

The Battle of the Neighbourhood – Optimal Restaurant location

Qianqian Wu

November 29, 2020

1. Introduction

1.1 Background

Thousands of people dream of starting and owning a successful restaurant. In almost every case, the path to entrepreneurial success starts with choosing the right location for the new restaurant. Where the restaurant is located can be just as important to success as the menu, marketing and customer reviews. “In the brick-and-mortar retail world, it is said that the three most important decisions are location, location and location,” affirms Irene Dickey, a lecturer in management and marketing at the University of Dayton’s School of Business. The restaurant expert Lorri Mealy explains that one reason restaurants fail is a poor choice in location. Therefore, an optimal location has a disproportionate effect on new restaurants that are not established yet with a core of loyal customers and can give the restaurant another push toward success.

London is one of the world’s most important global cities and exerts a considerable impact upon the arts, commerce, education, entertainment, finance, media, tourism etc. London is considered to be the most multicultural city in the world, and this is reflected onto the different cuisines on offer. The purpose of this project is to use Data Science and Machine Learning approach to find an optimal restaurant location in London.

1.2 Problem

The first step of opening a new restaurant is to **define the concept of the restaurant and common market targets**. According to the Google Map results, there are currently only 5 Uyghur Restaurants in Greater London, but four of which are located in the outskirts of London. Therefore, there is a significant market gap in London City Center. Therefore, this Data Science project will focus on **finding an optimal location for a new Uyghur Restaurant** (hereafter referred to as the ‘New Restaurant’) in London.

Common market targets include business professionals, urban hipster, families with children, sports enthusiasts, culture aficionados and fast-food customers. In addition, each area/location has different demographic makeup, consists of the workers and residents in the vicinity and tourists. Due to the characteristics of Uyghur food being fast to prepare, unique, **common market targets is defined to be business professionals, families with children, culture aficionados and fast-food customers**. In terms of ethnic group, Uyghur cuisine reflects the cooking styles of many ethnic groups of the Xinjiang region in China and is regarded one of the most popular cuisines among Chinese. The Uyghur food is predominantly halal. Therefore, **in addition to Chinese workers, residents and tourists, Muslim population and tourists can be potential customers of the New Restaurant**.

Next step is to **identify the minimum target market population** the New Restaurant will need using two formulas below. The Minimum target market population is calculated by daily restaurant

capacity multiplied by days of a month. The daily restaurant capacity can be calculated by the desired seating capacity and number of seating.

Minimum target market population = daily restaurant capacity x days of a month

Daily restaurant capacity = desired seating capacity x number of seatings

It is assumed that the New Restaurant can accommodate 40 people at once (desired seating capacity) and the number of seating is 3. According to the formula above, the daily restaurant capacity is 120.

The maximum days of a month is 31. Therefore, minimum target market population is calculated using 120 multiplying with 31, which equals to 3720. This means, **the minimum target market population is 3720**. Ideally, the new restaurant should be located in communities that have a high density of the target market and beyond the minimum target market population (3720).

While choosing an optimal restaurant location, the following influencing factors needs to be considered :

1. Accessibility: When looking at a restaurant location, the amount or accessibility of nearby parking, ease of access by car, foot traffic and surrounding traffic patterns should be taken into consideration to make it as easy as possible for customer to visit the new restaurant. The target locations should acquire enough traffic to fill tables organically, without any help from digital marketing or promotional efforts.

2. Visibility: Making the New Restaurant or restaurant sign visible to the public is like free advertising. It reminds the passers-by that a restaurant exists, and they should stop by for dinner sometime.

3. Target Demographics: A target demographic is a specific grouping of consumers that you market your products toward. Demographics are defined by specific characteristics such as age, gender or income level. A combination of all these factors creates the best demographic picture to help determine how a certain location would affect the New Restaurant. For each potential space, the following aspects will be considered:

- Is this location suitable for my concept?
- What is the population of this area?
- How many customers can I expect?
- Is there enough expandable income in this area to support my restaurant?
- What is the average age of people who live in this area?
- What is the average household income of people who live in this area?
- What is the ethnic breakdown of this neighbourhood?

