

WHERE TO PLACE TOURIST INFORMATION CENTERS IN LONDON

Reynard Jusuf

IBM Data Science Capstone, Coursera

October 2019

Business Problem

A new travel company want to open multiple tourist information offices in London. This data analysis would use Foursquare location data to determine the best places to put them in. We will determine areas where there is a high concentration of tourist attractions and cluster them.

Data

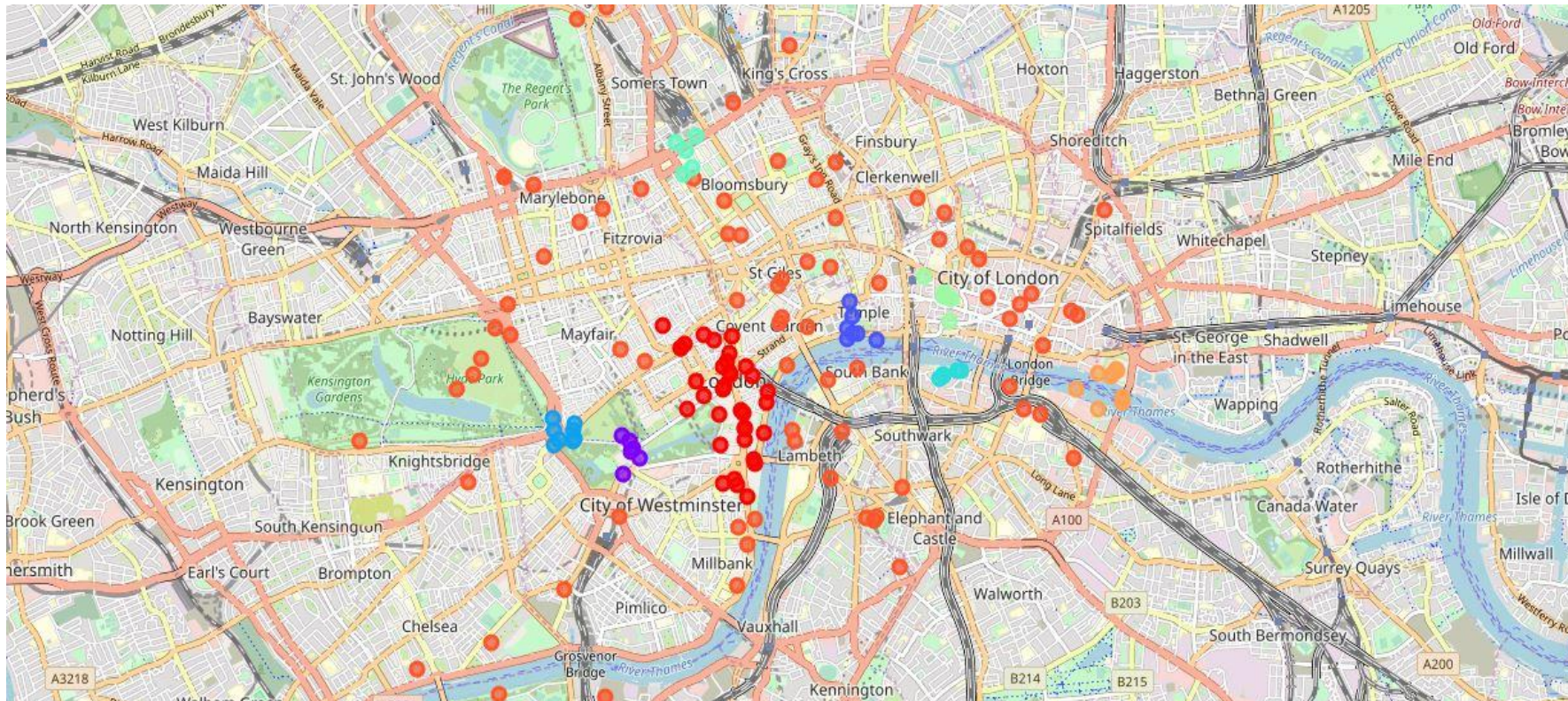
The Data we will use is the Foursquare location data. The Foursquare API allows us to explore areas by categories. In this case, we will use two categories that are frequently visited by tourists: Museums and Landmarks. We will use the Foursquare API to explore from the center of London. The latitude and longitude of London will be obtained by a simple google search.

Methodology

- First, we use the Foursquare API to explore London, specifically searching for Museums and Landmarks. While technically we can search for both of them using the same call, we will use two separate calls for each. The reason is each call is limited to 100 results. If we search both in the same call, we will get 100 totals. If we separate them, we can get up to 100 each. We set the radius to 4 KMs from the center of London.
- After getting the locations of the museums and landmarks, we will cluster them using DBSCAN. This method does not require us to state the initial number of clusters. Instead, it will do that for us. It will also find outliers that don't belong to a cluster. I would like to reference Professor Geoff Boeing for the method to cluster based on locations. His paper can be found here: <https://osf.io/preprints/socarxiv/nzhdc/>
- After clustering, we will clean up our dataframe by removing outlier rows. If there is any cluster that looks too large, we will cluster them further into two separate clusters using KMeans, since this time we know the number of clusters we need.
- The resulting clusters will be mapped into a folium map of London where the travel company can use to determine the location of their information offices.

