Capstone Project_London Tech Bar

Topic: To Find the Best Location to Operate A Futuristic High-Tech Bar

Date: 25 April 2020
Author: X FAN

1. Introduction

1.1 Background

This project is a Capstone Project of Coursera Course - 'Applied Data Science Capstone, IBM Data Science Certificate'. In this project, we will explore Central London and try to find the best location to operate a futuristic high-tech bar.

Everyone loves London. London has a long history and is now the financial and technological capital of Europe. The pace of work and life in London is very fast, especially in Central London, which is made of 'Western Central London' and 'Eastern Central London' and where most of the top schools, financial institutions, international enterprises in London cluster there.

The advancement of science and technology has brought convenience to people but also great pressure. Although there are already many pubs, cafes, and bars in Central London for workers to relax their stress, many stores have built for a long history and lack leisure atmosphere that incorporate smart technology. We hope to open a futuristic bar where daily life and digital technology can coexist, so that young and middle-age science and technology professionals can also enjoy the welfare and happiness of life brought by technological progress. They can conveniently find the bar, have a drink after work and make geek friends face-to-face, not just on GitHub, to achieve a real work-life balance.

1.2 Business Problem

However, different boroughs and locations in Central London have distinct culture background, business environment, public cognition and functional infrastructures. For businessmen who want to open a futuristic bar with high-technology theme and interaction elements in Central London, site selection is particularly important.

There are 4 key factors needed to be considered for site selection:

1) Target working place cluster: Industrial cluster of technology startups and Internet companies
2) Target people cluster: People there have high income/ customer consumption level, but also with high working press
3) Accessibility: Convenient access for people who work nearby
4) Culture of having a drink after work: Several pubs or bars already existed, so people have habits to drink and relax there

1.3 Who Would Be Interested In

Businessmen or investors who want to operate or invest a futuristic high-tech bar in Central London.

2. Data

2.1 Data needed

We need to extract/generate following data from corresponding data sources:

- List of Locations in London, United Kingdom with borough name and postcode
- Coordinates information (latitude and longitude) of these locations
- Information of venues around these locations, such as bars, pubs, coffee shops

2.2 Extracting Data

Part1: Scrapping Location Name, London Borough, Post Town, Postcode District, etc information for each locations in London via Wikipedia.

Data set URL: https://en.wikipedia.org/wiki/List_of_areas_of_London

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
...	...	...	...	...	...	...
528	Woolwich	Greenwich	LONDON	SE18	020	TQ435795
529	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4	020	TQ225655
530	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12	020	TQ225815
531	Yeading	Hillingdon	HAYES	UB4	020	TQ115825
532	Yiewsley	Hillingdon	WEST DRAYTON	UB7	020	TQ063804

533 rows × 6 columns

Part2: Using 'geocoder.arcgis' to get latitude and longitude of these locations via Geocoder package.

	Location	Borough	Postcode	Post town	Latitude	Longitude
0	Aldwych	Westminster	WC2	LONDON	51.513307	-0.117092
1	Bloomsbury	Camden	WC1	LONDON	51.520740	-0.123100
2	Charing Cross	Westminster	WC2	LONDON	51.509010	-0.124720
3	Covent Garden	Westminster	WC2	LONDON	51.515440	-0.126930
4	Holborn	Camden	WC1	LONDON	51.516510	-0.119680

```
1  central_data.shape
```

(18, 6)

Part3: Using Foursquare API to get venue data related to these locations.

```
# input the coordinates of all the locations in Central London, create a new dataframe called city_venues.

central_venues = getNearbyVenues (locations = central_data['Location'],
                                  latitudes=central_data['Latitude'],
                                     longitudes=central_data['Longitude']
                                   )
```

```
Aldwych
Bloomsbury
Charing Cross
Covent Garden
Holborn
King's Cross
St Giles
St Pancras
Aldgate
Angel
Barbican
Blackfriars
Clerkenwell
Farringdon
Finsbury
St Luke's
Temple
Tower Hill
```

```
print(central_venues.shape)
central_venues.head()
```

```
(513, 7)
```

