Coursera IBM Data Science Professional Capstone Project

TITLE: **The Battle of London Neighbourhoods**

Student Name: **Lucky Munemo**

PAGE INTENTIONALLY LEFT BLANK

CONTENTS

Page

INTRODUCTION

The purpose of this project was to investigate best locations in London, United Kingdom, for buy-to-let and/or short-term rental property investment through application of data science and machine learning (k-Means algorithm) tools and techniques. Analysis results from this project would help property/real estate investors to make smart and profitable decisions on building their property portfolios.

The majority of property or real estate investors will say that the three most important factors when investing in a property are; first is location, second is location and third is location. This notion is common whether the investor is doing traditional buy-to-let (sometimes known as single lets), Housing in Multiple Occupancy (HMOs) or Short-term lets (serviced accommodation, holiday lets, Airbnb, booking.com, Expedia, laterooms.com, HomeAway.co.uk, TripAdvisor etc).

Why is location so important? Firstly, location is constant and cannot be changed while the property itself can be changed by renovations. Secondly, the location determines the supply and demand for rental properties. Third, location determines desirability and future value of your property due to proximity to value adding amenities such as shops, transport, pubs, lifestyle centres, entertainment, cafes, restaurants etc. Fourth, Location determines the optimal rental strategy i.e buy-to-let, HMO or short term lets. Fifth, location determines profitability. Cities are generally more expensive than rural areas for example although it does not necessarily mean that cities are always profitable.

The investigation was carried out using data from Airbnb listings, crime data and residential property prices from London official government website, London official geolocation data, trending venues data from Foursquare.com API. The computing resources used were IBM Watson Studio platform with 1 vCPU, 4GB RAM, 2GB data storage, Jupiter Notebook and Python programming language. The Jupiter Notebook is available on Github .

The analysis focused on discovering best locations for short-term rentals in London. Exploration determined which neighbourhoods and Boroughs have the highest number of Airbnb listings, types of properties listed, price per night distribution, average property purchase price, crime rates distribution. The top 4 neighbourhoods were then explored for the most trending venues and then divided into clusters using Foursquare.com API (Application Programming Interface) and k-Means algorithm was implemented to the top 4 neighbourhoods' most trending venues. K-Means algorithm is a simple, effective and popular clustering algorithm for unsupervised machine learning which groups similar data points into a pre-defined number of groups. The number of groups is the only input required from user when using k-Means algorithm.

It was concluded that a buy-to-let or short-term rental property investor would better select a location where property purchase price is relatively low or average, average crime rates, high short-term rental listings (although high competition but maintains good occupancy rates due to high demand), good proximity to amenities (such as shops, pubs, bars, restaurants, fitness and wellness centres, coffee shops and cafes). The recommended price range per night was less than £200.

This report consists of Introduction, Datasets, Methodology, Analysis and Exploration, Results, Conclusions and Recommendations.

THE DATASET

The dataset used in this project was obtained from www.insideairbnb.com . The raw dataset for listings (dated 14 September 2019) was selected because it contains all the necessary data that are required for this project. Any unnecessary data will be discarded during the data cleaning process. The most important data required for this project are Post Code, Borough, Neighbourhood, Latitude, Longitude was obtained from https://www.doogal.co.uk/UKPostcodesCSV.ashx?region=E12000007 . Average property price data was extracted from https://data.london.gov.uk/download/average-house-prices/b1b0079e-698c-4c0b-b8c7-aa6189590ca4/land-registry-house-prices-borough.csv . London crime data was extracted from https://data.london.gov.uk/download/recorded_crime_rates/c051c7ec-c3ad-4534-bbfe-6bdfee2ef6bb/crime%20rates.csv . Trending venues data for the top neighbourhoods were obtained from https://foursquare.com/developers/apps via the API. Note that a developer's account is required for Foursquare.com API. This additional data was used for exploring, segmenting and clustering the neighbourhoods and the most trending venues within the top neighbourhoods. The London Neighbourhoods json file is also available and will be used within this project.

METHODOLOGY

Data cleaning and feature selection

The data obtained from www.insideairbnb.com was very raw but the best available. The dataset was converted into a pandas dataframe which had 85273 rows by 106 columns. Figure 1 shows some of the rows and columns.

```
In [6]: #Converting the data into a pandas dataframe with all columns. This will be useful for column selection during data cleaning.
pd.set_option('display.max_columns', len(london_raw_data.columns))
pd.set_option('display.max_rows', 200)
df_london_raw = pd.DataFrame(london_raw_data)
df_london_raw.head(2)
```

Out[6]:

od_overview	notes	transit	access	interaction	house_rules	thumbnail_url	medium_url	picture_url	xl_picture_url	host_id	
10 minutes by Victoria Sta...	No Smoking (very strict) Check-in time is afte...	Tons of buses (24hrs) go into central London f...	Guest will have access to the entire apartment	No interaction with guests as you book the ent...	No Smoking (very strict) No pets are allowed i...	NaN	NaN	https://a0.muscache.com/im/pictures/1d720898-c...	NaN	43039	https://www.airbnb.com/use
rk is a friendly ɔot commun...	For art lovers I can give guest my Tate Member...	The flat only a 10 minute walk to Finsbury Par...	Guest will have access to the self catering ki...	I like to have little chats with my guest over...	I'm an artist and have my artwork up on the wa...	NaN	NaN	https://a0.muscache.com/im/pictures/ffb507b7-9...	NaN	54730	https://www.airbnb.com/use

```
In [7]: df_london_raw.shape
Out[7]: (85273, 106)
```

Figure 1. Dataframe showing a snapshot of raw data extracted from insideairbnb.com

After dropping unwanted columns and rows with missing values the dataframe was reduced to 65608 rows by 11 columns as shown in Figure 2.

```
In [20]: # delete every row which has NaN value(s) in any of the nightly price column.
df_data = df_data.dropna(how='any',subset=['Price'])
df_data.head()
```

Out[20]:

	Neighbourhood	Borough	PostCode	Latitude	Longitude	Property_Type	Room_Type	Price	Weekly_Price	Monthly_Price	Location_Review_Scores
0	Brixton	Lambeth	SW9 8DG	51.46225	-0.11732	Apartment	Entire home/apt	88.0	645.0	2350.0	9.0
1	LB of Islington	Islington	N4 3	51.56802	-0.11121	Apartment	Private room	65.0	333.0	1176.0	9.0
2	Chelsea	Kensington and Chelsea	SW3	51.48796	-0.16898	Apartment	Entire home/apt	100.0	600.0	2250.0	10.0
3	Fitzrovia	Westminster	W1T4BP	51.52098	-0.14002	Apartment	Entire home/apt	300.0	1378.0	NaN	10.0
4	Battersea	Wandsworth	SW11 5GX	51.47298	-0.16376	Townhouse	Entire home/apt	175.0	1050.0	3500.0	9.0

```
In [21]: #Checking the dataframe shape
df_data.shape
Out[21]: (65608, 11)
```

