Capstone Project

MOVING TO LONDON – BATTLE OF THE NEIGHBORHOOD

APPLIED DATA SCIENCE CAPSTONE

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

Table of Contents

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:
- 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
- As an international student the fees for masters and living abroad is very expensive thus must find a location with an acceptable rent

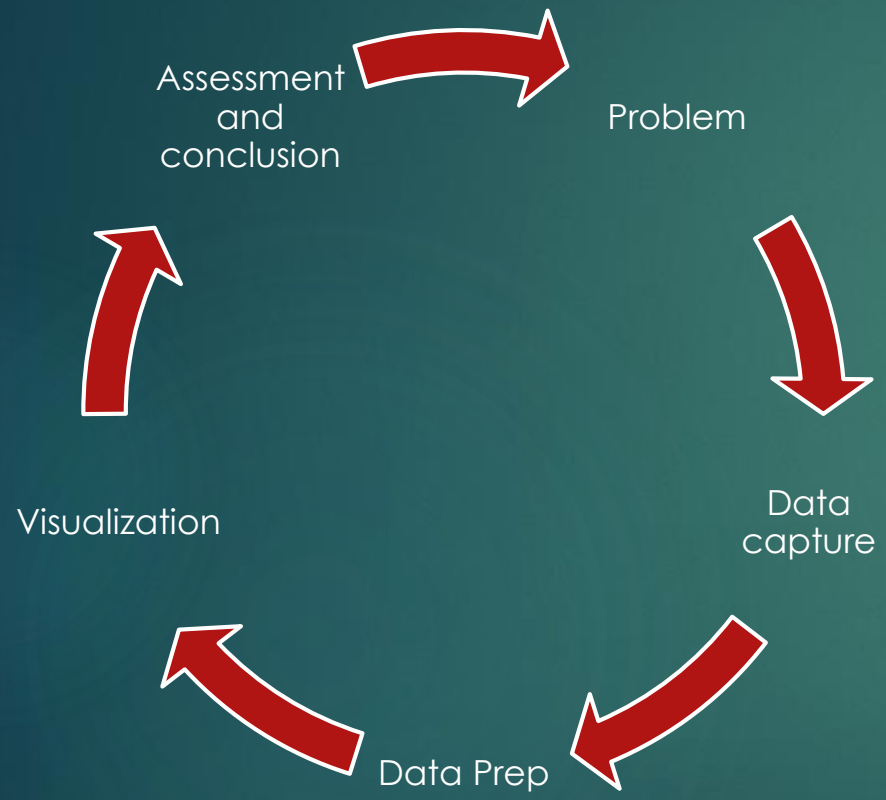
DATA SOURCES

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
1. 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
1. Rent Data per Location: Rent data per each location.
   https://data.london.gov.uk
1. General source of data : general data and research was conducted through the research using google.com
   https://google.com

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

Methodology (Problem)

Business Problem

1) Understanding the problem:

a) Approach:

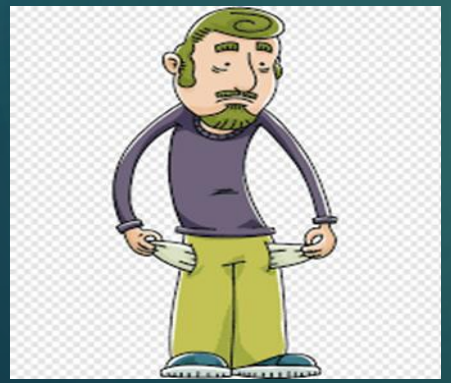
i) Interview my friend and understand his requirements and concerns

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



Methodology (Data Capture)

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):

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):

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

Methodology (Data Preparation)

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

Methodology (Visualization) 1 of 2

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

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

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:

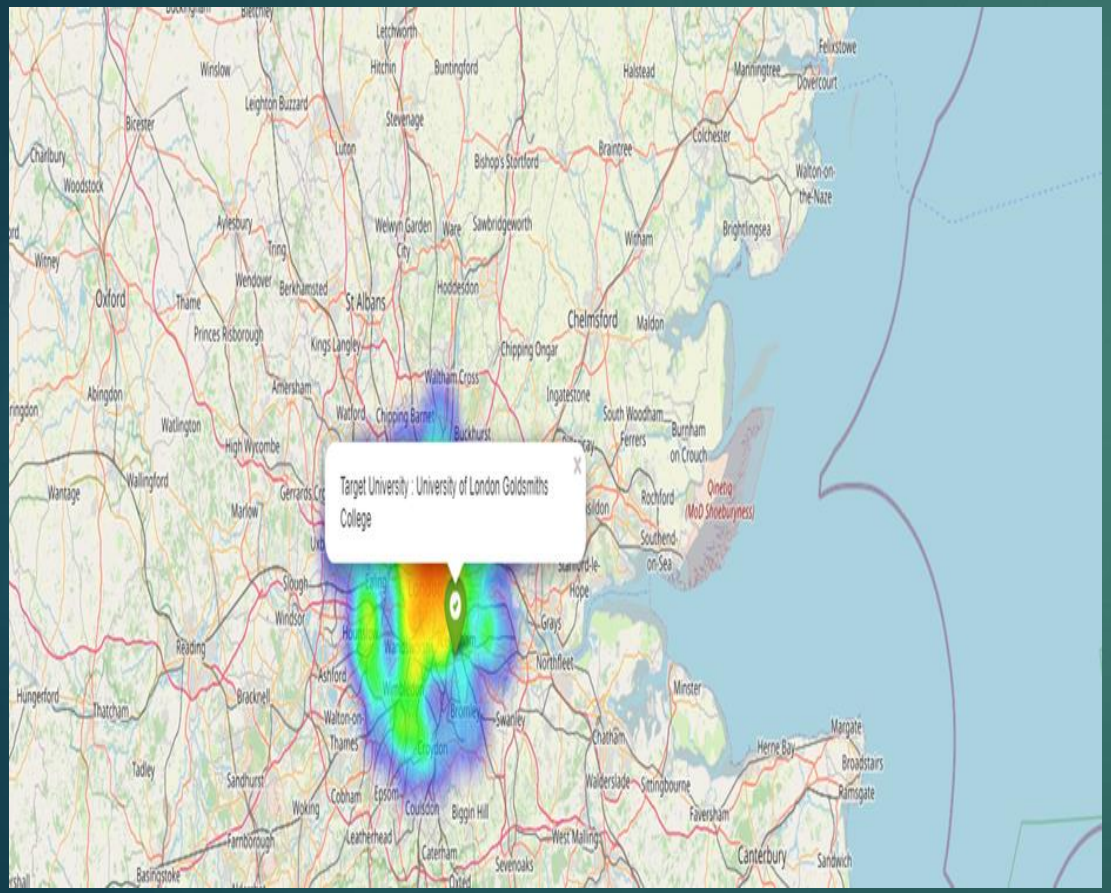
i) Read and upload Longitude and Latitude data

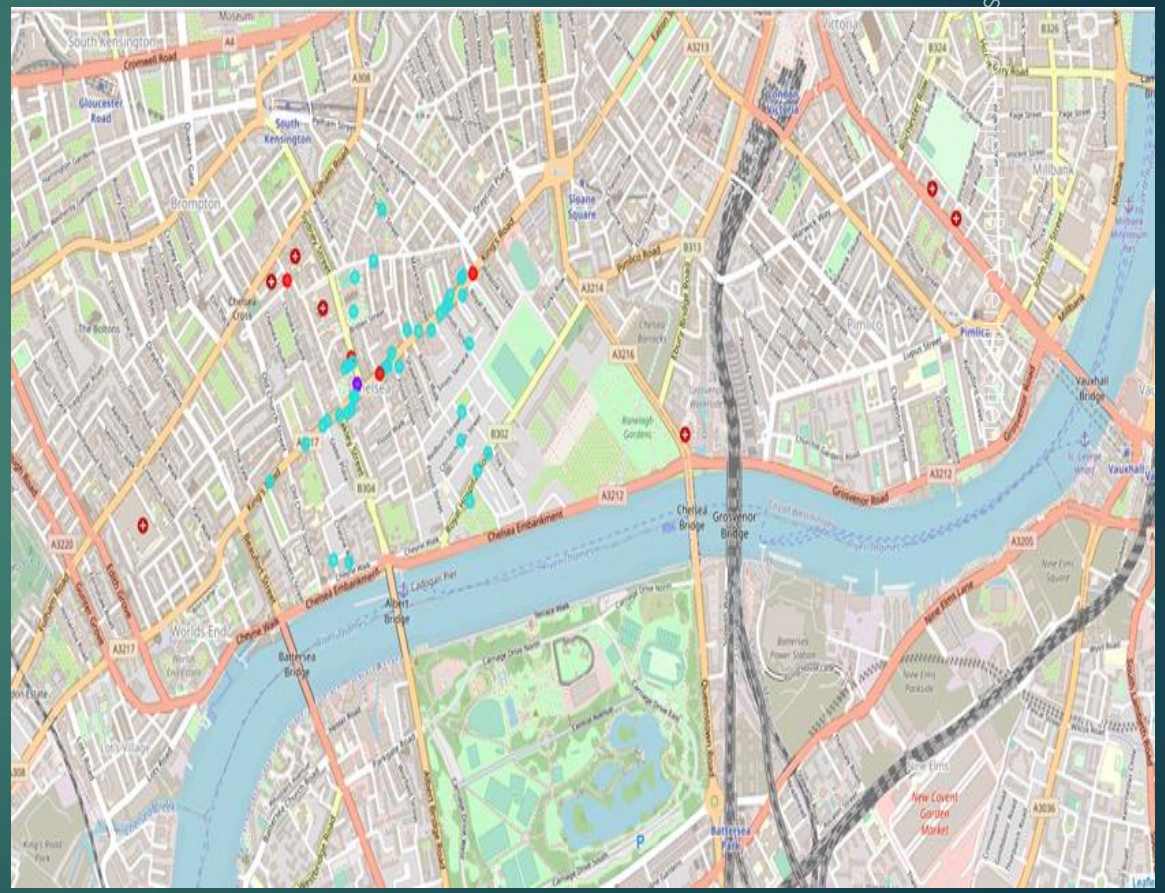
ii) Read and upload Rent data

iii) Merge both maps based on the mutual identifier of areas

iv) Create Heat Map using Folium

Methodology (Visualization) 2 of 2





Methodology (Assessment)

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.

Methodology (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.