4. Safety/Crime Rates: Extra insurance, exterior lighting and loss of customers due to safety concerns can spell disaster for the success of a new restaurant. Hence, it is critical to recognise and address the issues of safety and crime rates early on.

5. Complimentary Businesses: By “complementary businesses”, it is referred to the neighbouring businesses that could share their market with the New Restaurant but aren’t competition. The activities of complementary businesses actually drive diners to the New Restaurant. Simply being in close proximity to them can increase the foot traffic. Complementary businesses include but aren’t limited to stadiums, theatres, business districts, transportation hubs, and malls. For example, Office

buildings can bring hungry professionals to the New Restaurant for lunch. Restaurant near the train stations can attract commuters. In the vicinity of Theatres and stadiums, restaurants can capture attendees pre-and-post performance or a big game.

6. Direct and Indirect Competitors: When choosing a location, it is prudent to make sure that the market is not over-saturated. Some competition is healthy, but if there are multiple restaurants with the same cuisine and concept as the New Restaurant, we need to look for another location. In addition, it is important to make sure that businesses nearby are thriving. This indicates that there's an appetite for consumption. Furthermore, identifying gap is one of the key points for a successful restaurant. If there's a way we can refine the New Restaurant concept based on a market gap, the odds of success heighten.

7. Property price: in London, property price or rent play a significant role in deciding the optimal restaurant location. One of the biggest expenses is the rent on the restaurant premises. The more popular the location and the more it is likely to cost.

In this project, we will focus on find an optimal location for a new Uyghur Restaurant based on the influencing factors described above. This report will be targeted to stakeholders who are interested in opening a Uyghur Restaurant in London, United Kingdom.

1.2 Analytic Approach

In this project, we will use Data Science and Machine Learning methods to locate the target areas and analyse each target area and list top 15 optimal locations for a new Uyghur Restaurant in London.

In order to filter out inappropriate locations, the followings criteria are implemented to narrow down the target areas in Greater London.

Table 1: Influencing Factors and Selection Criteria.

Influencing Factors		Criteria
Accessibility		<ul style="list-style-type: none"> Public Transport Accessibility Score is a method used in United Kingdom transport planning to assess the access level of geographical areas to public transport.
Visibility		<ul style="list-style-type: none"> Distance to the nearest station
Target Demographics	<ul style="list-style-type: none"> Business professionals, employees Middle class Ethnic group: Asian and other mixed 	<ul style="list-style-type: none"> Job density Household Median Income Average Salary (£30000-£80000) Ethnic group population
Safety/Crime Rates		<ul style="list-style-type: none"> Crime rates
Complimentary Businesses		<ul style="list-style-type: none"> Number of active businesses Number of venues near the target areas
Direct and Indirect Competitors		<ul style="list-style-type: none"> Distance to the nearest station=≤ 0.3 km Number of restaurants within 500m radius of target locations ≤ 1 Number of Chinese and Turkish restaurants within 500m radius of target locations=0
Property price/rent		<ul style="list-style-type: none"> Median House Price

As there is a number of factors which need to be considered while deciding, clustering method will be used to reduce the scale of the data and narrow down the targeted locations. Clustering will be used to create centres of zones that contain targeted postcode areas. Clustering is an unsupervised machine learning algorithm. It can group data based on the similarity to each other and partition the data into mutually exclusive groups. Neighbourhood/ Areas segmentation will partition a city or countries into groups of areas that have similar characteristics. It is a significant strategy, as it allows the business to narrow down the targeted locations while choosing the optimal location . In addition, clustering is also an effective analytical approach to deriving segments and groups from large datasets. For instance, the target market population can be grouped based on age, gender, income, ethnic groups etc.