Figure 2. Cleaned dataframe for Airbnb listings in London

Addressing the London geographical location data problem

It can be observed from Figure 2 that London Post Codes are not the easiest to process for data science applications. They have different lengths and formats. To add to the problem, Airbnb seems to have allowed users to enter the postcodes themselves without format validation at data entry. This resulted in a toxic cocktail of geolocation data for London in the dataframe extracted from insideairbnb.com shown in Figure 2. There was now a serious need get official geolocation data for London. First attempt was to visit https://en.wikipedia.org/wiki/London_postal_district to look for official references. The was some interesting history but that was not very useful to my project. After searching, it was finally settled for https://www.doogal.co.uk/UKPostcodesCSV.ashx?region=E12000007 where after selecting the relevant features for the project, the resulting dataframe is shown in Figure 3.

```
In [28]: #Renaming columns
df_london_postcode1.columns = ('PostCode', 'Neighbourhood','Borough','Neighbourhood Postcode','Latitude', 'Longitude',
                    'Nearest Train Station', 'Distance to Station', 'London Zone')
df_london_postcode1.head()
```

Out[28]:

	PostCode	Neighbourhood	Borough	Neighbourhood Postcode	Latitude	Longitude	Nearest Train Station	Distance to Station	London Zone
0	BR1 1AA	Bromley Town	Bromley	BR1	51.401546	0.015415	Bromley South	0.218257	5
1	BR1 1AB	Bromley Town	Bromley	BR1	51.406333	0.015208	Bromley North	0.253666	4
2	BR1 1AD	Bromley Town	Bromley	BR1	51.400057	0.016715	Bromley South	0.044559	5
3	BR1 1AE	Bromley Town	Bromley	BR1	51.404543	0.014195	Bromley North	0.462939	4
4	BR1 1AF	Bromley Town	Bromley	BR1	51.401392	0.014948	Bromley South	0.227664	5

Figure 3. London official geolocation data.

The London geolocation dataframe was merged with Airbnb listings dataframe to ensure that each Airbnb listing had the correct and official geolocation data attributes. After cleaning the data, the resultant dataframe is shown in Figure 4.

```
In [34]: df_merged_london2.head()
```

Out[34]:

	PostCode	Neighbourhood	Borough	Neighbourhood Postcode	Latitude	Longitude	Nearest Train Station	Distance to Station	London Zone	Property Type	Room Type	Price	Location Review Scores
0	BR1 1HW	Bromley Town	Bromley	BR1	51.400948	0.013574	Bromley South	0.282734	5	Apartment	Entire home/apt	69.0	10.0
1	BR1 1HW	Bromley Town	Bromley	BR1	51.400948	0.013574	Bromley South	0.282734	5	House	Entire home/apt	69.0	10.0
2	BR1 1HW	Bromley Town	Bromley	BR1	51.400948	0.013574	Bromley South	0.282734	5	Bungalow	Entire home/apt	89.0	10.0
3	BR1 1JB	Bromley Town	Bromley	BR1	51.402570	0.014396	Bromley South	0.351683	5	House	Private room	25.0	9.0
4	BR1 1NA	Bromley Town	Bromley	BR1	51.405563	0.015534	Bromley North	0.323242	4	Apartment	Entire home/apt	119.0	10.0

Figure 4. Merged Airbnb listings and official London geolocation dataframes

Acquiring and cleaning London crime data

Short-term rental market usually thrives on good experiences for guests. Crime is one of the main things which usually ruins the experience for short-term rental guests. This fact made it necessary to consider London crime data for each Borough. Although, a going granular would be more detailed, it was deemed unnecessary for the scope of this project, but can be added to future improvements of this project.

London crime data was sourced as a .csv file from the London government website. The data contained all records from 1999 to 2017 for each London Borough as shown in Figure 5.

	Code	Borough	Year	Offences	Rate	Number_of_offences
6640	E09000019	Islington	2016-17	Other Notifiable Offences	1.9	437
6641	E09000020	Kensington and Chelsea	2016-17	Other Notifiable Offences	1.9	299
6642	E09000021	Kingston upon Thames	2016-17	Other Notifiable Offences	0.8	146
6643	E09000022	Lambeth	2016-17	Other Notifiable Offences	1.7	570
6644	E09000023	Lewisham	2016-17	Other Notifiable Offences	1.4	427
6645	E09000024	Merton	2016-17	Other Notifiable Offences	1.0	207
6646	E09000025	Newham	2016-17	Other Notifiable Offences	1.6	558
6647	E09000026	Redbridge	2016-17	Other Notifiable Offences	1.1	338
6648	E09000027	Richmond upon Thames	2016-17	Other Notifiable Offences	0.8	159
6649	E09000028	Southwark	2016-17	Other Notifiable Offences	1.3	414
6650	E09000029	Sutton	2016-17	Other Notifiable Offences	1.0	204
6651	E09000030	Tower Hamlets	2016-17	Other Notifiable Offences	1.9	591
6652	E09000031	Waltham Forest	2016-17	Other Notifiable Offences	1.3	360

Figure 5. London raw crime data

For the purpose of this project, only latest data were considered from 2016-2017. One other thing to note here is that the London Metropolitan Police defined the London Boroughs slightly differently as can be observed from the 'Borough column'. For example, the definition for Inner London and Outer London is generally open for debate. For consistency, in this particular project, there was need to merge the crime, geolocation and Airbnb dataframes in order to maintain the officially defined London Boroughs.

The next step was to keep only 2016-2017 crime data, total offences only and drop all rows with missing data. Since this project is not about crime analysis, it was decided that the best approach was to extract the all recorded offences for 2016-2017, total number of offences for each Borough and corresponding crime rate. The resulting clean data will be used for further analysis later on in the project is shown in Figure 6.

```
In [52]: # Creating new dataframe with only 2016-2017 crime data but with total offences only
         df_london_crime3 = df_london_crime2[df_london_crime2['Type of Offences'] == 'All recorded offences']
         df_london_crime3.head()
```

Out[52]:

	Borough	Year	Type of Offences	Crime Rate	Total Offences
0	Bromley	2016-17	All recorded offences	52.7	17243
10	Lewisham	2016-17	All recorded offences	68.4	20654
20	Lambeth	2016-17	All recorded offences	90.8	29787
30	Croydon	2016-17	All recorded offences	67.4	25777
40	Greenwich	2016-17	All recorded offences	71.6	20023

```
In [53]: #Checking dataframe dimensions
         df_london_crime3.shape

Out[53]: (32, 5)
```

Figure 6. Clean London crime dataframe

Obtaining and cleaning London property prices

Taken in a vacuum, short-term rental data from Airbnb only tells half the story. The next step is combining the data with research on the least-expensive housing neighbourhoods. Finding a sweet spot between low mortgage (or leasing) rates and high returns on investment would be the best strategy.

