

INVESTMENTS TO CONSIDER IN THE EASTERN DISTRICT OF THE CITY OF LONDON COUNTY

CHRISTOBEL ACHEN

01 JULY 2019

1. Introduction

1.1 Background

London, the capital of England is a modern city with a long history stretching as far back as the Roman times. It is considered to be one of the most important global cities and has been termed the worlds' most powerful, most visited as measured by international arrivals with the busiest city airport system as measured by passenger traffic, most expensive, most innovative and investment friendly. It ranks 26 out of 300 major cities for economic performance making it one of the top leading investment destinations. Hosting more international retails and ultra-high net worth individuals than any other city, it is among the largest metropolitan area GDP. The population within the London commuter belt is the most populous in the European Union with 14,040,163 inhabitants in 2016.

With this population and economic atmosphere, investment opportunities are endless. Complete research is required though before embarking on any economic activity to ensure maximum profitability. The area of interest for this research is the East end of London which is a subset of East London consisting of areas close to the ancient City of London.

1.2 Problem

Data required to explore the East end of London is majorly the existing postal codes in the area as a unique identifier, the wards which are four in this particular area out of the 25 in the whole London, the train stations closest to particular postcodes since it is a major transportation system in the city, the already existing business venues and their traffic frequency. This project aims to find out what business ventures are most profitable in the east end of London using the mentioned data.

1.3 Interest

A lot of entrepreneurs would be interested in studying a good number of investment opportunities before embarking on one that they are comfortable with and expect full returns on their investment from.

2. Data acquisition and cleaning

2.1 Data sources

The data set with London's postcodes, Counties, Districts, Wards, Geospatial information, train stations and how close they are to the postcodes can be found at the following url <https://www.doogal.co.uk/UKPostcodesCSV.ashx?area=London>. To get venues in the wards, I used foursquare data.

2.2 Data cleaning

Data retrieved from the csv file was read into jupyter notebooks. For the most part the data set was clean except some postcodes were no longer in use, however since the purpose of this research is to use the codes as a unique identifier, this development did not affect the dataset so all postcodes were retained.

2.3 Feature selection

After cleaning the data, there were 320426 samples and 44 features. Eight of the 44 features required namely Postcode, Latitude, Longitude, County, District, Ward, Nearest station and Distance to nearest station were selected.

On selecting the rows in the 'E1 xxx' Postcode of 'City of London' district, I realized a redundancy between the County feature and District feature so the County feature was dropped since the one required county was already acquired. After all that, 82 samples and 7 features remained

From the foursquare data, venue name, venue categories, venue latitude and longitude were retrieved and used in analysis.

3. Methodology

This project aimed to find out what business ventures were most profitable in the East end of London city. The data used was sourced from uniform resource locator "<https://www.doogal.co.uk/UKPostcodesCSV.ashx?area=London>" and foursquare venues data. Of the data acquired, 320426 samples and 44 features, 8 features namely Postcode, Latitude, Longitude, County, District, Ward, Nearest station and Distance to nearest station were selected. The samples in the Postcode 'E1 XXX' and County 'City of London' were then selected which left a redundancy between 'County' and 'District' hence 'County' feature was dropped because it was a known factor. The final data frame was left with 82 samples and seven features. We then visualized the samples on a map. The venues which were within 200 meters radius around each Ward was explored using four squares API and a total of 2164 venues returned with 57 unique venue categories. The top 10 venues in each Ward were later retrieved using a defined function. One hot encoding was later used to convert the data into a form that could be provided to ML algorithms for better prediction. KMeans was lastly employed to group the venues into four clusters with similar properties and the results visualized on a map. The tools used were IBM Cloud jupyter notebooks where libraries like pandas, numpy, folium, matplotlib and the like, where used to process and visualize the data.

4. Results

The eastern district of 'City of London' county consists of 4 Wards namely Bishopsgate, Tower, Aldgate and Portsoken. This was visualized using a map.