2. Data Description

2.1 Data sources

Two kinds of data are required for this project: location data and demographic data in each borough and post code area. I collected demographic data of all Boroughs and post code areas in Greater London from the following websites:

Table 2: Data Source and Description

Data Source	Data Name	Data Description
https://data.london.gov.uk	London Borough Profiles	The csv file describes the demographic profiles of 33 London Boroughs. It contains data such as population, household, household median income, job density, active businesses, crime rates, median house price, public transport accessibility etc.
https://data.london.gov.uk	Ethnic Groups Borough	Data format is xls. It describes the population of ethnic groups (White, Asian, Black and Mixed/Other) in all London Boroughs.
https://www.doogal.co.uk	London Postcodes	Data format is csv. It contains information postcode, latitude, longitude, nearest station, distance to station, borough, rural/urban areas, average income.

Location data is data describing places and venues such as their geographical location in form of geographical coordinates (latitude and longitude value), their category, nearby venues etc. For this Data Science project, I will use the Foursquare location data segment Greater London and group the London Boroughs/Neighbourhoods into clusters to narrow down the targeted areas and get deeper insights into a neighbourhood's characteristics. Furthermore, the Foursquare location data also helps us to locate the New Restaurant near to complimentary businesses e.g. substation, museum, theatre etc. and to the areas with great accessibility and visibility. Moreover, the Foursquare location data also can be used to carry out the competitive analysis to find out whether the market is oversaturated (whether is restaurant nearby and what kind of restaurant are there near the targeted areas) or to identify the market gap.

2.2 Data Cleaning

Data download from multiple sources were read into three data frames. However, there are several problems with the datasets.

1. First, all three datasets contain much more data than we need for analysis and some column names are too long and contains plenty of special characters. Therefore, I selected the key columns (see table 1 criteria) and assigned them to a new data frame and cleaned the column names of these key columns.
2. Some columns have incorrect data type. For example, household median incomes of all boroughs contain pound sign and comma, which have been removed. In addition, columns household, household median income, crime rates, median house price and public transport accessibility have data type object that have been converted to numeric data type int or float.
3. Several missing values are identified. Missing values, different approaches to deal with missing values as well as reasons are summarised as below.

Table 3: Approaches to deal with missing values

Missing value	Approach	Reason
crime rate of city of London	replaced by the median of crime rates of all other London boroughs	Due to the paramount importance of the London Borough of City of London and limited number of London boroughs, dropping the whole row is considered as inappropriate.
population of ethnic groups (White, Asian, Black, Mixed_Other) in City of London	dropped the rows which contains the missing values.	Population of individual ethnic groups can not be replaced by mean values. Because I found out that after replacing the population of each ethnic groups in City of London by the mean values, the sum is bigger than the original total population in London Borough of City of London which contradicts the facts.

2.3 Feature selection

After reading the London Borough Profile data into the data frame, it is found that there are 38 rows and 84 columns. As Greater London has only 33 boroughs, there is redundancy in rows. Based on the analytical approach and methodology described above and examining the meaning of each features, only the following features are selected for this analysis. Kept features, dropped features and reason for dropping features are outline in the table 4.

Table 4: Kept Features and Feature Dropping Process

Data	Kept Features	Dropped Features	Reasons for dropping features
London Borough Profiles	'Code', 'Borough', 'Inner_Outer_London', 'Population', 'Household', 'Household_Median_Income', 'Jobs_Density', 'Active_Businesses', 'Crime_Rates', 'Median_House_Price', 'Public_Transport_Accessibility'	rest of the columns	They don't have impact on the success of a restaurant.
Ethnic Groups Borough	'Borough', 'White', 'Asian', 'Black', 'Mixed_Other', 'Total'.	rest of the columns	They are redundancies and can not be used for data analysis.
London Postcodes	'Postcode', 'Latitude', 'Longitude', 'District', 'Rural/urban', 'Nearest station', 'Distance to station', 'Average Income'	rest of the columns	They don't have impact on the success of a restaurant.