The London property prices was sourced as a .csv file from the London government website and the dataframe contained house price data from 1995 to 2017. For this project purposes, only property price data for only 2016-2017 was considered for fair comparison with crime data for the same period. The impact of Brexit uncertainty was not be considered for this project. The resulting clean data will be used for further analysis later on in the project.

```
In [94]: # Creating new dataframe with only 2016-2017 property mean price data
         df_property_mean_price = df_house_prices1[df_house_prices['Measure'] == 'Mean']
         #Renaming column names
         df_property_mean_price.columns = ['Borough','Year','Measure','Mean_Price']
         df_property_mean_price.head()

         /opt/conda/envs/Python36/lib/python3.6/site-packages/ipykernel/__main__.py:2: UserWarning:
         ex.
           from ipykernel import kernelapp as app
```

Out[94]:

	Borough	Year	Measure	Mean_Price
7965	City of London	Year ending Dec 2017	Mean	950760
7966	Barking and Dagenham	Year ending Dec 2017	Mean	301518
7967	Barnet	Year ending Dec 2017	Mean	667593
7968	Bexley	Year ending Dec 2017	Mean	357779
7969	Brent	Year ending Dec 2017	Mean	578705

Figure 7. London average property prices 2016-2017

Map generation

Nominatim was used to convert London into its geolocation coordinates (latitude and longitude) for the City of London. The Folium library was used to generate the map as shown in Figure 8. This map was used for various visualizations throughout the project.

```
In [99]: #Obtaining the values of latitude and longitude for London, United Kingdom
         address = 'London, UK'

         geolocator = Nominatim(user_agent = "london_explorer")
         location = geolocator.geocode(address)
         latitude = location.latitude
         longitude = location.longitude
         print('The geographical coordinates for City of London are {}, {}.'.format(latitude, longitude))

           The geographical coordinates for City of London are 51.5073219, -0.1276474.

In [100]: #Creating the map of London using folium and the values of latitude and longitude
          london_map = folium.Map( location = (latitude, longitude), zoom_start = 11)
          london_map

Out[100]:
```

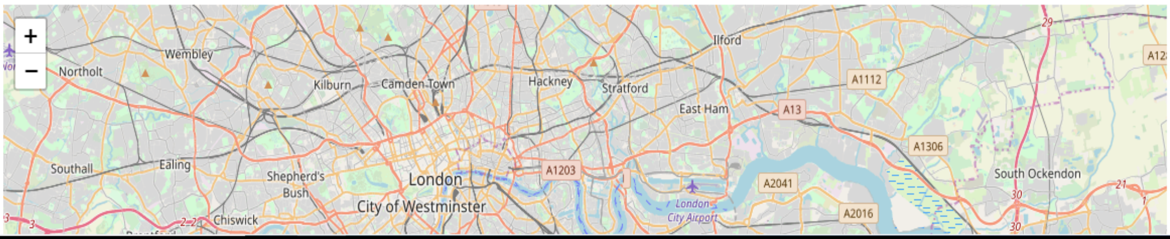

Figure 8. London map generated.

The London map was re-used to plot Airbnb listings on the map to make it interactive such that when the user clicks a circle marker, details of the neighbourhood and Borough pop-up as shown in Figure 9. However, due to limited computing power, only 500 listings have been added to the London map. With more computing power, all 65000 listings can be added.

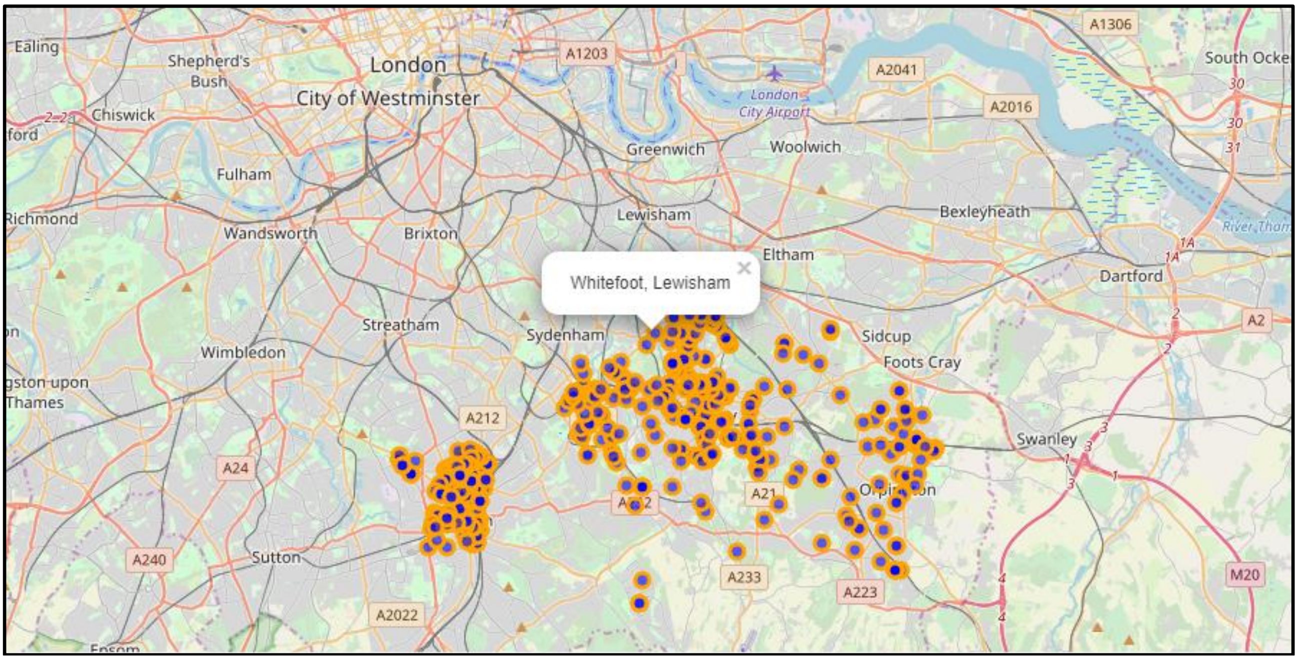

Figure 9. Airbnb listings on map.

A map in Figure 9 with 65000 circle markers on it would look very busy and messy. Also, there is no way of knowing whether the listing was a shared room, private room or entire property. A better way to improve the map visualization was to cluster the Airbnb listings into clusters by room type as shown on an interactive map in Figure 10. Again, due to limited computing power, only 2000 points were shown on the map.

Figure 10. Airbnb listing clusters by room type and location.

