 5 Boroughs are the safest ones as compared to the other Boroughs of London; we will further try to eliminate the areas with crime so as to find the most secure venues for our target audience
  - If we take all the 92 Wards from the shortlisted 5 safe Boroughs, there may still be a possibility that some of the Venues could fall in the "unsafe" Ward of that particular safe Borough
  - Therefore, in order to avoid such a scenario and to ensure that the Venues found are from the most secure areas of London, another layer of safety will be added to identify the 10 Most Safest Wards within each of the 5 Most Safest Boroughs
  - Thus, out of a total of 615 Wards in the whole of London, we will shortlist only the 50 Most Safest Wards

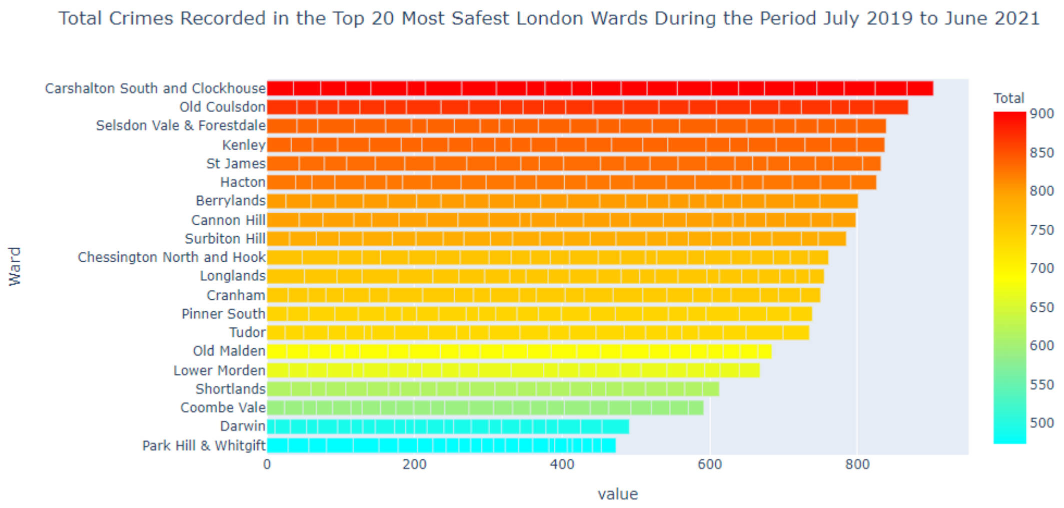
- Bar Chart of the Total Crimes Recorded During the Period July 2019 to June 2021



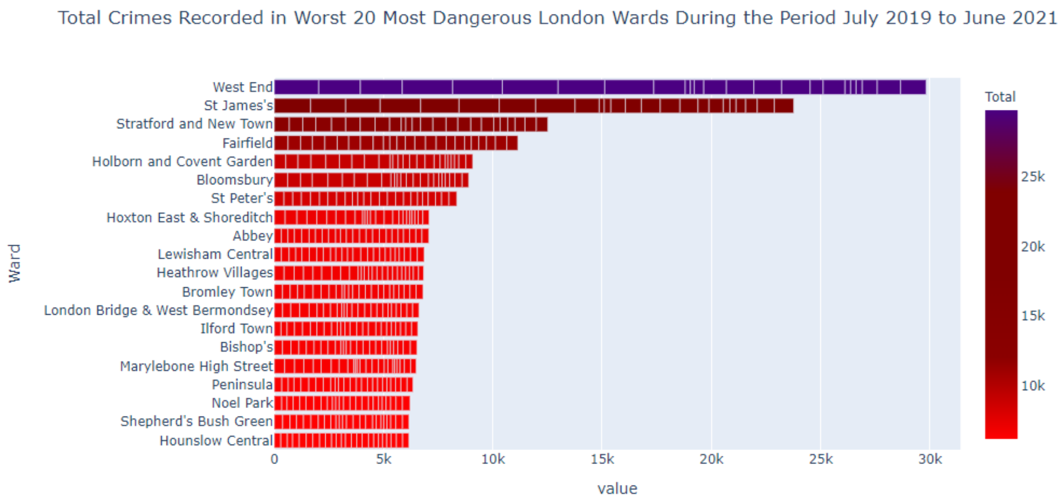
- Horizontal Bar Chart of the Total Crimes Recorded During the Period July 2019 to June 2021



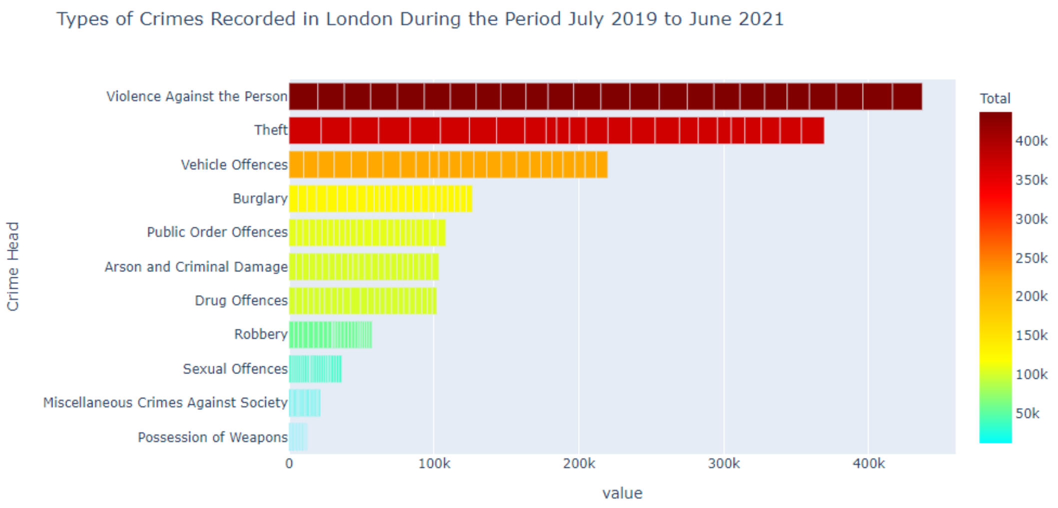
- Horizontal Bar Chart for the Total Crimes Recorded in the Top 20 Most Safest London Wards During the Period July 2019 to June 2021



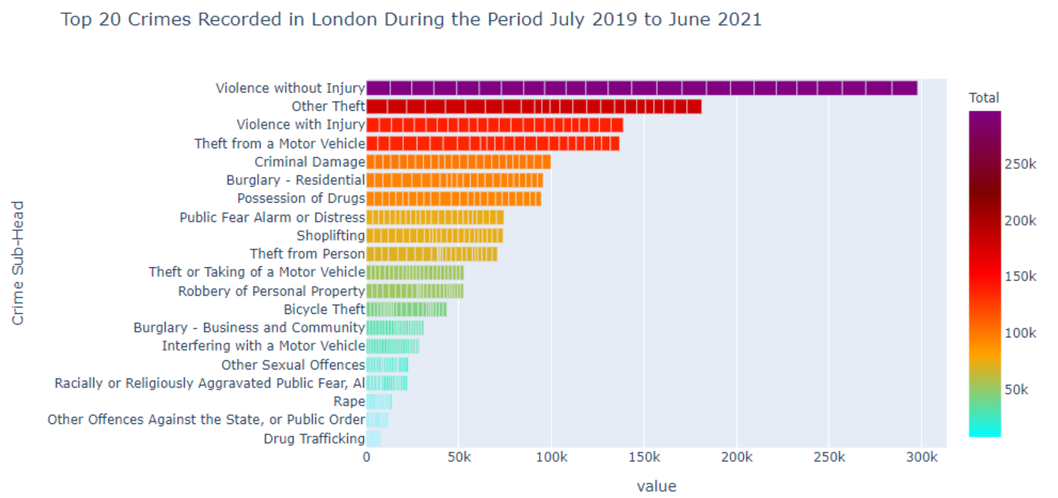
- Horizontal Bar Chart for the Total Crimes Recorded in Worst 20 Most Dangerous London Wards During the Period July 2019 to June 2021



