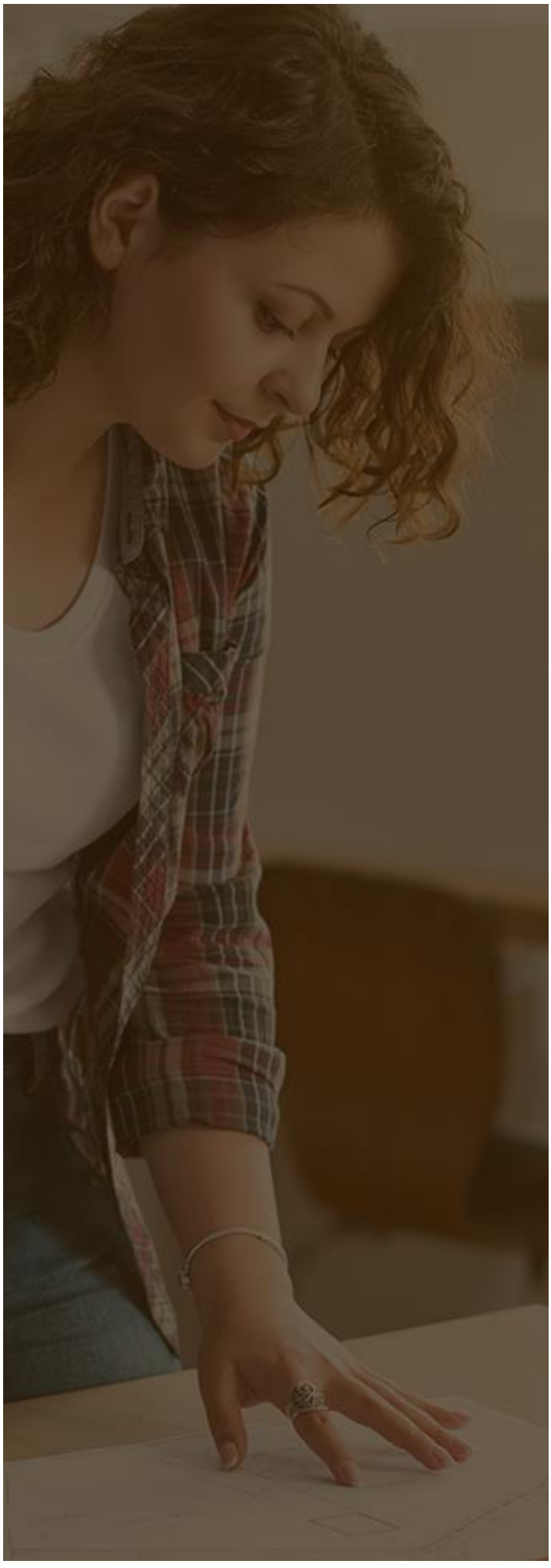
# Battle of the Neighborhoods

## Relocation to London

*IBM Data Science Capstone project report*



# TABLE OF CONTENTS

# INTRODUCTION

This is report will be addressing the requirement of IBM Data Science provided in partnership with Coursera

The report will consist of the following:

1. Introduction (executive summary) where you discuss the business problem and who would be interested in this project.
2. Data where you describe the data that will be used to solve the problem and the source of the data.
3. Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
4. Results section where you discuss the results.
5. Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
6. Conclusion section where you conclude the report.

# 1. EXECUTIVE SUMMARY (INTRODUCTION)

**My friends Daughter is planning to study her Masters in London thus he requested my help to identify the best place to live in London in term of having acceptable rent and a lot of venues to cover all his kid needs.**

**Business Problem:**

**Finding the best area for international student to live in London keeping the following in mind:**

1. **Since my friend's daughter has never been in London nor actually lived alone the target location must include a good variety of venues e.g. Restaurant, super markets …etc**
2. **As an international student the fees for masters and living abroad is very expensive thus must find a location with an acceptable rent**

# 2. DATA SOURCES

When looking at the allocated time and research budget available best approach was to get data from free dependable data provider.

Data Sources and description:

1- Venues data: data is needed to find venues per area for London city in addition to the rating of each venue to be able to pinpoint the best location.
The best available data source with enough data and at no cost was getting the data through Foursqaure.com.

**https://foursquare.com**

2- Spatial data (Latitude and Longitude) for all London locations: this data was needed to be able to get the needed Foursquare data and conduct the needed analysis

**https://data.london.gov.uk**

3- Rent Data per Location: Rent data per each location.