Exploring property type distribution

An investor would want to know what property types are mostly listed and in which neighbourhoods? There were over 40 property types and the most reasonable way to proceed was to pick the top 15 property types and then bundle the rest into one category as "Other" property type as shown in Figure 11.

```
In [80]: # Replacing other categories with 'other'
         df_merged_london2.loc[~df_merged_london2['Property Type'].isin(['House', 'Apartment', 'Townhouse', 'Condominium','Serviced apartment',
                                                                          'Loft','Bed and breakfast', 'Guest suite', 'Guesthouse','Boutique hotel',
                                                                          'Hostel','Bungalow','Hotel','Cottage','Aparthotel']),
                                              'Property Type'] = 'Other'
         #df_merged_london2.head()
         df_merged_london2['Property Type'].value_counts()

Out[80]: Apartment             36092
         House                 12157
         Townhouse              2030
         Condominium            1835
         Serviced apartment     1279
         Loft                    512
         Bed and breakfast       439
         Other                   299
         Guest suite             233
         Guesthouse              183
         Boutique hotel          130
         Hostel                  127
         Bungalow                 89
         Hotel                    68
         Cottage                  53
         Aparthotel               50
         Name: Property Type, dtype: int64
```

Figure 11. Property types grouping

The best approach was decided to be an interactive map with different coloured circles for each property type. Due to limited computing power, only top 7 property types and the 2500 points were plotted on the map. The user can select the points to show on the map by selecting/unselecting by property type as shown in Figure 12.

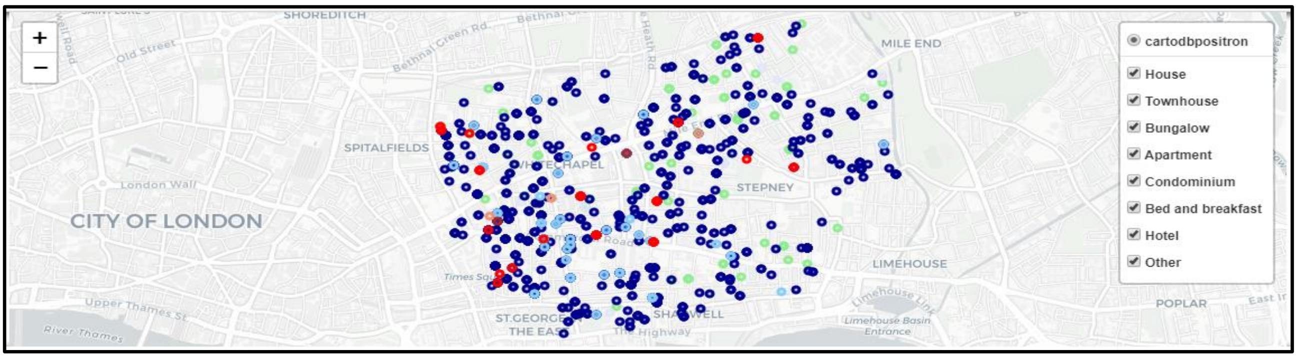

Figure 12. Top seven property types distribution

Proximity to certain amenities and trending venues

This is an important consideration when property investors, especially short-term rental, to consider for their targeted market demographics and psychographics. After identifying the top performing neighbourhoods, trending venues within a 200m radius of each Airbnb listings within each neighbourhood were extracted from www.foursquare.com via their API. For each neighbourhood, machine learning K-Means Clustering algorithm was used to identify top most common venues. Due to limited computing resources, only 4 neighbourhoods clusters were produced. With more resources, the elbow method would have been used to determine the optimal k-value.

ANALYSIS

Firstly, the price data was analysed to determine the price range (i.e maximum and minimum). The price range was too big starting from £0 – 20000 per night. A decision was made split the data by price ranges; £10-1000 and £1000-20000, plotted a histogram of listings frequency against price, in order to determine the most common price range and easily spot any outliers as show in Figures 13 and 14.

```
import matplotlib.pyplot as plt
# Distribution of prices from £10 to £1000
plt.figure(figsize=(20,4))
df_merged_london2.Price.hist(bins=100, range=(10,1000))
plt.margins(x=0)
#plt.axvline(200, color='orange', linestyle='--')
plt.title("Airbnb advertised prices per night in London from £10 to £1000", fontsize=16)
plt.xlabel("Price (£)")
plt.ylabel("Number of listings")
plt.show()
```

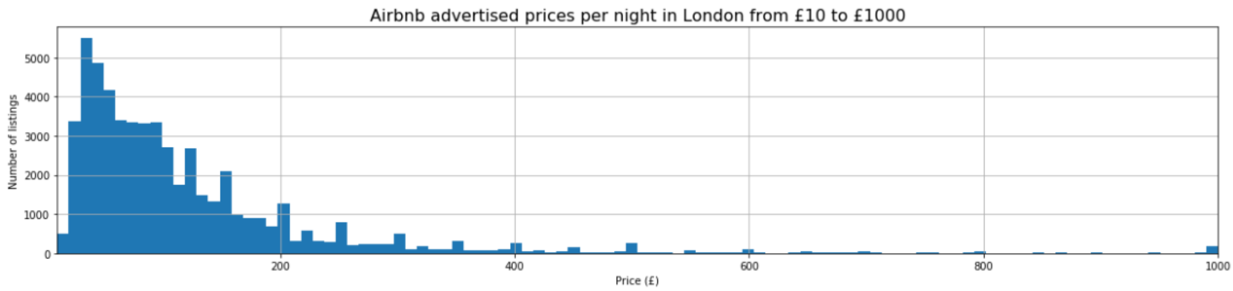

Figure 13. Histogram of Airbnb listings vs price per night (£10 - £1000)

```
# Distribution of prices from £1000 to £20000
plt.figure(figsize=(20,4))
df_merged_london2.Price.hist(bins=100, range=(1000,2000))
plt.margins(x=0)
#plt.axvline(200, color='orange', linestyle='--')
plt.title("Airbnb advertised prices per night in London from £1 000 to £20 000", fontsize=16)
plt.xlabel("Price (£)")
plt.ylabel("Number of listings")
plt.show()
```

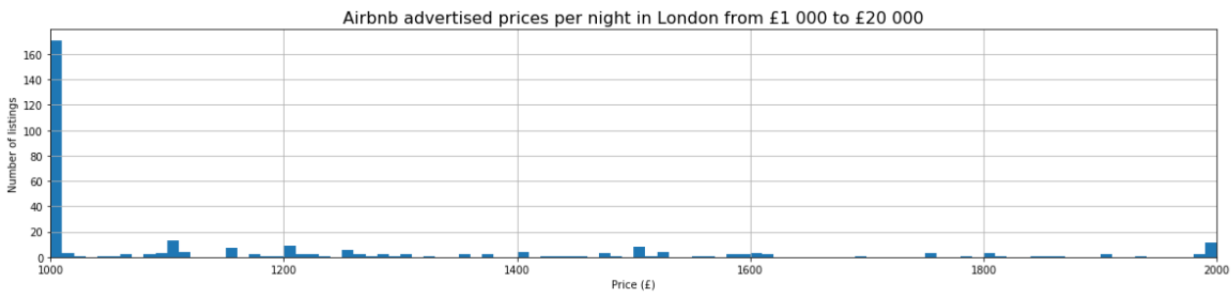

Figure 14. Histogram of Airbnb listings vs price per night (£1000-20000)

It was clear that the majority of prices per night were in the range of £10-200. All the data outside of that range was not considered for further analysis.