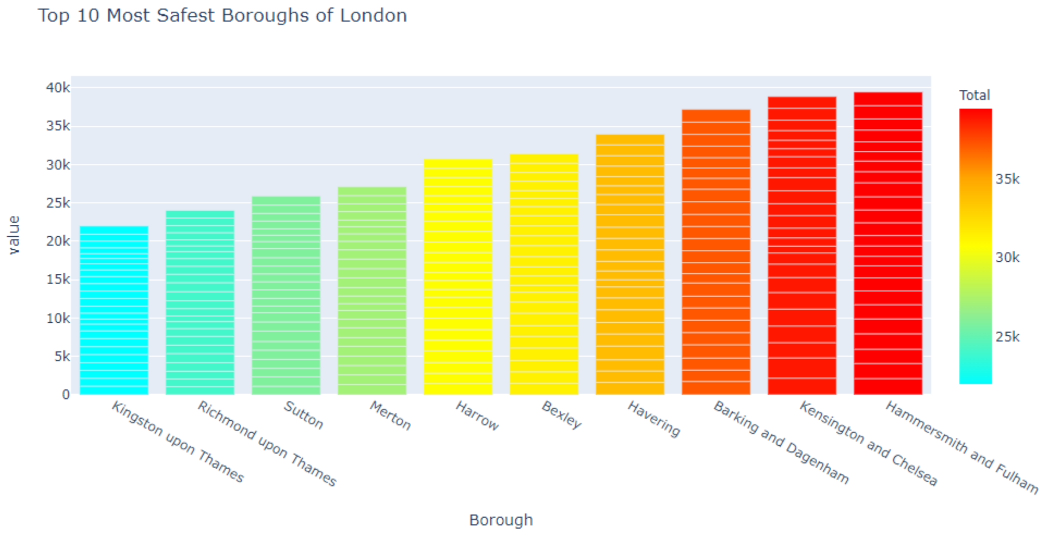
- Horizontal Bar Chart for the Types of Crimes Recorded in London During the Period July 2019 to June 2021



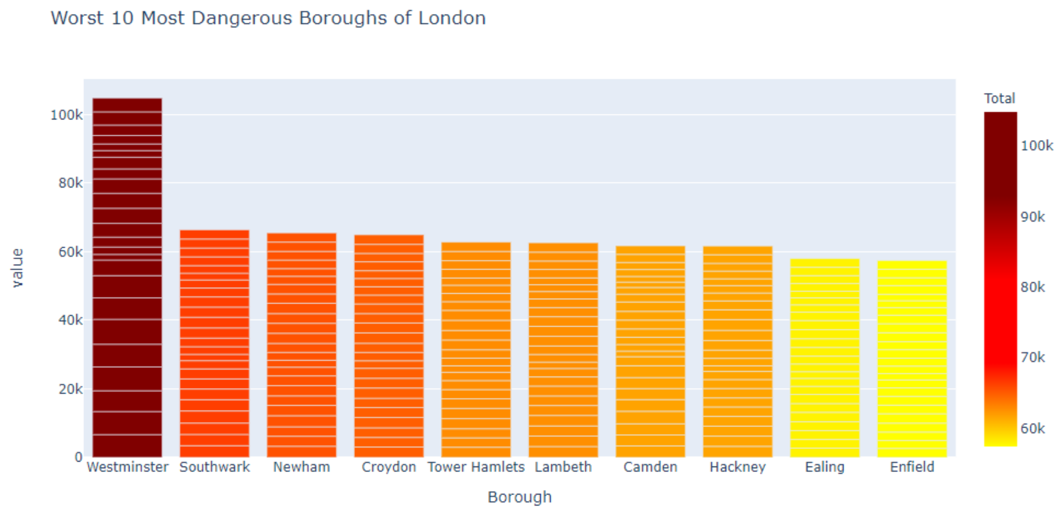
- Horizontal Bar Chart for the Top 20 Crimes Recorded in London During the Period July 2019 to June 2021



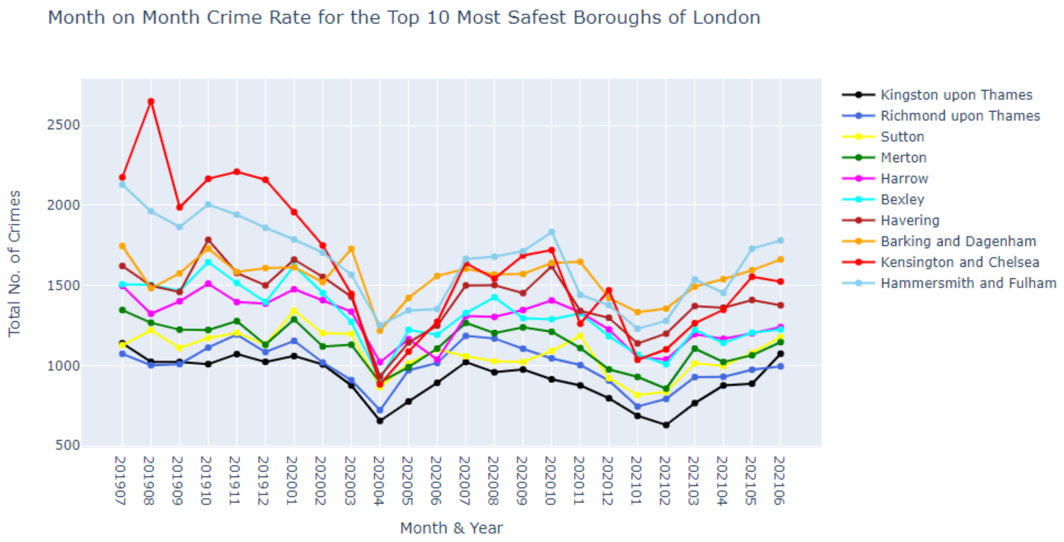
- Bar Chart for the Top 10 Most Safest Boroughs of London



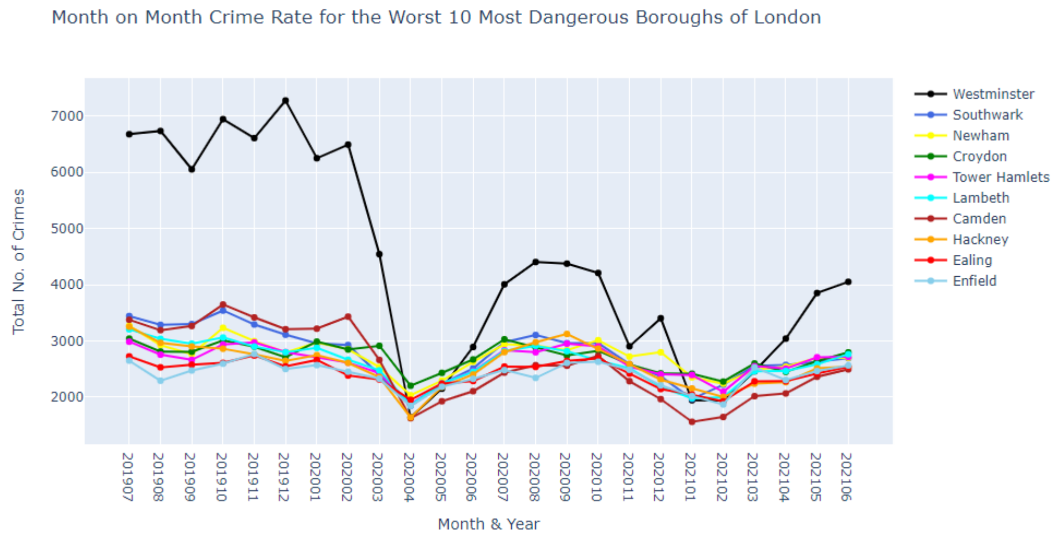
- Bar Chart for the Worst 10 Most Dangerous Boroughs of London