**https://data.london.gov.uk**

4- General source of data : general data and research was conducted through the research using google.com

**https://google.com**

# 3. METHODOLOGY

**Analysis Tools:**

Python jupyter notebook will be the main tool used for data preparation and analysis using the following libraries and its modules:

1- Panda library for data analysis
2- Numpy library to handle data in a vectorized manner
3- Requests library to handle requests
4- Random library for random number generation
5- Geopy Library to convert an address into latitude and longitude values
6- Folium library for plotting maps
7- Sklearn library for clustering
8- Matplotlib for plotting

**Analysis:**

Following are the analysis steps that led to the conclusion:

1) Understanding the problem:
   a) Approach:
      i) Interview my friend and understand his requirements and concerns
      ii) Talk to his Daughter to assess her character, her skills and her outlook on the journey to come
   b) Findings:
      i) Monthly Budget for rent = 1,000 GBP
      ii) Target University : University of London Goldsmiths College - 8 Lewisham Way, New Cross, London SE14 6NW, UK
2) Getting the data:
   a) Data was acquired through the following ways
      i) Direct scarping from Google to get the list of London locations (extract below):

Out[3]:

| | Neighborhood |
|---|---|
| 0 | Chelsea |
| 1 | Camden Town |
| 2 | Westminster |
| 3 | Soho |
| 4 | Canary Wharf |
| 5 | Covent Garden |

ii) Rent and geospatial data was read directly for London Government website reading the CSV file through API provided links and getting the latitude/ Longitude(extract below):

| | | | |
|---|---|---|---|
| 0 | Chelsea | 51.487542 | -0.168220 |
| 1 | Camden Town | 51.542305 | -0.139560 |
| 2 | Westminster | 51.500444 | -0.126540 |
| 3 | Soho | 51.513163 | -0.131175 |
| 4 | Canary Wharf | 51.504895 | -0.019001 |

iii) Venues information were acquired through direct reading of data from 4Square.com using the developer option

3) **Data Preparation:**
   a) In general data acquired was clean and ready for analysis – some dropping of columns and column name changes were done using python note book.
   b) Latitude and Longitude data were acquired by mapping the locations from Google using Geolocator
4) **Visualization:**
   a) **Venues Maps**
      i) Getting Venues data from Foursquare based on Latitude and Longitude table information
      ii) OneHot Encoding and most common venues

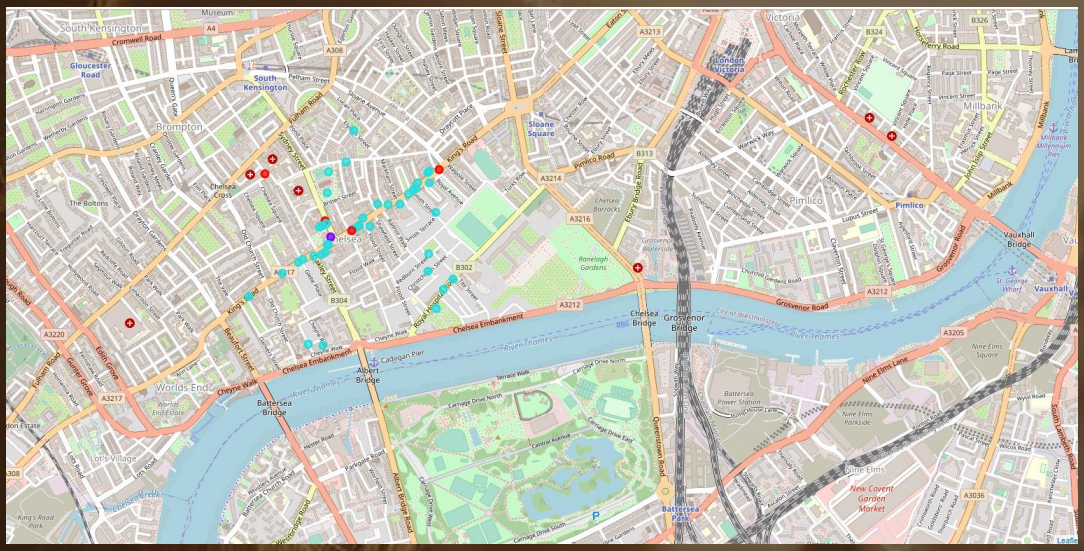
98]:

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | ... | 41th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Balham | Watch Shop | Yoga Studio | Farmers Market | Dumpling Restaurant | Electronics Store | English Restaurant | Ethiopian Restaurant | Event Space | Exhibit | ... | Gym / Fitness Center |
| 1 | Belgravia | Hotel | Café | Boutique | Gastropub | Park | Clothing Store | Lounge | Tea Room | Pakistani Restaurant | ... | Design Studio |
| 2 | Bermondsey | Pub | Grocery Store | Farmers Market | Coffee Shop | Indonesian Restaurant | Trail | Bus Stop | Brewery | Café | ... | Donut Shop |
| 3 | Bethnal Green | Coffee Shop | Pub | Cocktail Bar | Hotel | Café | Flower Shop | Pizza Place | Wine Bar | Breakfast Spot | ... | Dog Run |
| 4 | Blackheath | Pub | Bakery | Café | Pizza Place | Bar | Greek Restaurant | Gastropub | French Restaurant | Himalayan Restaurant | ... | Dog Run |
| 5 | Bristol | Bar | Café | Pub | Coffee Shop | Italian Restaurant | Burger Joint | Indian Restaurant | Asian Restaurant | Steakhouse | ... | Ice Cream Shop |
| 6 | Brixton | Coffee Shop | Pub | Pizza Place | Social Club | Caribbean Restaurant | Cocktail Bar | Convenience Store | Plaza | Park | ... | Dive Bar |
| 7 | Camberwell | Café | Pub | Bus Stop | Grocery Store | Coffee Shop | Gym / Fitness Center | Gastropub | Bakery | Park | ... | Cricket Ground |

iii) Using KMeans for the clustering of the neighborhoods based on Foursquare table

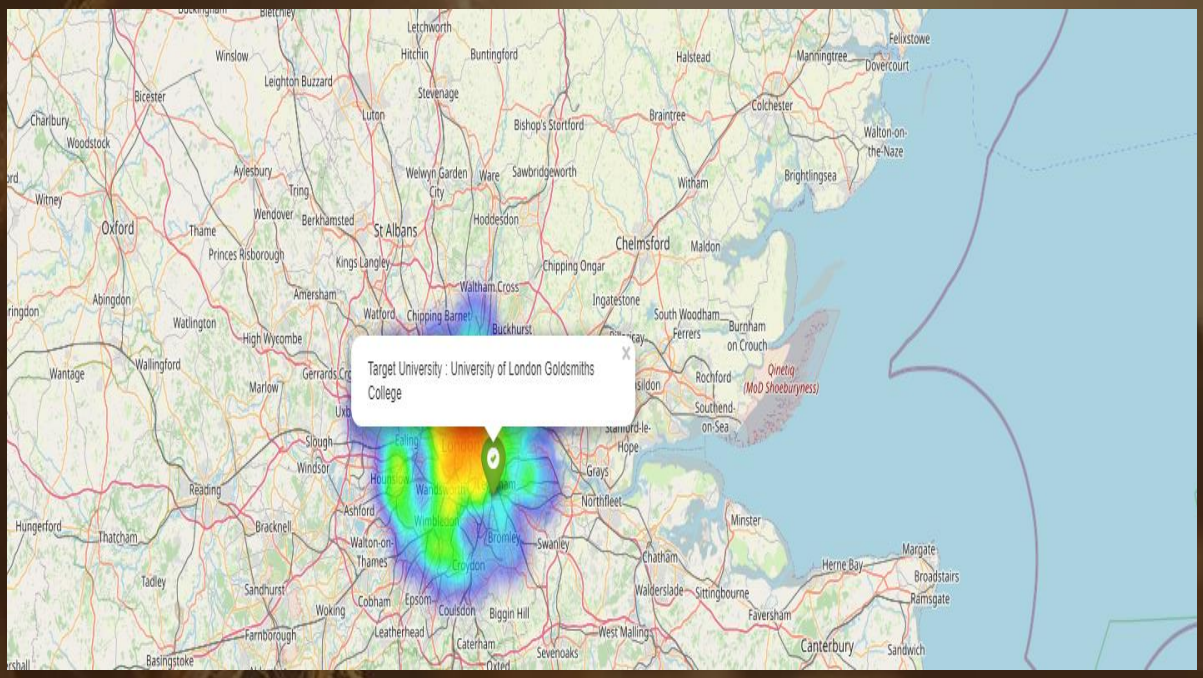
| | Neighborhood | Latitude | Longitude | Cluster Labels | Common Venue | Common Venue | Common Venue | Common Venue | Common Venue | Common Venue | ... | Common Venue | Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Chelsea | 51.487542 | -0.168220 | 0.0 | Bakery | Pub | French Restaurant | Japanese Restaurant | Park | English Restaurant | ... | Ethiopian Restaurant | Dog Run |
| 1 | Camden Town | 51.542305 | -0.139560 | 0.0 | Pub | Coffee Shop | Burger Joint | Café | Ice Cream Shop | Italian Restaurant | ... | Fish & Chips Shop | Middle Eastern Restaurant |
| 2 | Westminster | 51.500444 | -0.126540 | 0.0 | Coffee Shop | Pub | Outdoor Sculpture | Café | Historic Site | Garden | ... | Dog Run | Dive Bar |
| 3 | Soho | 51.513163 | -0.131175 | 0.0 | Coffee Shop | Bakery | Theater | Ice Cream Shop | Japanese Restaurant | Plaza | ... | Mediterranean Restaurant | English Restaurant |
| 4 | Canary Wharf | 51.504895 | -0.019001 | 0.0 | Coffee Shop | Burger Joint | Sandwich Place | Shopping Mall | Bakery | Steakhouse | ... | Multiplex | Fish Marke |
| 5 | Covent Garden | 51.512874 | -0.122544 | 0.0 | Theater | Coffee Shop | Burger Joint | Bakery | Hotel | Restaurant | ... | Deli / Bodega | Spanish Restaurant |