Question: What is the price per night distribution for London Boroughs and neighbourhoods?

The pyplot function within the matplotlib library was used to plot a barchart of Airbnb listing price per night against each Borough in Figure 15.

For neighbourhoods it was not very practical to easily visualise them on a barchart for example because there are 636 neighbourhoods in London. Therefore only top 30 neighbourhoods were plotted on a bar chart shown in Figure 16.

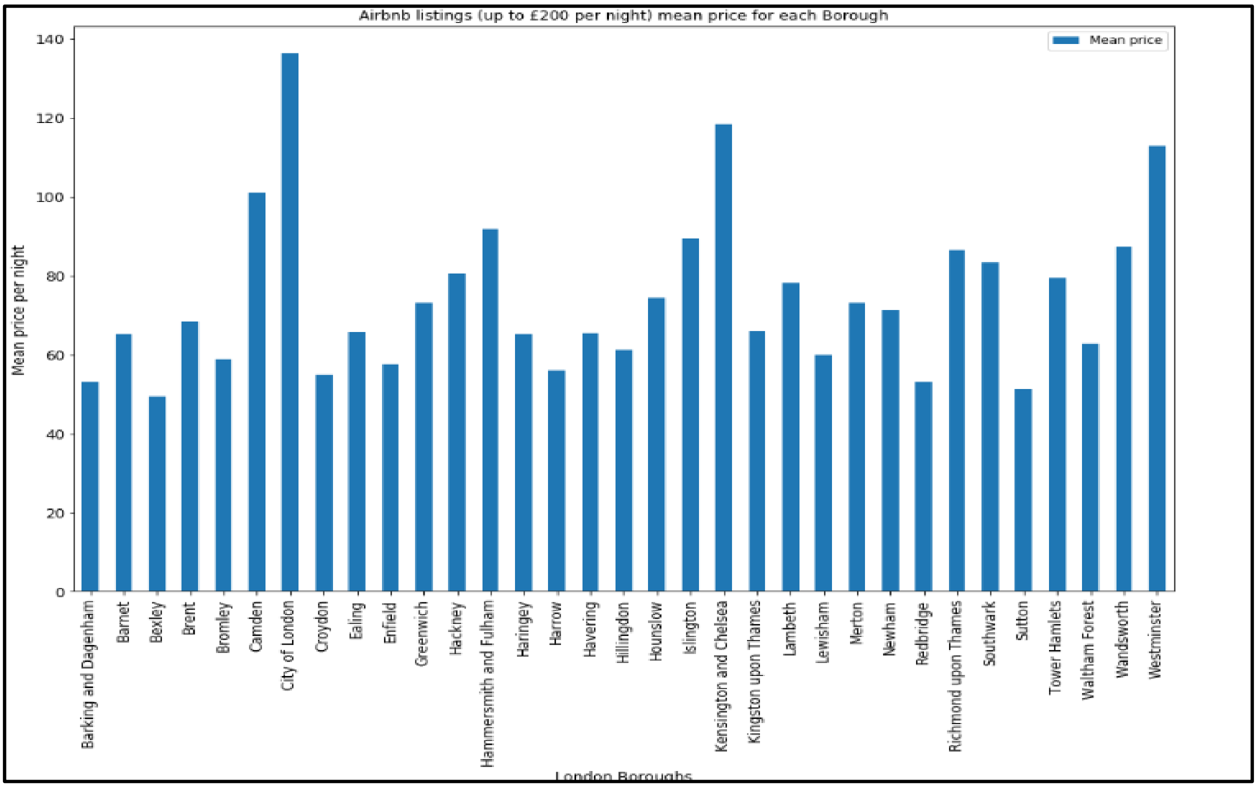

Figure 15. Airbnb mean price per night for each London Borough

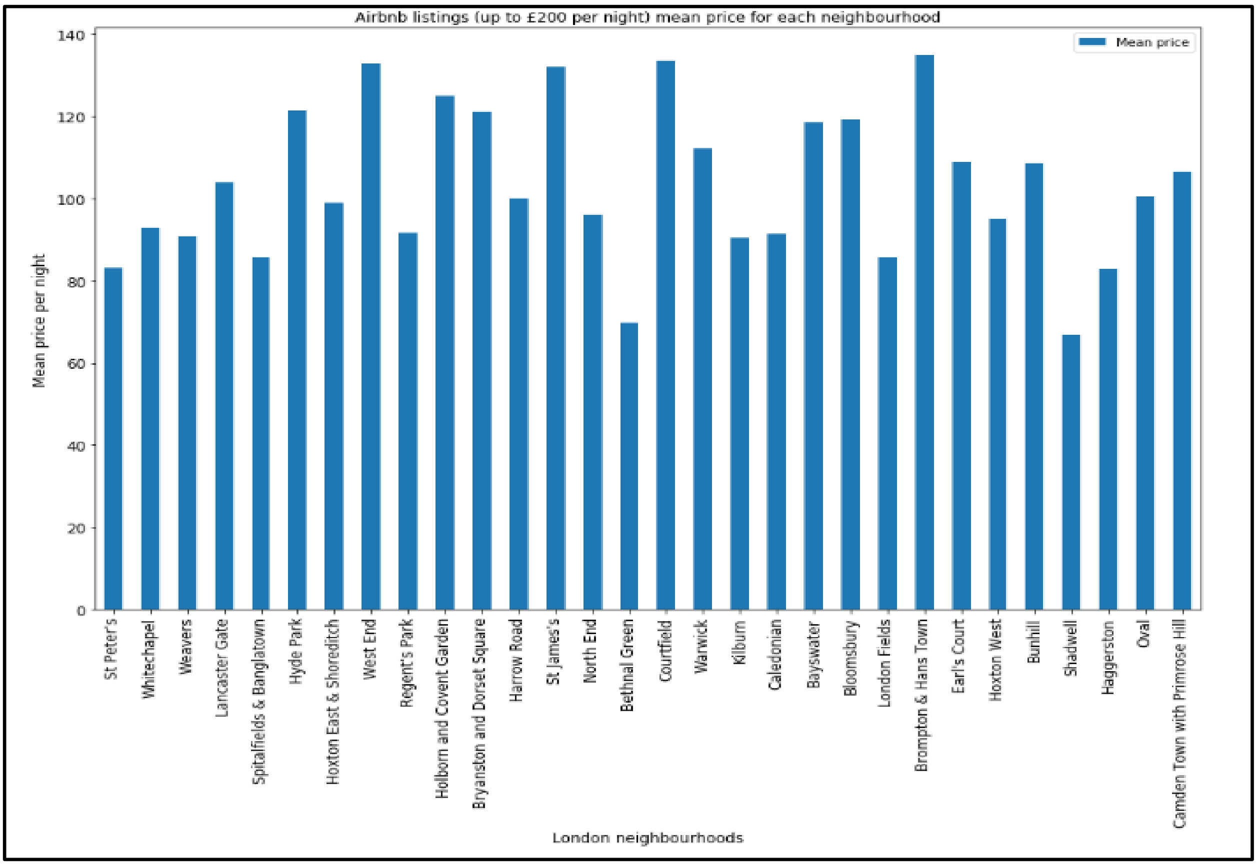

Figure 16. Airbnb listings mean price per night for top 30 London neighbourhoods.