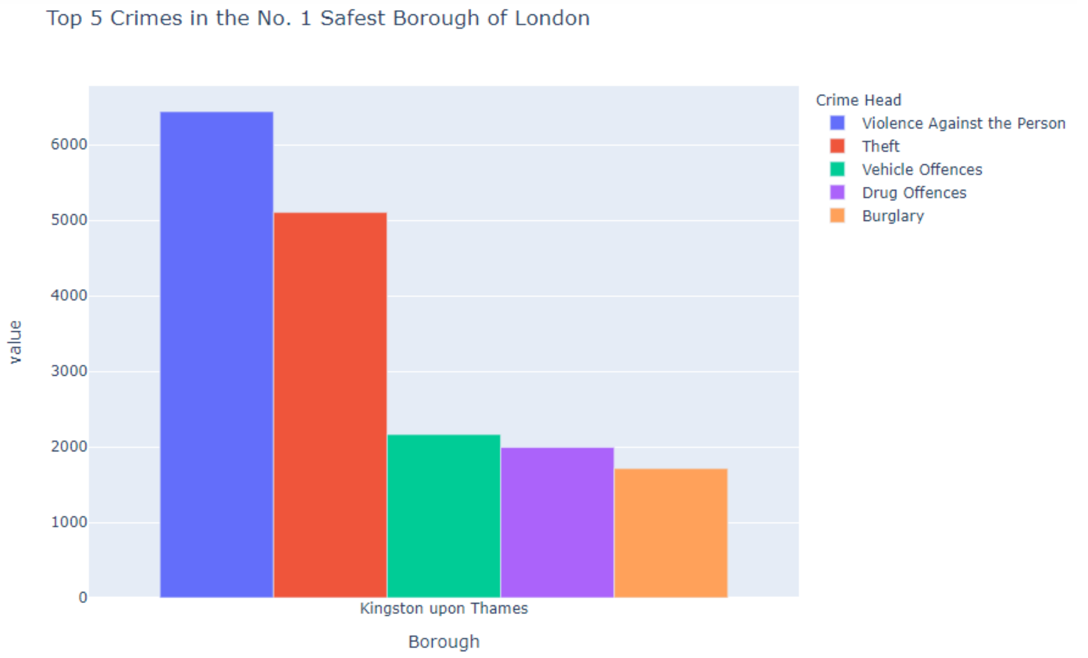
- Line Chart for the Month-on-Month Crime Rate for the Top 10 Most Safest Boroughs of London



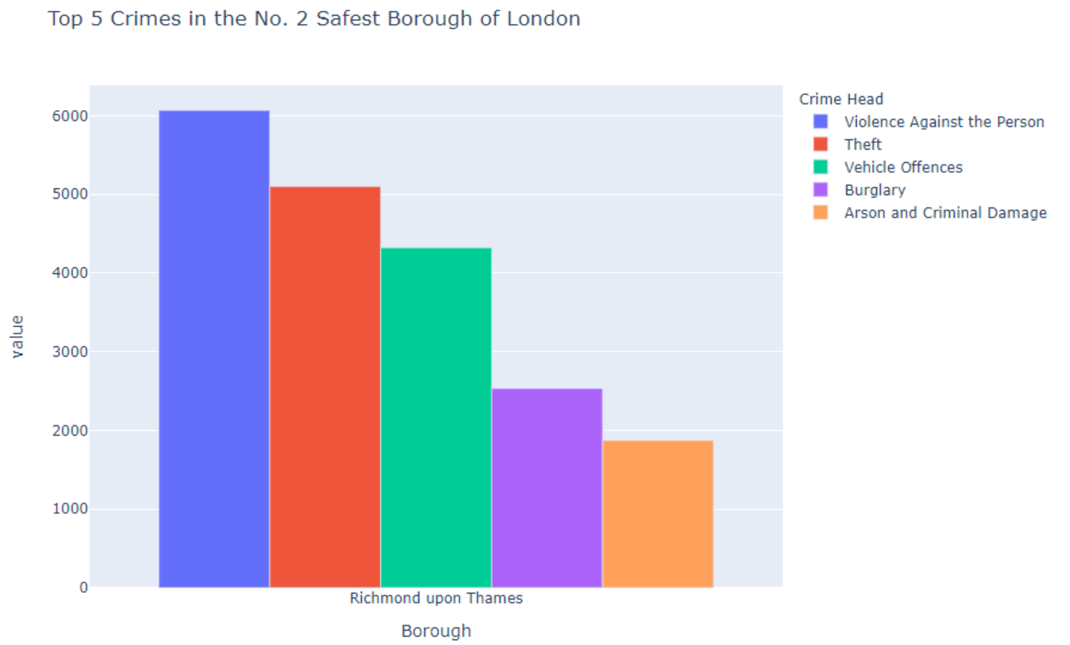
- Line Chart for the Month-on-Month Crime Rate for the Worst 10 Most Dangerous Boroughs of London



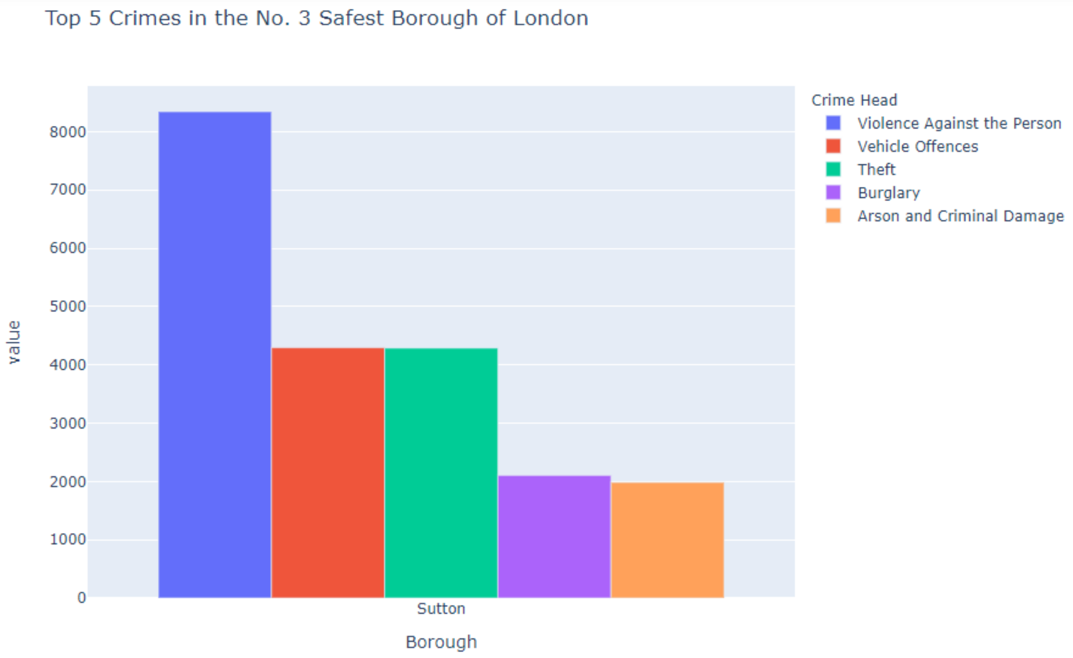
- Bar Chart for the Top 5 Crimes in the No. 1 Safest Borough of London, i.e., Kingston upon Thames



- Bar Chart for the Top 5 Crimes in the No. 2 Safest Borough of London, i.e., Richmond upon Thames