3. Exploratory Data Analysis

3.1 Inner London Boroughs and Outer London Boroughs

Through exploratory data analysis, it is found that there are 14 boroughs in Inner London. They are City of London, Camden, Hackney, Hammersmith and Fulham, Haringey, Islington, Kensington and Chelsea, Lambeth, Lewisham, Newham, Southwark, Tower Hamlets, Wandsworth, Westminster. 19 boroughs are located in Outer London, they are Barking and Dagenham, Barnet, Bexley, Brent, Bromley, Croydon, Ealing, Enfield, Greenwich, Harrow, Havering, Hillingdon, Hounslow, Kingston upon Thames, Merton, Redbridge, Richmond upon Thames, Sutton, Waltham Forest (Figure 1).

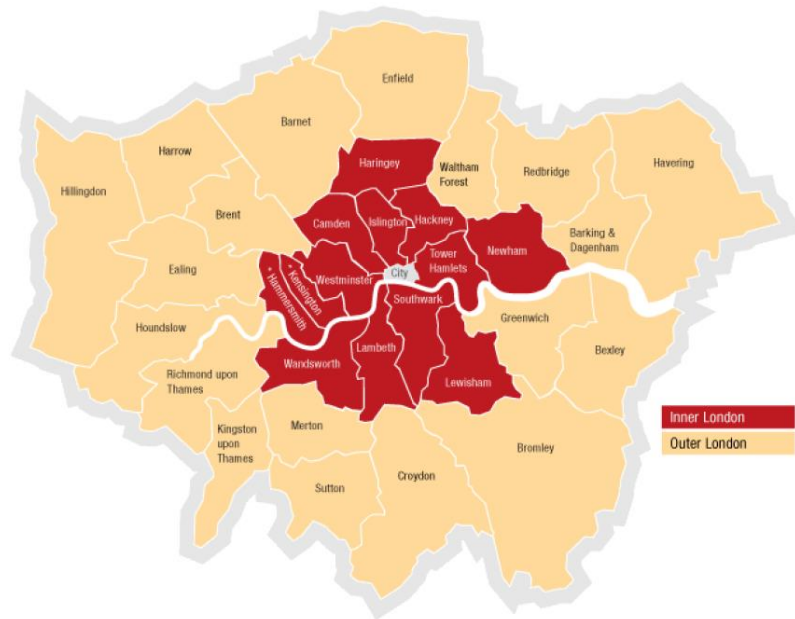


Figure 1: Inner London Boroughs and Outer London Boroughs

3.2 London Boroughs with Highest Public Transport Accessibility and Job Density

Public Transport Accessibility is a detailed and accurate measure of the accessibility of a point to the public transport network, taking into account walk access time and service availability. It measures the density of the public transport network at any location within Greater London. The first bar chart shows the top 7 boroughs that possess the highest public transport accessibility score. As shown in the bar chart below, City of London is the best connected with other locations in London via public transport followed by London Borough of West Minster. Kensington and Chelsea, Camden and Islington have similar public transport accessibility score (5.8) on third place followed by Lambeth and Tower Hamlets with public transport accessibility of 5.

In Greater London, the majority of the jobs (over 85%) concentrate in the London Borough City of London. It is worth mentioning that if the target customers are the business professionals, City of London is an advantageous location for a new restaurant as the job density is ten times as much as in other boroughs. Westminster and Camden are on the second and third place in terms of job density, however having job density only around 5% and 3% respectively.