Question: What is the average property purchase price in each London Borough?

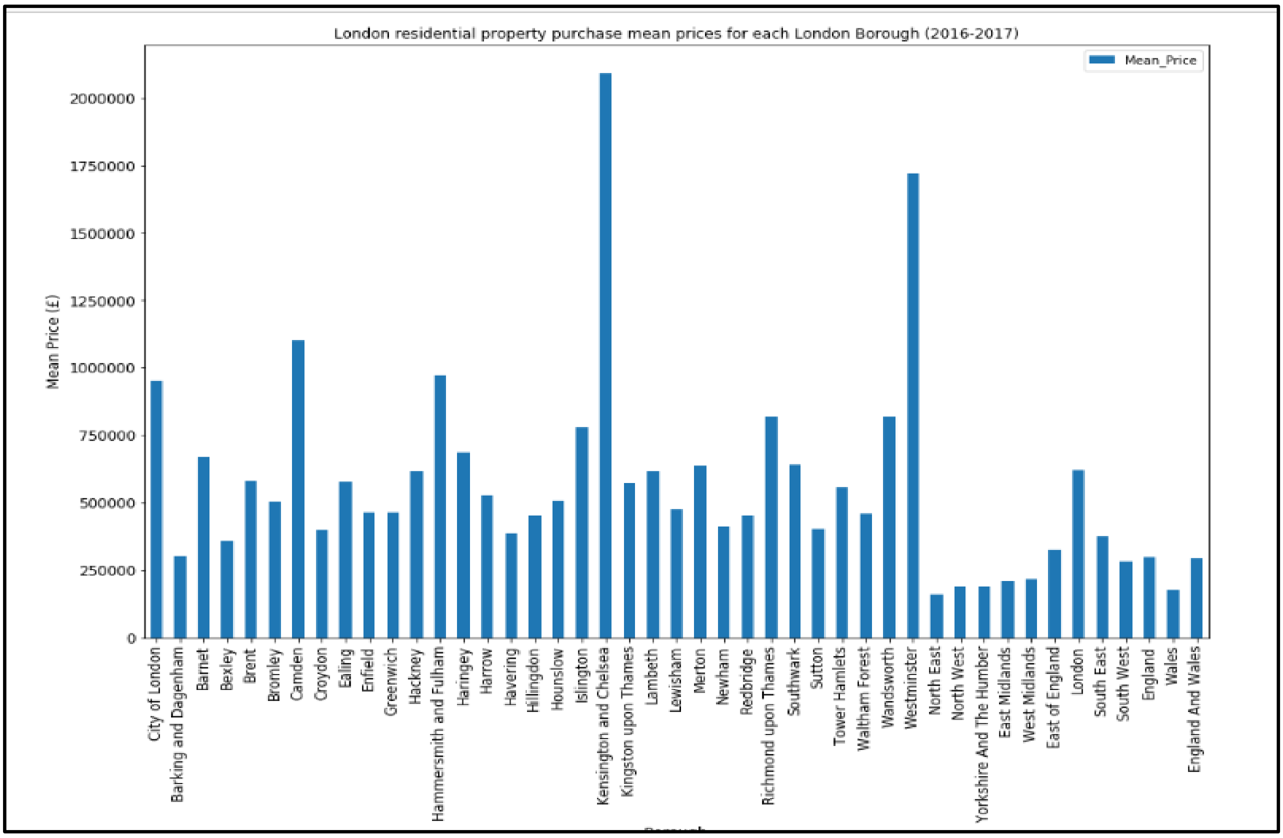

Figure 17. Average property purchase price in for each Borough

Question: What is the crime distribution for each London Borough?

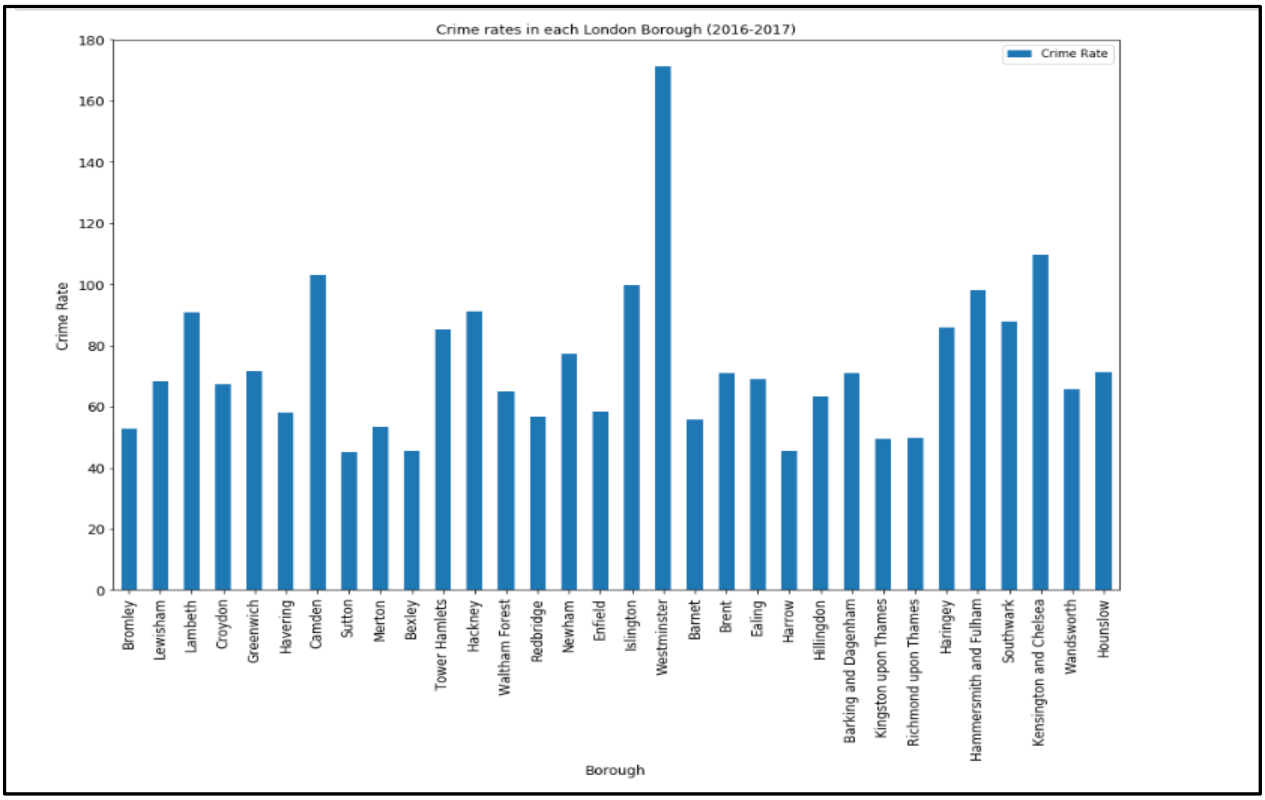

Figure 18. Crime rates in each London Borough

Question. Which London Boroughs and neighbourhoods have the highest Airbnb listings?