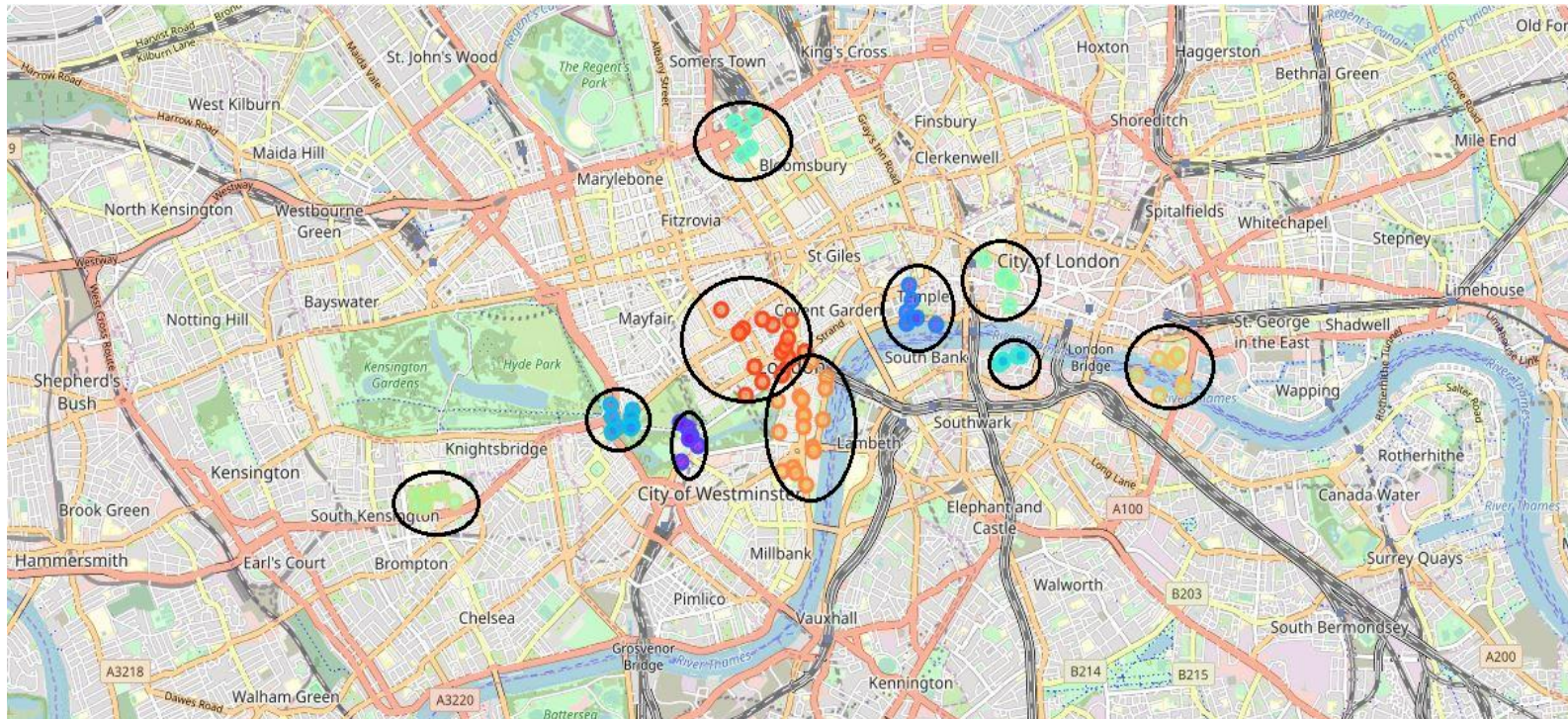
Result: DBSCAN Clusters

- The DBSCAN result yields 9 clusters + outliers



Results: Splitting Red Cluster and Removing Outliers

- We can see that the red clusters is too big so we split it into 2 with KMeans. We also remove the outliers



Results: The 10 Clusters

The 10 locations to put tourist information centers are:

1. Near Buckingham Palace
2. Near Temple
3. Near Wellington Arch
4. South Bank
5. Near Museum of Zoology
6. Near St. Paul's Cathedral
7. South Kensington
8. Near Tower of London / Tower Bridge
9. Westminster
10. Trafalgar Square / Leicester Square / Picadilly Circus

Discussion

The 10 clusters are well suited to have tourist information centers because they contain famous places. We recommend that they sell souvenirs and snacks in their information centers for extra income.

Conclusion

- The Foursquare API serves its purpose to locate tourism venues in London.
- The many data science libraries are very useful to successfully cluster the venues.
- This experiment can be repeated in many other cities around the world.