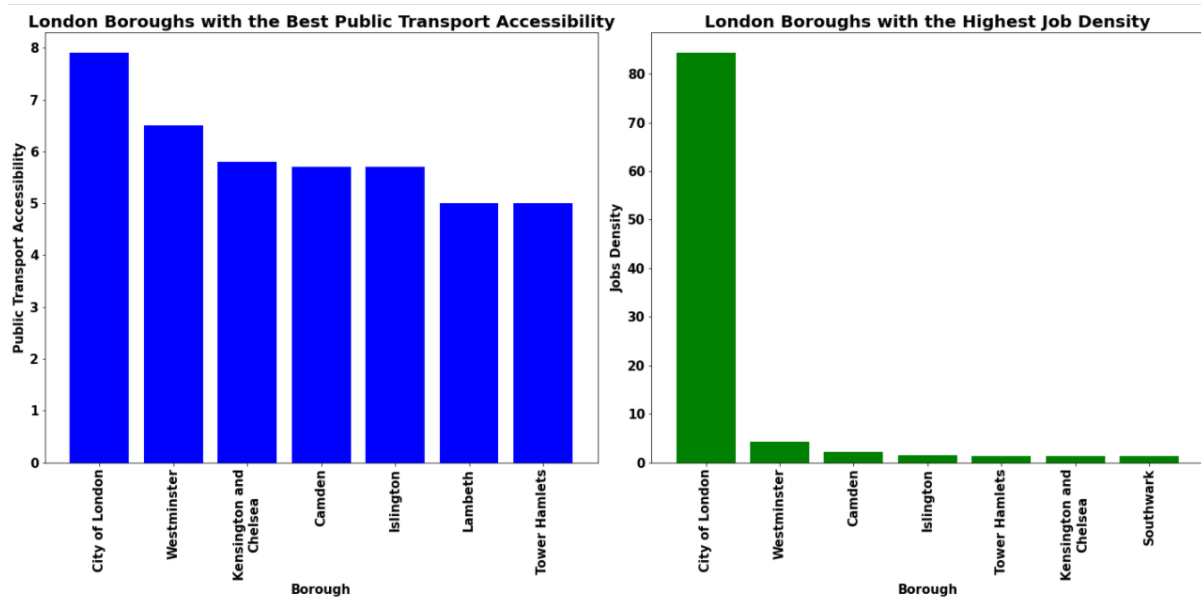


Figure 2: Top 7 London Boroughs with the Best Public Transport Accessibility and Highest Job Density

3.3 Richest London Boroughs

Household Median Income is the combined gross income of all members of a household who are 15 years or older. It is an important parameter indicating an area's standard of living. In this analysis, it is used to indicate, how likely that people would go to a restaurant for dinner instead of cooking at home. The higher the Household Median Income, the higher the discretionary income (the amount of income a household or individual has to invest, save, or spend after taxes and necessities are paid) will be and the more likely that people will go out for dinner. As shown in the bar chart below, the residents in City of London have the highest Household Median Income (£63000) which is followed by Kensington and Chelsea with approximately £55000 and Richmond upon Thames with approximately £53000. People who live in Westminster is as affluent as the residents in Wandsworth who have around £48000 Household Median Income.

The number of active business is a significant index to demonstrate the business in the areas are thriving and there's an appetite for consumption. It is noteworthy that City of London possesses highest number of active businesses which is almost as twice as the number of active businesses in Westminster in second place. The prosperity level in Camden is as high as in Islington, which

followed by Tower Hamlets with approximately 25000.