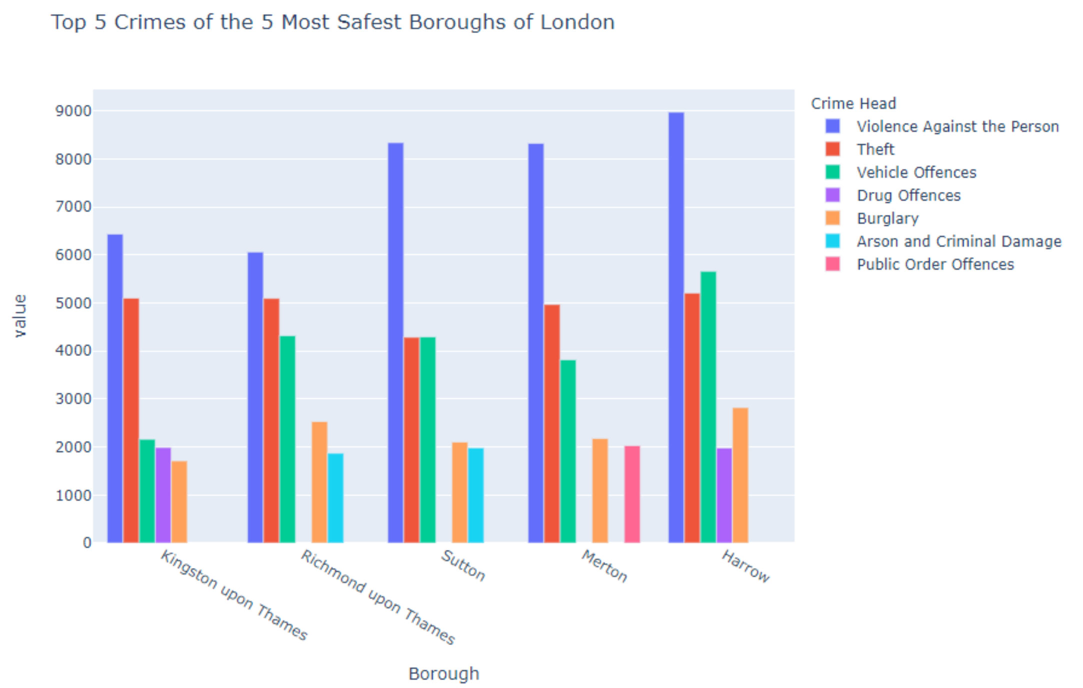
- Bar Chart for the Top 5 Crimes in the No. 3 Safest Borough of London, i.e., Sutton



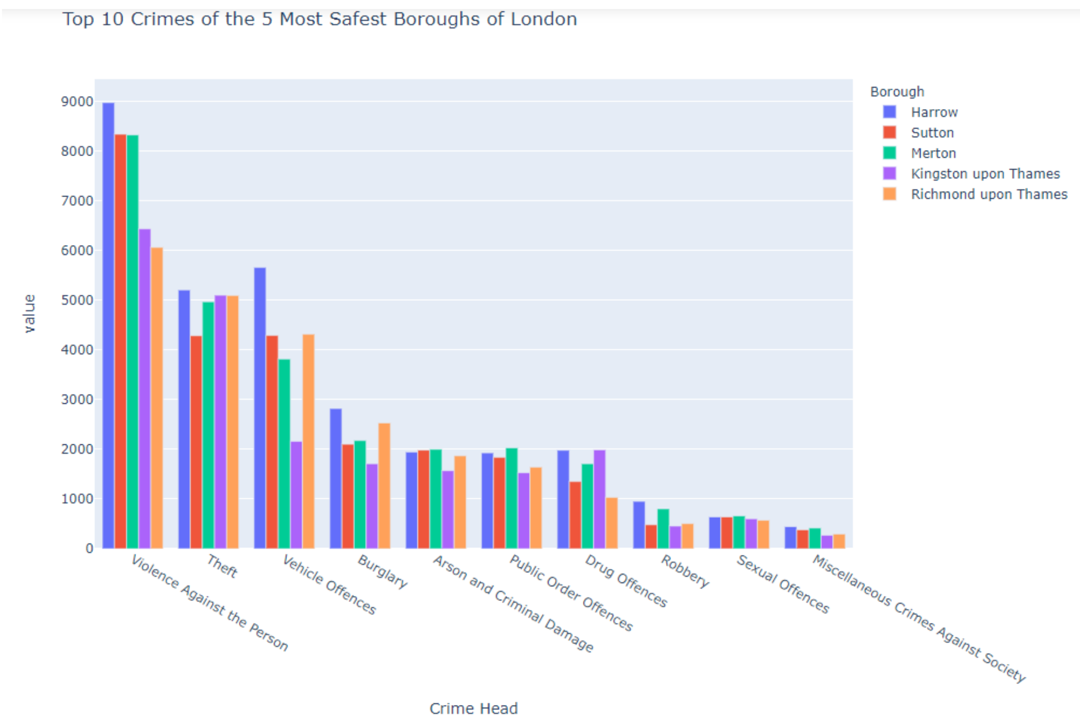
- Bar Chart for the Top 5 Crimes in the No. 4 Safest Borough of London, i.e., Merton

- Bar Chart for the Top 5 Crimes in the No. 5 Safest Borough of London, i.e., Harrow

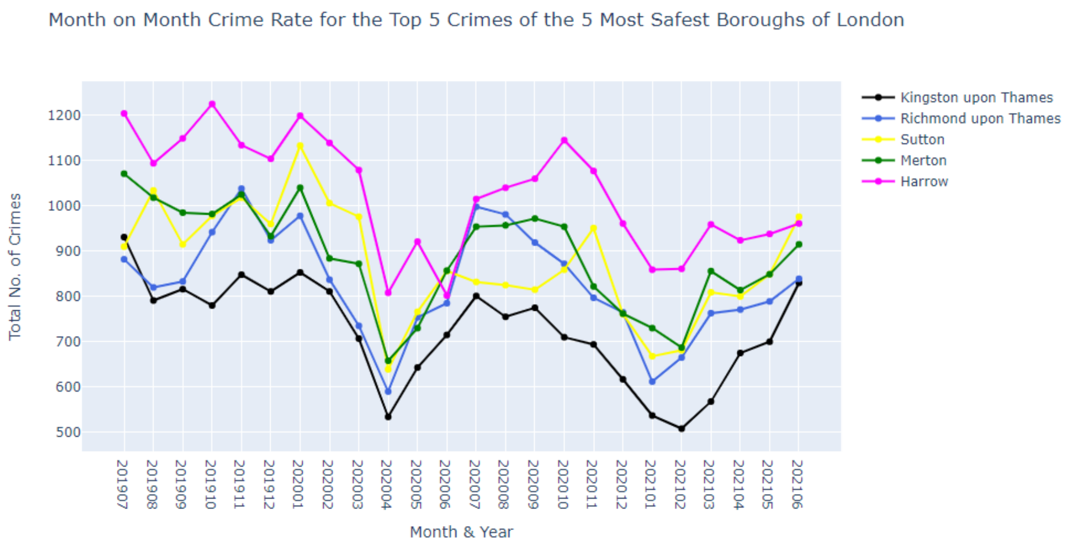
- Grouped Bar Chart for the Top 5 Crimes in the 5 Most Safest Boroughs of London Grouped By "Boroughs"



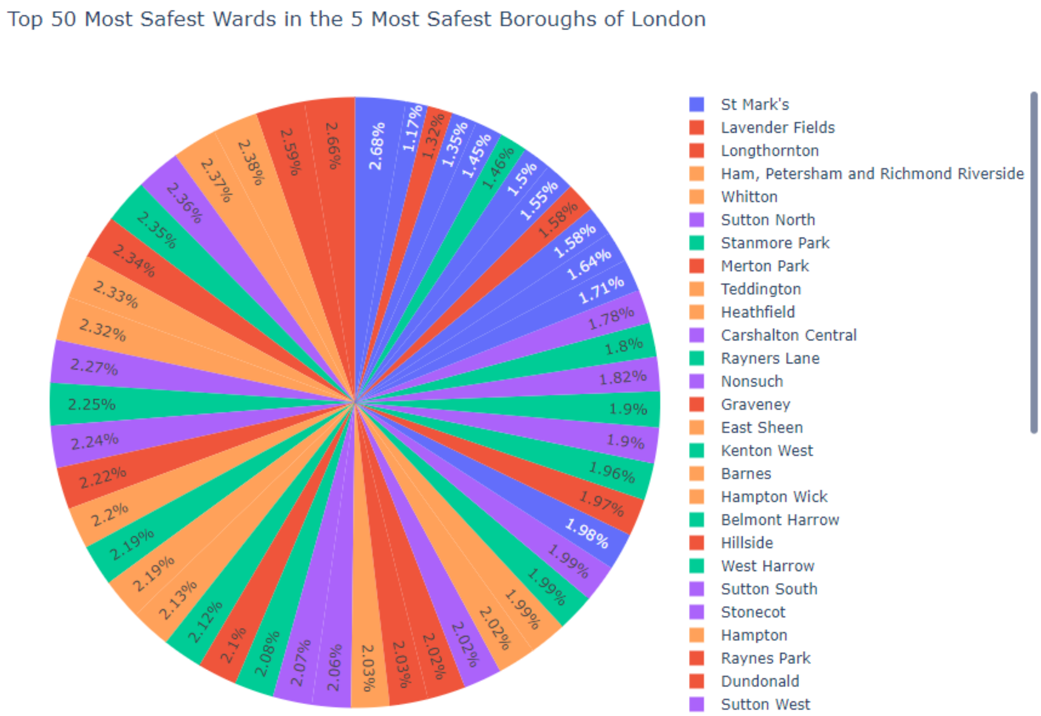
- Grouped Bar Chart for the Top 10 Crimes in the 5 Most Safest Boroughs of London Grouped By "Crime Head"



