CAPSTONE PROJECT

Open A Café In London

Project Introduction

In modern society, coffee shop plays an important role in people's lives. It has became a popular place at where people choose to chat, meet friends and get relaxed. In this project, we are interested in opening a musical theme cafe in London. We would collect the data through "geopy" library in Python and the Foursquare API and conduct our analysis through machine learning (K-Means clustering method).

Contents

PART ONE

Background

• As the capital of the Great Britain, there are more than six million people live in London.

• Attracting more than 17 million tourists every year makes London one of the most popular cities in the world.

• The average growth rate of coffee shop industry in the UK is 6.1% from 2014 to 2019.

• IBIS World's report expected that the average annual industry revenue raising rate will be 4.8% from 2019 to 2024.

• Alongside Broadway in New York, London's West End theatre represents the highest quality theatre in the world.

• There are more than 40 theatres crowded within a less than one square mile of land called "Theatreland".

PART TWO

Methodology

In this project we will detect areas within the **2km radius** circle from the West End centre (Leicester Square) that have low cafe density.

- Step one: Collect the required data: location of every cafe within 2km from West End centre.
- Step two: Calculate and explore cafe density across different areas of West End.

  Heatmaps will be used to identify a few promising areas close to centre with low number of cafes in general and focus our attention on those areas.
- Step three: Focus on the most promising areas found above, we will find locations with **no cafe in radius of 150 metres**. Clusters will be created for these good locations by **K-Means** clustering method. Cluster centres will be our final recommendation location which should be a nice starting point for further exploration and research.

PART THREE

Data Collection

Factors

Based on definition of our problem, factors that will influence our decision are:

- Number of existing cafes in the neighbourhood
- Distance of neighbourhood from West End centre (theatre centre)

We decided to use **regularly spaced grid of locations**, cantered around city centre, to define our neighbourhoods.

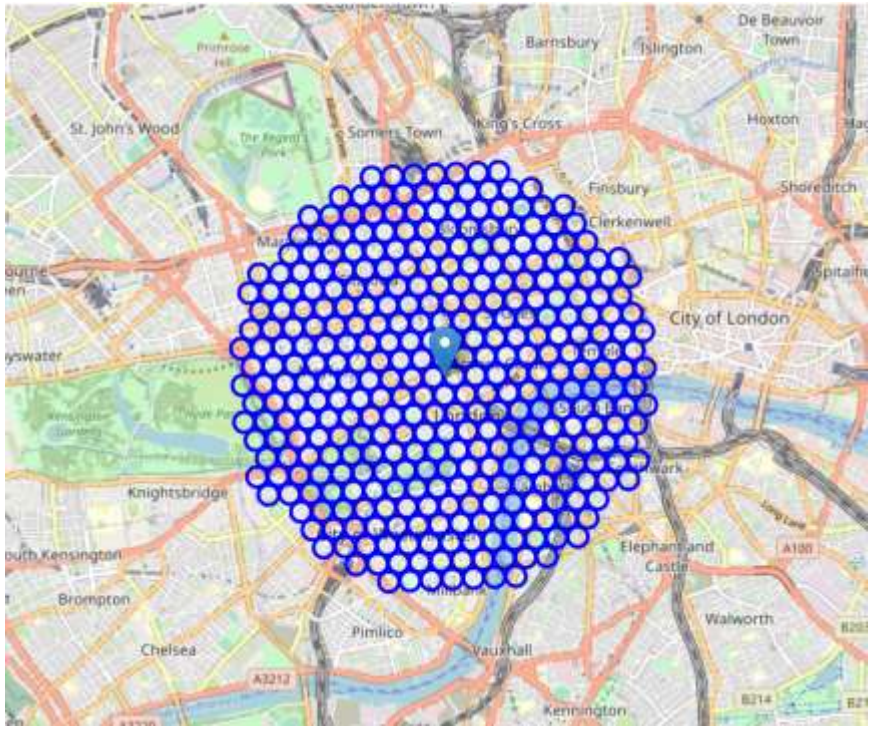
Data Demand and Source

- Coordinate of West End centre will be obtained using **geolocator. Geocode** of well known London location (**Leicester Square**)
- Centres of candidate areas will be generated algorithmically and approximate addresses of centres of those areas will be obtained using **geolocator. Geocode.**
- Number of cafes and location in every neighbourhood will be obtained using **Foursquare API**