There are 2164 venues in this area. Bishopsgate has the highest number of venues totaling 841 followed by Portsoken 801, Tower 348, and lastly Aldgate 174. The top 3 venues in each ward are as follows:

TOP THREE VENUES IN THE EASTERN DISTRICT OF 'CITY OF LONDON' COUNTY

Bishopsgate		Portsoken		Aldgate		Tower	
Venue	Freq	Venue	Freq	Venue	Freq	Venue	Freq
Coffee shop	0.07	Pub	0.10	Salad Place	0.07	Hotel	0.18
Wine Bar	0.06	Coffee Shop	0.09	Cocktail	0.07	Pub	0.13
Mediterranean Restaurant	0.06	Hotel	0.06	Gym/Fitness Centre	0.06	Gym/fitness Centre	0.10

The hotel in Tower had the most frequency with a difference of 0.12 with the venues with low frequency.

Using KMeans to group the venues into four clusters, they all tended to their specific Wards in that the first cluster tended to Aldgate, the second to Tower, the third to Bishopsgate and the fourth to Portsoken.

On further evaluating the clusters, the following was observed

CLUSTER EVALUATION

	Aldgate	Tower	Bishopsgate	Portsoken
1st most common venue	Cocktail Bar	Hotel	Coffee Shop	Pub
2nd most common venue	Salad Place	Pub	Wine Bar	Coffee Shop
3rd most common venue	Gym/Fitness Center	Gym/Fitness Center	Mediterranean Restaurant	Hotel
4th most common venue	Italian Restaurant	Coffee Shop	Italian Restaurant	Italian Restaurant
5th most common venue	Cafe	Café	Cafe	Cocktail Bar
6th most common venue	Indian Restaurant	American Restaurant	Salad Place	Café
7th most common venue	Sandwich Place	Italian Restaurant	Plaza	Sandwich Place
8th most common venue	Coffee Shop	Wine Bar	Wine shop	Indian Restaurant
9th most common venue	Mediterranean Restaurant	Bar	Organic Grocery	Gym/Fitness Center
10th most common venue	Middle Eastern Restaurant	Cocktail Bar	Szechuan Restaurant	Pizza Place

- The first cluster is around Aldgate ward with closest postcode to Aldgate station 0.16 miles and furthest 0.24 miles away
- The second cluster is around Tower ward with closest postcode to Tower Gateway station being 0.00 miles and furthest being 0.15 miles away
- The third cluster around Bishopsgate ward with closest postcode to Aldgate station 0.23 miles and furthest 0.28 miles, closest postcode to Liverpool Street station being 0.18 miles and furthest being 0.24 miles away
- The fourth cluster around Portsoken ward with closest postcode to Aldgate station 0.09 miles and furthest 0.23 miles, closest postcode to Tower Gateway station being 0.19 miles and furthest being 0.21 miles away

5. DISCUSSION

- The Tower Ward is right next to the station which explains the high frequency in traffic to the hotel. The best business venture in this ward would be accommodation or coffee shop. Travellers love their coffee.
- The Bishopsgate ward has a high volume of venues compared to the rest which in turn brings in more traffic. Since it is close to two stations, this could be an added advantage. With Coffee shop at the top, this business venture wouldn't hurt your wallet. You have to make it very delicious though so that they keep coming back and forget your competitors. You could also open a wine bar or Mediterranean restaurant only if you can cook Mediterranean food or hire someone to.
- In Aldgate ward, the Salad place and cocktail bar have the same frequency so either one could be first common venue or second common venue which would be my choices if ever I wanted to invest in the ward
- In Portsoken ward, second in place in respect to number of venues because of its proximity to two stations like Bishopsgate, the best business idea would be Pub or Coffee shop. People love the hotel at Tower Ward so I wouldn't venture into that here.

6. CONCLUSION

In this project, the investments to consider in the Eastern district of the 'City of London' County were examined. The Postcode, Latitude, Longitude District and Ward were identified as the important features. KMeans clustering was used to further analyze the Wards and what they had to offer to an entrepreneur. The results can be useful in that an entrepreneur can decide on what to invest in using these results.