- Line Chart for the Month-on-Month Crime Rate for the Top 5 Crimes in the 5 Most Safest Boroughs



- Pie Chart for the Top 50 Most Safest Wards in the 5 Most Safest Boroughs of London

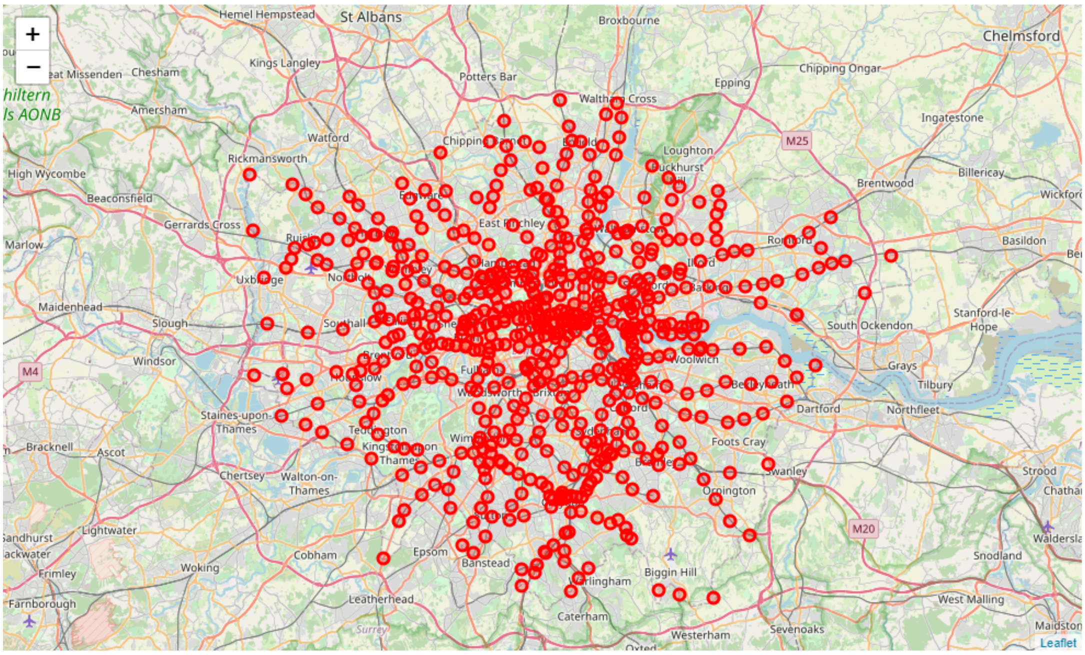


d. Collecting the Coordinates and Plotting them on the Map of London [Index]

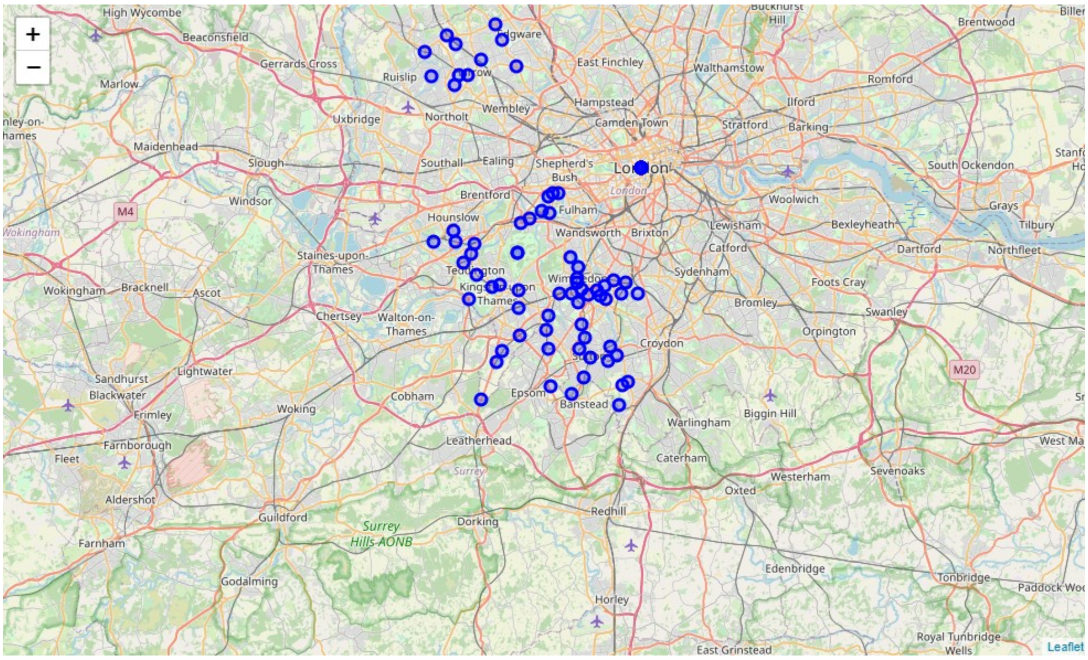
- Once we are done with identifying the safest Boroughs and Wards of London, we will extract the Postcodes of the different neighbourhoods in London
  - It should be noted that even though the dataset already has the Latitude and the Longitude data available as a part of the originally downloaded dataset, we will still be using the ArcGIS API to re-fetch the coordinates of the preferred locations
  - The dataset was processed to identify the postcodes of the Top 50 safest Wards of London
  - After cleaning the data, we find that there are still 10,083 records of the Postcodes for the Top 50 safest Wards of London
  - Now, we could have used these 10,083 Postcodes of the Top 50 Wards of London to find their coordinates, but the process of fetching the coordinates for so many postcodes would have taken a lot of time
  - Hence, it is necessary to reduce the number of records further
  - In order to further reduce the number of locations, the neighbourhoods having distance nearest to a station in these safe Wards will be selected
  - This process will not only reduce the number of locations, but it will also greatly assist the target audience, as finding venues that are nearer to the stations will reduce their travel time and will also be more convenient to them

- Since the number of stations that fall in these safe Wards are 81, after processing this requirement, the number of Postcodes reduce to 81 from 10,083
- Further, since we have selected only those Venues that are nearest to the Station, we will rename the column "Nearest Station" to "Neighbourhood" as all these neighbourhoods are very close to the respective Stations
- This dataset of Postcodes will then be used to fetch the geographical coordinates, i.e., the Latitude and Longitude, of the different neighbourhoods within the Top 50 safest Wards of London
- As discussed earlier, we will use the ArcGIS API to collect the Latitude and Longitude coordinates of the neighbourhoods based on their postcodes
- These coordinates will then be used to plot these locations on the Map of London

- Plotting All Stations of London on the Map of London



- Plotting Stations in the Safest Wards of London on the Map of London



e. Identifying Venues around the Safe Neighbourhoods of London [Index]

- The Latitude and Longitude coordinates will be linked with the Foursquare API to identify the different venues near these neighbourhoods
- In order to get the required information, we provide the Foursquare API with the Latitude and Longitude coordinates of the preferred neighbourhood
- Based on the Latitude and Longitude coordinates, the Foursquare API acquires information about different venues within each neighbourhood
- The data retrieved from the Foursquare API contains information of venues, which are within the radius of 500 metres of the latitude and longitude of said postcode
- The following information is obtained for each venue,
    - **Neighbourhood:** Name of the Neighbourhood
    - **Neighbourhood Latitude:** Latitude of the Neighbourhood
    - **Neighbourhood Longitude:** Longitude of the Neighbourhood
    - **Venue:** Name of the Venue
    - **Venue Category:** Category of the Venue
    - **Venue Latitude:** Latitude of the Venue
    - **Venue Longitude:** Longitude of the Venue
- In order to understand this information better, we will analyse the data using the Groupby function

f. Segmenting Neighbourhoods of London by Common Venue Categories [Index]