	Location	Location Latitude	Location Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aldwych	51.513307	-0.117092	The Courtauld Gallery	51.511584	-0.117559	Art Museum
1	Aldwych	51.513307	-0.117092	Lundenwic	51.512823	-0.118343	Coffee Shop
2	Aldwych	51.513307	-0.117092	The Delaunay	51.513181	-0.117988	Restaurant
3	Aldwych	51.513307	-0.117092	Aldwych Theatre	51.513053	-0.118993	Theater
4	Aldwych	51.513307	-0.117092	Novello Theatre	51.512280	-0.119322	Theater

```
print('There are {} uniques categories.'.format(len(central_venues['Venue Category'].unique())))
```

```
There are 142 uniques categories.
```

3. Methodology

As mentioned above, different boroughs and locations in Central London have distinct culture background and business environment, people already have clear awareness of each location's function and the purpose of going to that location. We use K-means clustering to find similar locations in Central London that meet our requirements on site selection based on the venues and amenities around each location. K - means clustering is a modelling way of unsupervised machine learning that clusters data based on predefined cluster size.

4. Analysis

We analyze these 4 clusters to identify the characteristics of each location in Central London.

Cluster 1: Social activities; Culture and Art; Tourism

Cluster 1 is made of 4 locations in Western Central London. The most common venues in these locations are theater, coffee shop. hotel, dessert and bakery shop. We can image that people and tourists in these locations gather together to watch operas, buy ice creams and

desserts in the street shops and settle down in nearby hotels after a whole day touring.

Cluster 2: Pub; Creative lifestyle
Cluster 2 is made of 7 locations in Eastern Central London. The most common venues in these locations are pub, coffee shop, hotel, gym/fitness center, garden/park, salad/falafel restaurant, as well as venues that are not common in other clusters such as indie movie theater, performing arts venue, cocktail bar. They present a creative, energetic, young and vigorous lifestyle.

Cluster 3: Far away from the center of Central London in geographical location
Cluster 3 is made of 1 location in Eastern Central London but far away from the center of Central London. We could obviously leave out this cluster.

Cluster 4: Traditional lifestyle; Transportation
Cluster 4 is made of 6 locations in both Eastern and Western Central London. As St Luke, St Giles and Finsbury are also far away from the center of Central London according to the map, we leave out these three locations. The most common venues in the remaining three locations are café, coffee shop, hotel, restaurant from various countries, as well as venues that are not common in other clusters such as train station. They present a traditional European lifestyle.

5. Results and Discussion

After analyzing these 4 clusters' characteristics, and taking the 4 key factors for site selection raised in the introduction part - Target working place cluster, Target people cluster, Accessibility, Culture of having a drink after work - into consideration, we can see 4 Eastern Central London locations in Cluster 2 (Aldgate, Barbican, Blackfriars, Clerkenwell) may meet our requirements. First, there are many pubs so that people there already have the habits on drinking after work, know there are bars and pubs in that borough and will be more likely to try a new bar. Second, the number of gym/fitness center, salad/falafel restaurant and juice bar in Cluster 2 is significantly larger than that of other clusters, indicating that people there choose to keep an energetic and healthy living habit. Moreover, venues for young and fashion people such as indie movie theater, performing arts venue, cocktail bar are especially common in Cluster 2, meaning that people there will try modern and futuristic things. Therefore, we prefer locations in Cluster 2.

6. Conclusion

The aim of this project is to assist businessmen or investors in making decisions to select a propriate location to operate or invest a futuristic high-tech bar in Central London, by exploring venues around target locations and get preliminary portrait and business environments around them. We start by processing Central London Locations' basic and coordinates information extracted from Wikipedia and Geocoder, then using Foursquare API to get the information of venues around those locations, and finally using K-means clustering

machine learning algorithm to cluster those locations based on the similarity of their venue category and frequency. We find locations in Cluster 2 present a creative, energetic, young and vigorous lifestyle so that preliminarily match the theme - 'balancing work and life in high-technology interaction' – of our futuristic high-tech bar. Further work should be done to take other factors, such as the number and distribution of high-tech companies, rental fee, theme of pub/bar's nearby, into consideration to shortlist the locations, and present more valuable information for site selection reference.