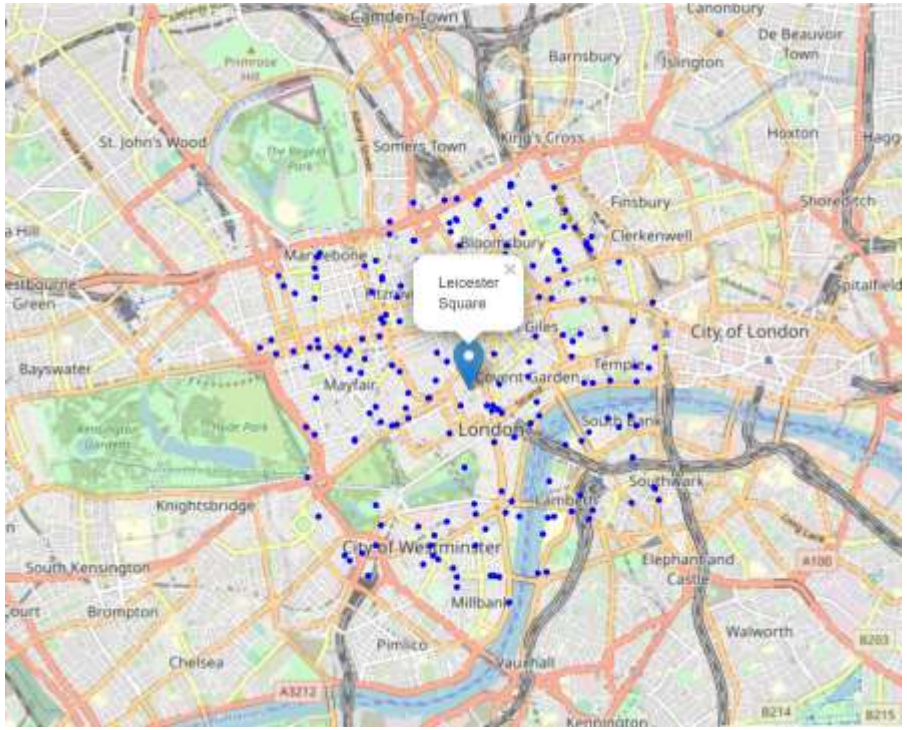
Data Describe

- **Leicester Square's coordinates:** Latitude, Longitude = 51.511, -0.131 (WGS 84) X,Y = 282753.38803952804, 5711255.921690965 (UTM)
- Within **2km** radius circle from Leicester Square we built **364** candidate cells which are **200** metres away from all there neighbours. (Hexagonal grid of cells)
- **198** cafes were found (Map of cafes within 2km radius)

Hexagonal grid of cells



Map of cafes within 2km radius



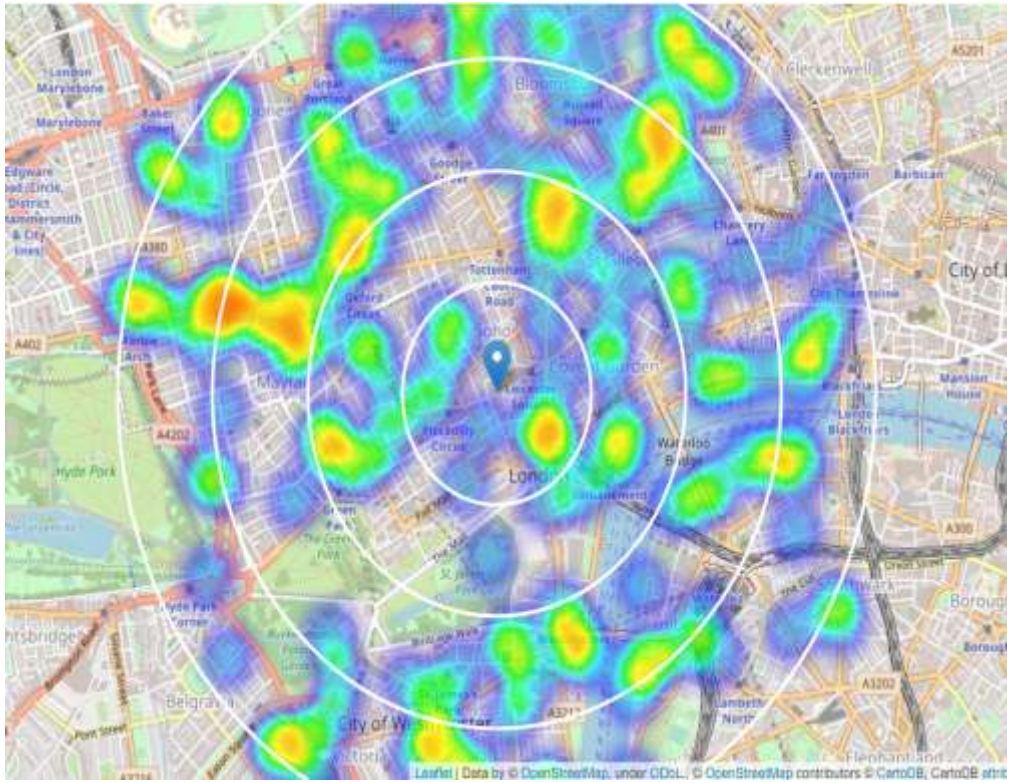
PART FOUR

Analysis

General Area Density Analysis

In our 2km area:

- On average 0.42 cafes in every area with radius 100 meters.
- Average distance to closest café from each area center is **138.25** meters.
- According to the general heatmap low café density area can be found in the **north, south, south-west and east** from Leicester Square (in the 1km circle).