- Here, we will use One Hot Encoding on the column Venue Category
- This will convert all the values in the column Venue Category to those many different columns

```
In [190]:
1 venues_london_ohe = pd.get_dummies(venues_london[["Venue Category"]], prefix = "", prefix_sep = "")
2 venues_london_ohe.head()
```

Out[190]:

	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	BBQ Joint	Bagel Shop	...	Train Station	Tram Station	Tree	Turkish Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restauran
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
2	0	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0
3	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0

5 rows × 159 columns

- We will then print each Neighbourhood along with the Top 8 Most Common Venues in that Neighbourhood

```
---------Hampton---------
                Venue  Frequency
0               Hotel        3.0
1             Theater        3.0
2         Art Gallery        2.0
3          Art Museum        2.0
4               Plaza        2.0
5    Outdoor Sculpture       2.0
6                 Pub        2.0
7  Japanese Restaurant       1.0


---------Hampton Court---------
                 Venue  Frequency
0               Garden        4.0
1                 Café        4.0
2                  Pub        3.0
3                Hotel        2.0
4   Italian Restaurant        2.0
5        Grocery Store        1.0
6  Other Great Outdoors       1.0
7             Fountain        1.0


---------Hampton Wick---------
            Venue  Frequency
0             Pub        3.0
1           Hotel        1.0
2   Train Station        1.0
3            Park        1.0
4  Clothing Store        1.0
5     Coffee Shop        1.0
6     Sports Club        1.0
7           Plaza        1.0
```

- After this, we will create a dataframe having the columns Neighbourhood and the Top 8 Most Common Venues in those Neighbourhoods

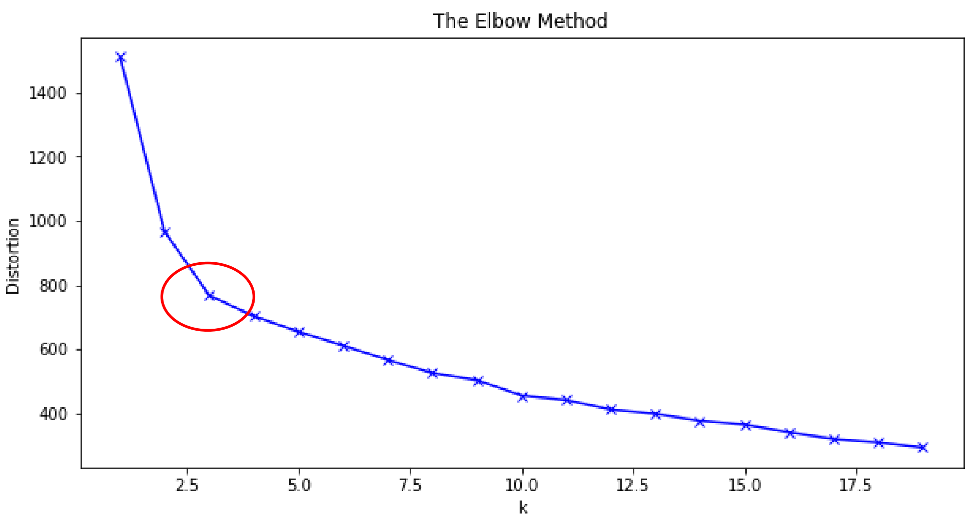
	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Ashtead	Clothing Store	Pub	Athletics & Sports	Auto Garage	American Restaurant	Optical Shop	Other Great Outdoors	Outdoor Sculpture
1	Banstead	Grocery Store	Indian Restaurant	Seafood Restaurant	Train Station	Fish & Chips Shop	Golf Course	Performing Arts Venue	Optical Shop
2	Barnes	Farmers Market	Food & Drink Shop	Pub	Park	Indie Movie Theater	Bookstore	Coffee Shop	Café
3	Barnes Bridge	Pub	Grocery Store	Coffee Shop	Farmers Market	Athletics & Sports	Gym / Fitness Center	Italian Restaurant	Harbor / Marina
4	Belgrave Walk	Grocery Store	Supermarket	Indian Restaurant	Irish Pub	Nature Preserve	North Indian Restaurant	Optical Shop	Other Great Outdoors

In [197]:
```
neighborhoods_venues_sorted.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 80 entries, 0 to 79
Data columns (total 9 columns):
 #   Column                 Non-Null Count  Dtype 
---  ------                 --------------  ----- 
 0   Neighbourhood          80 non-null     object
 1   1st Most Common Venue  80 non-null     object
 2   2nd Most Common Venue  80 non-null     object
 3   3rd Most Common Venue  80 non-null     object
 4   4th Most Common Venue  80 non-null     object
 5   5th Most Common Venue  80 non-null     object
 6   6th Most Common Venue  80 non-null     object
 7   7th Most Common Venue  80 non-null     object
 8   8th Most Common Venue  80 non-null     object
dtypes: object(9)
memory usage: 5.8+ KB
```

g. Clustering Neighbourhoods by Common Venues (K-Means Clustering) [Index]

- In order to assist our Target Audience to find venues of their choice in the safest neighbourhoods of London, we will be clustering the neighbourhoods using the K-Means Clustering Algorithm
- The K-Means Clustering Algorithm will cluster neighbourhoods with similar venues into different clusters
- We will first use the Elbow Method to identify the Optimal Number of Clusters
- **Elbow** method gives us an idea on what a good *k* number of clusters would be based on the sum of squared distance (SSE) between data points and their assigned clusters' centroids
- As per this method, the optimal number of clusters is achieved when the change in slope of the line becomes small
- Thu, we pick "k" at the spot where SSE starts to flatten out and forming an elbow



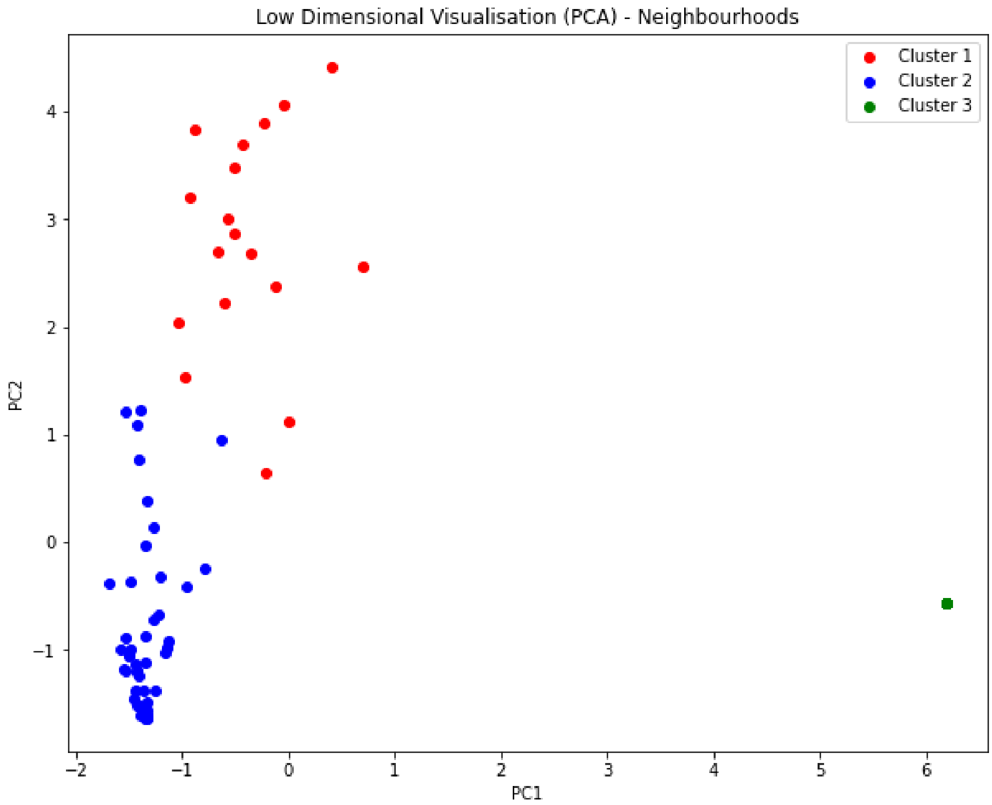
- After we have identified the optimal number of clusters, we will run the Machine Learning Algorithm to get the Cluster Labels