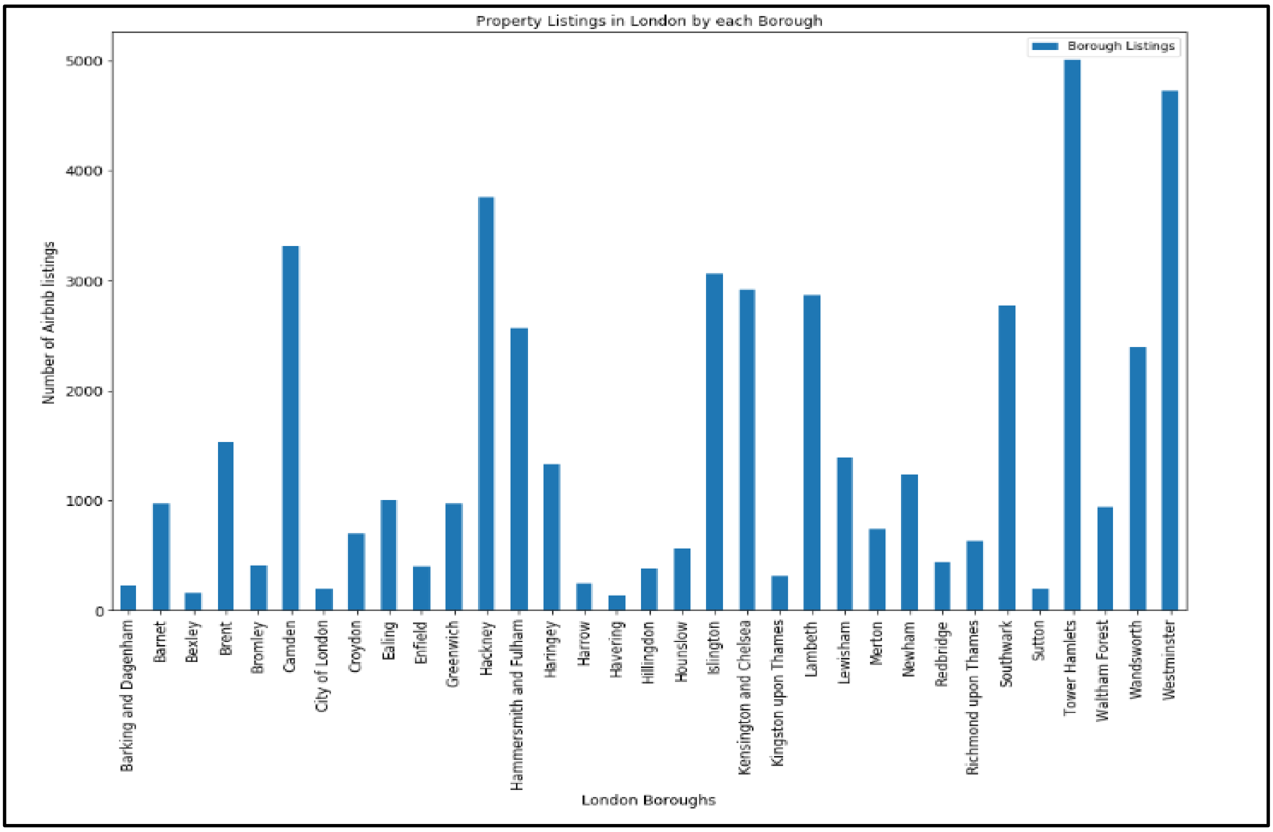

Figure 19. Airbnb listings for each London Borough

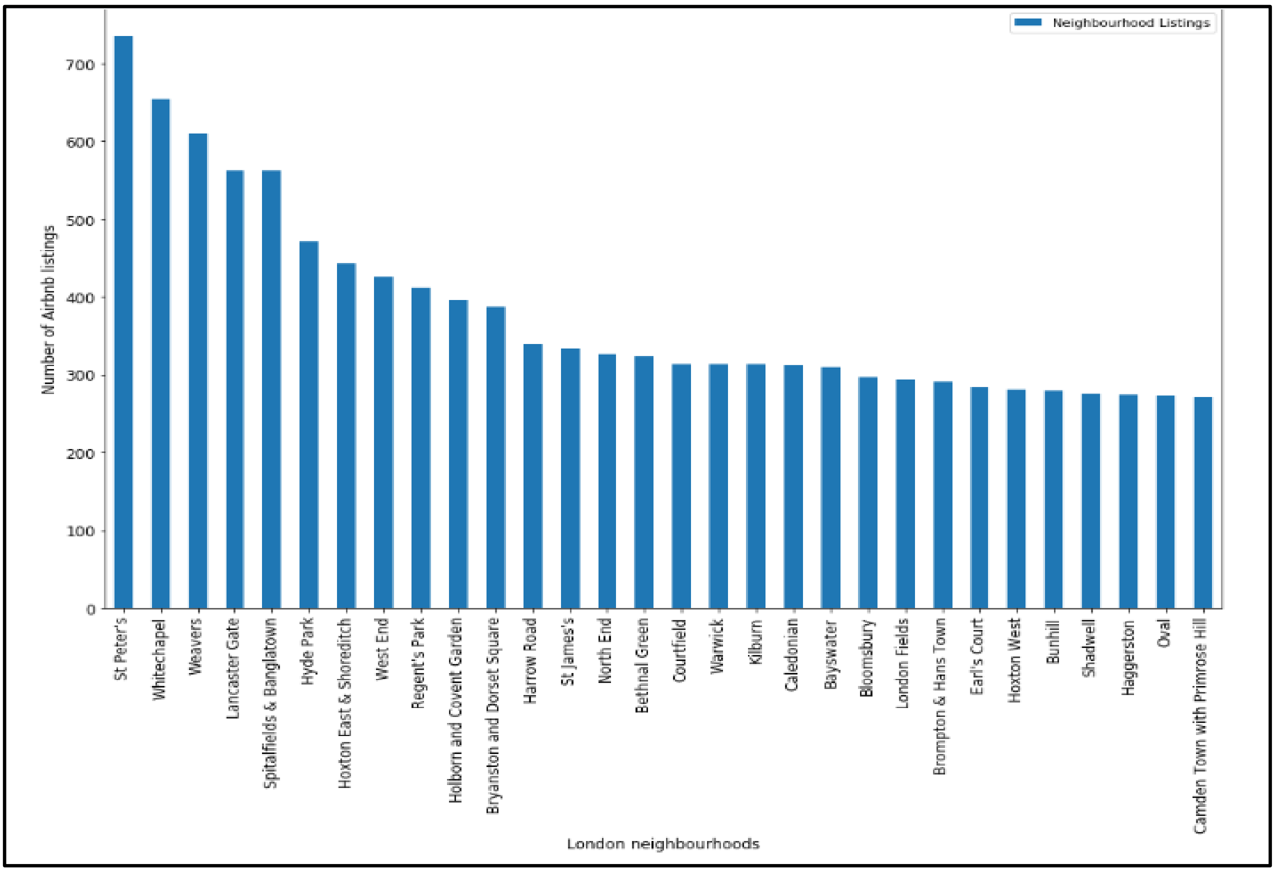

Figure 20. Airbnb listings for top 30 London neighbourhoods.

For neighbourhoods it was not very practical to easily visualise them on a bar chart for example because there are 636 neighbourhoods in London. Therefore, only top 30 neighbourhoods were visualised in a bar chart shown in Figure 20.

Question: What are most trending venues within a certain radius of each Airbnb listing for the top 4 neighbourhoods by number Airbnb listings.

One-hot encoding technique was used to analyse the data which was extracted from Foursquare.com via their API. There were more than 100 different categories of venues and it was decided to focus on the top 20 trending venue categories for each of the top 4 neighbourhoods with highest number of listings. Only 4 were selected due to limited computing resources.

Next step was to cluster the top 4 neighbourhoods using k-Means Clustering Algorithm.

RESULTS

From the results in Figure 15 it can be observed that the most expensive properties are in the Borough of Kensington and Chelsea followed by Westminster. Also, London generally has the most expensive properties compared to any other region within England and Wales.

For crime rates the top two were Westminster, Kensington and Chelsea while bottom two were Bexley and Havering.

The top 4 neighbourhoods with the highest Airbnb listings were St. Peter's, Whitechapel, Weavers and Lancaster gate and Spitalfields & Banglatown in joint fourth as shown in Figure 18.

Most common trending venues for the top 4 neighbourhoods are shown in Figure 21.

----Lancaster Gate----

	venue	freq
0	Hotel	0.23
1	Pub	0.13
2	Greek Restaurant	0.11
3	Café	0.08
4	Breakfast Spot	0.07
5	Garden	0.07
6	Indian Restaurant	0.05
7	Chinese Restaurant	0.05
8	Pharmacy	0.04
9	Coffee Shop	0.04
10	Sandwich Place	0.03
11	Kebab Restaurant	0.02
12	Plaza	0.02
13	Gym / Fitness Center	0.01
14	Restaurant	0.01
15	Malay Restaurant	0.01
16	French Restaurant	0.01
17	Burger Joint	0.01
18	Ice Cream Shop	0.01
19	Fountain	0.01

----St Peter's----

	venue	freq
0	Pub	0.19
1	Coffee Shop	0.10
2	Café	0.05
3	Cocktail Bar	0.04
4	Canal	0.04
5	Yoga Studio	0.04
6	Restaurant	0.03
7	Flower Shop	0.03
8	Art Gallery	0.03
9	Italian Restaurant	0.03
10	Bar	0.03
11	Pizza Place	0.02
12	Bakery	0.02
13	Park	0.02
14	Martial Arts Dojo	0.02
15	Brewery	0.02
16	Beer Bar	0.02
17	French Restaurant	0.02
18	Canal Lock	0.02
19	Platform	0.01

----Weavers----

	venue	freq
0	Café	0.09
1	Bagel Shop	0.09
2	Pub	0.09
3	BBQ Joint	0.08
4	Indie Movie Theater	0.08
5	Beer Bar	0.08
6	Furniture / Home Store	0.08
7	Coffee Shop	0.05
8	Turkish Restaurant	0.04
9	Italian Restaurant	0.04
10	Market	0.02
11	Restaurant	0.02
12	Cosmetics Shop	0.02
13	Salon / Barbershop	0.02
14	Indian Restaurant	0.02
15	Thrift / Vintage Store	0.02