Narrower Area Density Analysis

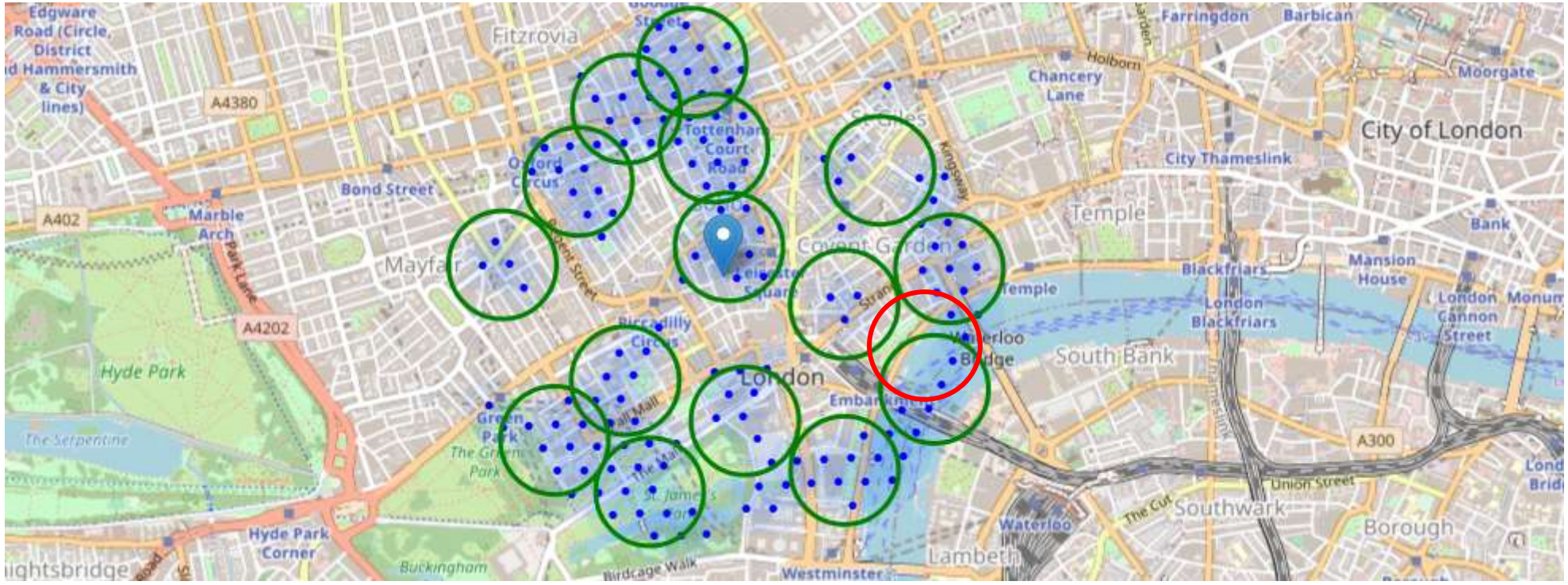
In the 1km circle we create 364 candidate neighborhoods which are 100m apart from their neighbors:

- There are **150 points** meet the condition of no café in radius of 150 meters.
- Most of them are in the City of Westminster and some of them are located in the borough of Camden.
- The City of Westminster and Camden boroughs are popular to not just tourists but also Londoners.

K-Means Clustering Analysis

• 15 clusters are generated for the 150 points above. These clusters represent most of the candidate locations and cluster centers are nicely in the middle of the zone “rich” with location candidates.

• It is worth noticing that one cluster (red circle) is in the river therefore it should not be in the final results.



PART FIVE

Results

- The final results are addresses for all 14 cluster centres except the one in the river (marked in grey):

PART SIX

Conclusion and Further Directions

• We found 14 possible locations with no café in 150 meters around within the 1km radius circle from Leicester Square.

• Further analysis should be based on specific characterizes of neighborhoods and locations in every recommended zone.

• Factors such as: levels of noise, distance to a particular theater, real estate availability, prices, economic dynamic of each neighborhood etc. should be taken into further consideration.

Thanks For Watching!

Yuxin Ding