	Neighbourhood	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
0	Ashtead	1	Clothing Store	Pub	Athletics & Sports	Auto Garage	American Restaurant	Optical Shop	Other Great Outdoors	Outdoor Sculpture
1	Banstead	1	Grocery Store	Indian Restaurant	Seafood Restaurant	Train Station	Fish & Chips Shop	Golf Course	Performing Arts Venue	Optical Shop
2	Barnes	0	Farmers Market	Food & Drink Shop	Pub	Park	Indie Movie Theater	Bookstore	Coffee Shop	Café
3	Barnes Bridge	0	Pub	Grocery Store	Coffee Shop	Farmers Market	Athletics & Sports	Gym / Fitness Center	Italian Restaurant	Harbor / Marina
4	Belgrave Walk	1	Grocery Store	Supermarket	Indian Restaurant	Irish Pub	Nature Preserve	North Indian Restaurant	Optical Shop	Other Great Outdoors

In [222]:
```
neighborhoods_venues_sorted.info()
```

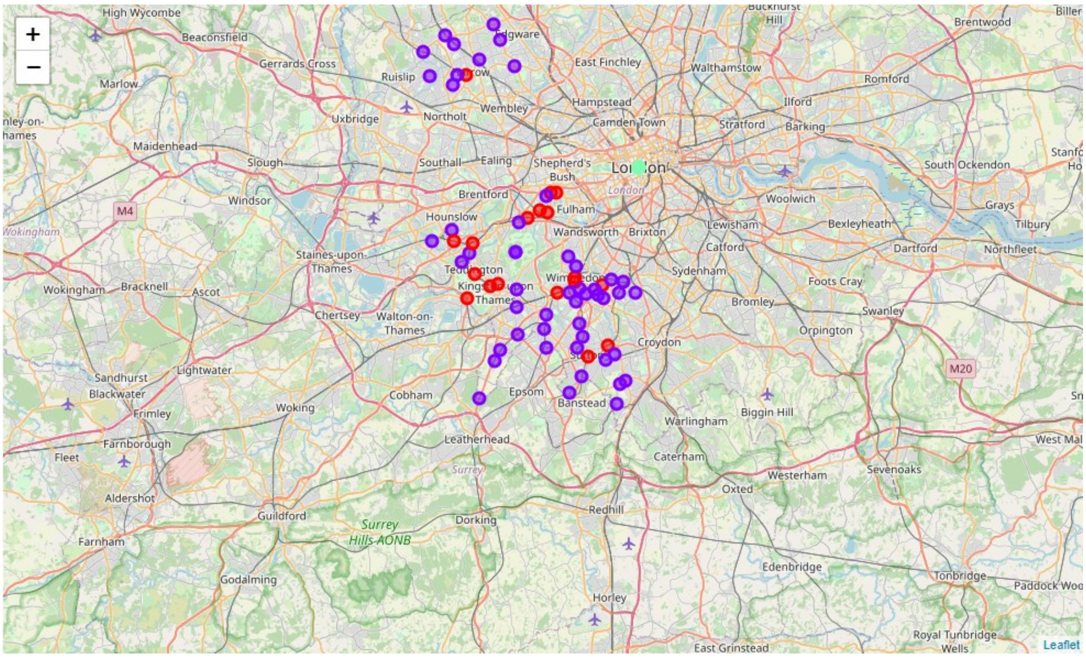
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 80 entries, 0 to 79
Data columns (total 10 columns):
 #   Column                 Non-Null Count  Dtype 
---  ------                 --------------  ----- 
 0   Neighbourhood          80 non-null     object
 1   Cluster Labels         80 non-null     int32 
 2   1st Most Common Venue  80 non-null     object
 3   2nd Most Common Venue  80 non-null     object
 4   3rd Most Common Venue  80 non-null     object
 5   4th Most Common Venue  80 non-null     object
 6   5th Most Common Venue  80 non-null     object
 7   6th Most Common Venue  80 non-null     object
 8   7th Most Common Venue  80 non-null     object
 9   8th Most Common Venue  80 non-null     object
dtypes: int32(1), object(9)
memory usage: 6.1+ KB
```

- Applying Dimensionality Reduction Techniques helps in visualising how the Clusters are related in the original high dimensional space
- Hence, in order to see how the Clusters are related in the original space, we will use Principal Component Analysis (PCA) to visualise the high dimensional data



- PCA also helps in finding if the features of the data are linearly related to each other
- It can be seen that the Explained Variance for the 10% of the Total Components, i.e., the first eight components, are able to preserve about 76% of the original information, thus, reducing the dimensionality of our data

- The clustered neighbourhoods will then be plotted on the Map of London



- **Cluster 1:**

	Neighbourhood	Ward	Borough	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
2	Barnes	Barnes	Richmond upon Thames	0.0	Farmers Market	Food & Drink Shop	Pub	Park	Indie Movie Theater	Bookstore	Coffee Shop	Café
3	Barnes Bridge	Barnes	Richmond upon Thames	0.0	Pub	Grocery Store	Coffee Shop	Farmers Market	Athletics & Sports	Gym / Fitness Center	Italian Restaurant	Harbor / Marina
8	Carshalton	Carshalton Central	Sutton	0.0	Pub	Grocery Store	Coffee Shop	Hotel	Train Station	Spa	Soccer Field	Park
13	Colliers Wood	Lavender Fields	Merton	0.0	Coffee Shop	Clothing Store	Sporting Goods Shop	Gym / Fitness Center	Convenience Store	Dry Cleaner	Electronics Store	Fast Food Restaurant
15	Dundonald Road	Dundonald	Merton	0.0	Burger Joint	Pub	Indian Restaurant	Bar	Spa	Multiplex	Sushi Restaurant	Burrito Place
20	Hammersmith (District)	Barnes	Richmond upon Thames	0.0	Pub	French Restaurant	Café	Lake	Park	Coffee Shop	Bar	Playground
22	Hampton Court	Hampton	Richmond upon Thames	0.0	Garden	Café	Pub	Hotel	Italian Restaurant	Grocery Store	Other Great Outdoors	Fountain
23	Hampton Wick	Hampton Wick	Richmond upon Thames	0.0	Pub	Hotel	Train Station	Park	Clothing Store	Coffee Shop	Sports Club	Plaza
25	Harrow-On-The-Hill	West Harrow	Harrow	0.0	Coffee Shop	Clothing Store	Sandwich Place	Fast Food Restaurant	Donut Shop	Pub	Pizza Place	Furniture / Home Store