16	Wine Bar	0.02
17	Garden	0.01
18	Hotel	0.01
19	Fried Chicken Joint	0.01

----Whitechapel----

	venue	freq
0	Hotel	0.11
1	Coffee Shop	0.08
2	Indian Restaurant	0.07
3	Middle Eastern Restaurant	0.06
4	Gym / Fitness Center	0.05
5	Sandwich Place	0.05
6	Italian Restaurant	0.05
7	Pizza Place	0.04
8	Korean Restaurant	0.04
9	Kebab Restaurant	0.04
10	Bar	0.04
11	Grocery Store	0.03
12	Movie Theater	0.03
13	Turkish Restaurant	0.03
14	Restaurant	0.03
15	North Indian Restaurant	0.02
16	Camera Store	0.02
17	Café	0.02
18	Park	0.02
19	Pub	0.02

Figure 21. Top 20 most common trending venues for the top 4 neighbourhoods by number of listings.

CONCLUSIONS

This project focused on which are the best locations to invest in property for short-term rentals in London. Exploration and analysis determined which neighbourhoods and Boroughs have the highest number of Airbnb listings, types of properties listed, price per night distribution, average property purchase price, crime rates distribution. The top 4 neighbourhoods were then explored for the most trending venues and then divided into clusters using k-Means Clustering.

St. Peter’s, Whitechapel, Weavers and Lancaster gate and Spitalfields & Banglatown had the highest listings. Most expensive properties are in the Boroughs of Kensington and Chelsea and Westminster. The same two Boroughs also topped the highest crime rates.
Most common trending venues near the Airbnb listing location clusters were pubs and bars, coffee shops and cafés, restaurants (different types), fitness and wellbeing venues (different types) and shops.

From the results of this project, it can be concluded that a buy-to-let or short-term rental property investor would better select a location where property purchase price is relatively low or average, average crime rates, high short-term rental listings (although increased competition but maintains good occupancy rates due to high demand), good proximity to amenities (such as shops, pubs, bars, restaurants, fitness and wellness centres, coffee shops and cafes). The recommended price range per night in London is up to £200.

RECOMMENDATIONS

It is recommended to include property rental data from data driven websites such as zoopla.co.uk to further reinforce the deal analysis for the property investors. In addition, to this, further granular analysis on performance of each property type within the same neighbourhood because an overall market may be generally profitable, some property types could be far more successful than others. Also, events and seasons are known to impact the short-term rental markets. Analysis these factors would improve the project. Additional data to calculate operating expenses, occupancy rates and non-sector specific data such as employment rates, tourist attractions, local business economy would also greatly assist property investors to make smart and profitable decisions.