iv) Create Maps based on the clusters




v) Getting the clusters size:
Cluster 0 : 4 Rows
Cluster 1 : 1 Rows
Cluster 2 : 49 Rows
Cluster 3 : 1 Row

b) **Rent Maps:**
i) Read and upload Longitude and Latitude data
ii) Read and upload Rent data
iii) Merge both maps based on the mutual identifier of areas
iv) Create Heat Map using Folium:


Target University : University of London Goldsmiths College
Letchworth
Hitchin
Buntingford
Halstead
Manningtree
Felixstowe
Dovercourt
Winslow
Leighton Buzzard
Bicester
Stevenage
Luton
Bishop's Stortford
Braintree
Colchester
Walton-on-the-Naze
Charlbury
Woodstock
Aylesbury
Tring
Welwyn Garden City
Ware
Sawbridgeworth
Witham
Brightlingsea
Witney
Oxford
Thame
Wendover
Berkhamsted
St Albans
Hoddesdon
Chelmsford
Maldon
Princes Risborough
Kings Langley
Chipping Ongar
Abingdon
Amersham
Waltham Cross
Ingatestone
Watford
Chipping Barnet
South Woodham Ferrers
Burnham on Crouch
Watlington
High Wycombe
Buckhurst
Wantage
Wallingford
Gerrards Cross
Marlow
Rochford
Southend-on-Sea
Slough
Windsor
Reading
Grays
Northfleet
Ashford
Hungerford
Thatcham
Bracknell
Walton-on-Thames
Bromley
Swanley
Chatham
Herne Bay
Margate
Broadstairs
Ramsgate
Tadley
Sandhurst
Woking
Cobham
Epsom
Coulsdon
Biggin Hill
Walderslade
Sittingbourne
Faversham
Farnborough
Leatherhead
Caterham
West Malling
Sevenoaks
Canterbury
Sandwich
Basingstoke
Oxted

# 4. RESULTS

Based on the mapping created for both the venues and the rent following are the results:

1- Looking at London venues maps and clustering results it obvious that downtown London (Chelsea area) is were all top-rated venues are located
2- Looking at the heat Map downtown London(Chelsea area) is where the highest rent is paid per meter
3- Looking at the location of Goldsmith college we will find that it is far from downtown London

In summary the first conclusion was to suggest Down Town London as the ideal location but based on Rent data/Distance to college this is not the case as if we weigh the importance of Venues vs available budget for sure Budget will have a bigger weight.

Since the college is in Lewis ham and rent is relatively acceptable (953 vs Budget of 1,000) in that area the final conclusion is to live there.

# 5. CONCLUSION

Based on the results that came out of mapping of data it is concluded that living in the outskirts of London is the best option as it still part of the top cluster in addition rent as per the heat map is acceptable.