```
In [216]:   1  cluster_1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 18 entries, 2 to 76
Data columns (total 12 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Neighbourhood          18 non-null     object
 1   Ward                   18 non-null     object
 2   Borough                18 non-null     object
 3   Cluster Labels         18 non-null     float64
 4   1st Most Common Venue  18 non-null     object
 5   2nd Most Common Venue  18 non-null     object
 6   3rd Most Common Venue  18 non-null     object
 7   4th Most Common Venue  18 non-null     object
 8   5th Most Common Venue  18 non-null     object
 9   6th Most Common Venue  18 non-null     object
 10  7th Most Common Venue  18 non-null     object
 11  8th Most Common Venue  18 non-null     object
dtypes: float64(1), object(11)
memory usage: 1.8+ KB
```

- ## Cluster 2:

	Neighbourhood	Ward	Borough	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
0	Ashtead	Chessington South	Kingston upon Thames	1.0	Clothing Store	Pub	Athletics & Sports	Auto Garage	American Restaurant	Optical Shop	Other Great Outdoors	Outdoor Sculpture
1	Banstead	Cheam	Sutton	1.0	Grocery Store	Indian Restaurant	Seafood Restaurant	Train Station	Fish & Chips Shop	Golf Course	Performing Arts Venue	Optical Shop
4	Belgrave Walk	Lavender Fields	Merton	1.0	Grocery Store	Supermarket	Indian Restaurant	Irish Pub	Nature Preserve	North Indian Restaurant	Optical Shop	Other Great Outdoors
5	Belmont	Belmont Sutton	Sutton	1.0	Pub	Asian Restaurant	Event Service	Train Station	Pet Store	Optical Shop	Other Great Outdoors	Outdoor Sculpture
6	Berrylands	Berrylands	Kingston upon Thames	1.0	Park	Pub	Platform	Coffee Shop	Train Station	Playground	Plaza	Pizza Place
7	Canons Park	Belmont Harrow	Harrow	1.0	Park	Indian Restaurant	Bar	Metro Station	North Indian Restaurant	Optical Shop	Other Great Outdoors	Outdoor Sculpture
9	Carshalton Beeches	Carshalton South and Clockhouse	Sutton	1.0	Train Station	Italian Restaurant	Grocery Store	Bakery	Pet Store	Optical Shop	Other Great Outdoors	Outdoor Sculpture
11	Chessington North	Chessington North and Hook	Kingston upon Thames	1.0	Indian Restaurant	Breakfast Spot	Platform	Convenience Store	Fast Food Restaurant	Fish & Chips Shop	Train Station	Grocery Store
12	Chessington South	Chessington South	Kingston upon Thames	1.0	Bar	Train Station	Golf Course	Supermarket	Playground	Platform	Newsagent	North Indian Restaurant

In [218]:
```
cluster_2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 50 entries, 0 to 80
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   Neighbourhood         50 non-null     object 
 1   Ward                  50 non-null     object 
 2   Borough               50 non-null     object 
 3   Cluster Labels        50 non-null     float64
 4   1st Most Common Venue  50 non-null     object 
 5   2nd Most Common Venue  50 non-null     object 
 6   3rd Most Common Venue  50 non-null     object 
 7   4th Most Common Venue  50 non-null     object 
 8   5th Most Common Venue  50 non-null     object 
 9   6th Most Common Venue  50 non-null     object 
 10  7th Most Common Venue  50 non-null     object 
 11  8th Most Common Venue  50 non-null     object 
dtypes: float64(1), object(11)
memory usage: 5.1+ KB
```

- **Cluster 3:**

	Neighbourhood	Ward	Borough	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
10	Cheam	Cheam	Sutton	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
21	Hampton	Hampton	Richmond upon Thames	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
29	Kenton	Kenton West	Harrow	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
32	Malden Manor	Old Malden	Kingston upon Thames	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
35	Morden	Merton Park	Merton	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
37	Morden South	Merton Park	Merton	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
40	New Malden	Coombe Vale	Kingston upon Thames	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
43	North Harrow	Headstone North	Harrow	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
47	Pinner	Pinner South	Harrow	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
48	Queensbury	Queensbury	Harrow	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant
50	Rayners Lane	Rayners Lane	Harrow	2.0	Hotel	Theater	Art Gallery	Art Museum	Plaza	Outdoor Sculpture	Pub	Japanese Restaurant

```
In [220]: 1 cluster_3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 12 entries, 10 to 65
Data columns (total 12 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Neighbourhood          12 non-null     object
 1   Ward                   12 non-null     object
 2   Borough                12 non-null     object
 3   Cluster Labels         12 non-null     float64
 4   1st Most Common Venue  12 non-null     object
 5   2nd Most Common Venue  12 non-null     object
 6   3rd Most Common Venue  12 non-null     object
 7   4th Most Common Venue  12 non-null     object
 8   5th Most Common Venue  12 non-null     object
 9   6th Most Common Venue  12 non-null     object
 10  7th Most Common Venue  12 non-null     object
 11  8th Most Common Venue  12 non-null     object
dtypes: float64(1), object(11)
memory usage: 1.2+ KB
```

- Based on the information collected, we will present our observations and findings, which will assist us in taking the necessary decisions

7. Links to Jupyter Notebook [Index]

- If, due to some reason, you are unable to view / open the Jupyter Notebook, Charts or Maps on GitHub, you may access the .ipynb file from the below mentioned links

- **Link to the Jupyter Notebook on IBM Cloud:**

  https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/ac0cc32a-2a4b-44c9-babc-2506856bace4/view?access_token=4bd1f1dea43b545d2abaa7391720e9b947f6a990af10990b92fe5490ec212b3a

- **Link to the Jupyter Notebook on Binder:**

  https://mybinder.org/v2/gh/vincyspereira/Coursera_Capstone/cd96eb73058886ece38132d0265b443e5aaecb58

  Note:

  - First click on the "*Week 5 – The Battle of Neighbourhoods (Part 2)*" folder
  - Next, click on the "*Capstone Project - The Battle of Neighbourhoods - London's Crime Rate Analysis and Clustering of the Safest Neighbourhoods of London.ipynb*" file to access the Jupyter Notebook
  - Lastly, click the 'File' Menu and then select 'Trust Notebook' to view the charts and maps

- **Link to the Jupyter Notebook using 'nbviewer':**

  https://nbviewer.jupyter.org/github/vincyspereira/Coursera_Capstone/blob/cd96eb73058886ece38132d0265b443e5aaecb58/Week%205%20-%20The%20Battle%20of%20Neighborhoods%20(Part%202)/Capstone%20Project%20-%20The%20Battle%20of%20Neighbourhoods%20-%20London's%20Crime%20Rate%20Analysis%20and%20Clustering%20of%20the%20Safest%20Neighbourhoods%20of%20London.ipynb

- **Link to the Jupyter Notebook on GitHub:**

  https://github.com/vincyspereira/Coursera_Capstone/blob/cd96eb73058886ece38132d0265b443e5aaecb58/Week%205%20-%20The%20Battle%20of%20Neighborhoods%20(Part%202)/Capstone%20Project%20-%20The%20Battle%20of%20Neighbourhoods%20-%20London's%20Crime%20Rate%20Analysis%20and%20Clustering%20of%20the%20Safest%20Neighbourhoods%20of%20London.ipynb

  Note:

  If you are unable to view the code / charts properly on GitHub, then you may either:

  - Click on the "Circle with Horizontal Line" symbol on the top right-hand corner to view the Jupyter Notebook with "nbviewer"

    OR

  - Click on the "Download" button to download the .ipynb file

8. Results and Discussion [Index]

- The aim of this project is to help the Migrants and Tourists who want to explore the safest neighbourhoods of London
- They can decide to stay or visit a specific neighbourhood based on their preferred cluster
- Based on the type of clusters, different people, i.e., families with children, young couples, executives, or tourists, can decide which neighbourhood is best suited for them
- **Cluster 1:**
  - This cluster is mostly made up of Pubs, Coffee Shops, Cafe, Multi-Cultural Restaurants, Bars, Gyms, Sports Clubs, Supermarkets, Grocery Stores, Shopping Plazas, Fast-food Joints, etc.
  - Thus, this cluster is most suitable for young couples and executives
- **Cluster 2:**
  - This is the biggest cluster from our Dataset
  - It is mostly made up of Supermarkets, Bakeries, Pharmacies, Auto Garages, Parks, Playgrounds, Sports Complexes, Multi-Cultural Restaurants, Ice Cream Parlours, Fish & Chips Shops, Pubs, various stores like, Grocery, Convenience, Clothing, Furniture, Pet, Optical, Electronics, Warehouse, etc., and Train Stations
  - It has almost everything that a family requires
  - Thus, this cluster seems to be most suitable for families with children

- **Cluster 3:**
    - This cluster is mostly made up of Hotels, Pubs, Theatres, Art Galleries, Art Museums, Outdoor Sculptures and Plazas
    - Thus, this cluster is most suitable for Tourists
- This segmentation is also proved right from the PCA Chart
- According to PCA, Cluster 1 and Cluster 2 seem to be Linearly Related, while Cluster 3 is not at all related to the other two clusters
- As can be seen above, Clusters 1 & 2 seem to suit Migrants, who intend to stay in neighbourhoods falling in those clusters, while Cluster 3 seems to suit Tourists, who intend to visit neighbourhoods falling in that cluster

9. Conclusion [Index]

- This Capstone Project will help families with children, young couples, executives, and tourists, to understand,
  - which are the safe Boroughs, Wards and Neighbourhoods of London
  - the most common venues in those neighbourhoods
  - the different types of neighbourhoods based on the cluster of venue categories
  - which neighbourhoods to choose as per their preference
- As can be seen from the data on clusters, the aim of the project to seems to have been fulfilled.

End of Report