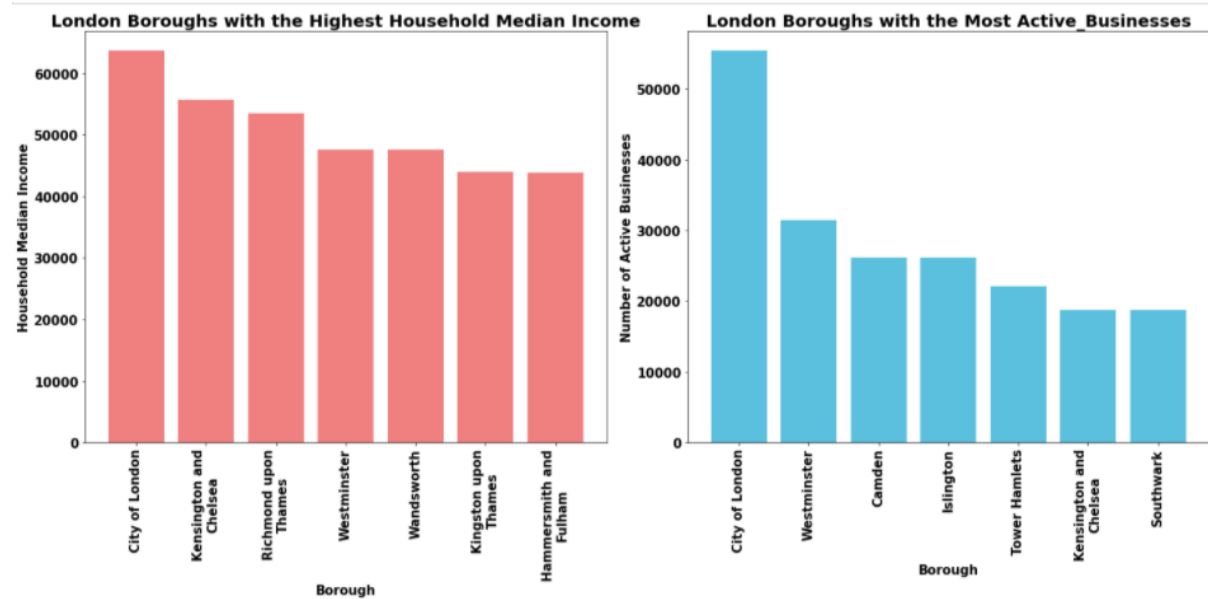


Figure 3: Top 7 Boroughs with the Highest Household Median Income and Number of Active Businesses.

3.4 Boroughs with highest Target Ethnic Group Population

The Ethnic Group Population data in London divided the people into White, Black, Asian, Mixed_Others. The target ethnic group is determined to be Asian and Mixed_Others.

On the one hand that Uyghur cuisine reflects the cooking styles of many ethnic groups of the Xinjiang region in China, middle Asian and is regarded one of the most popular cuisine among Chinese. The Uyghur food is predominantly halal. Therefore, **in addition to Chinese, Muslim can be potential customer of the New Restaurant.**

On the other hand, the people who belong to Mixed_Other have parents from other countries but were born and grow up in UK. They are often open-minded to other cultures and cuisine.

The bar chart below shows that the population of the target ethnic group is highest in Newham followed by Brent, Tower Hamlets, Ealing, Redbridge. Hillingdon and Harrow are on the sixth and seventh place.

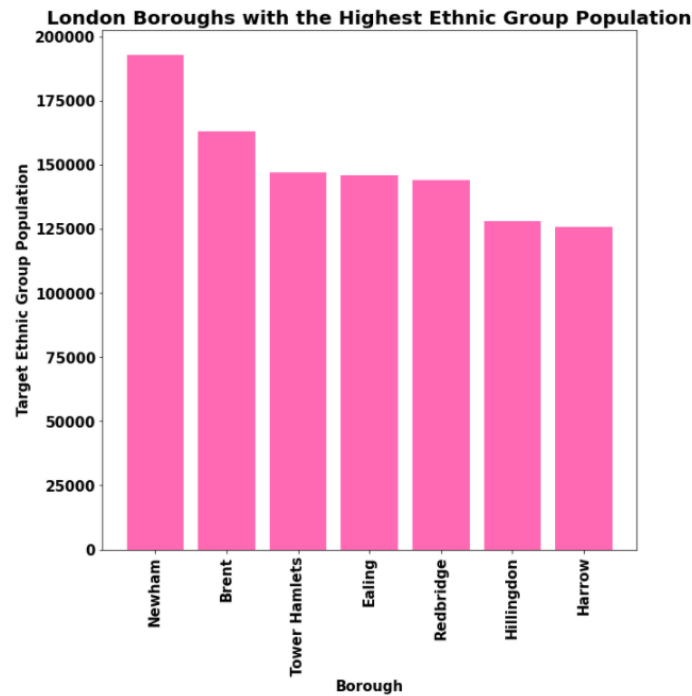


Figure 4: London Boroughs with the Highest Target Ethnic Group Population

3.5 Optimal Boroughs

On the basis of the data analysis described above, evaluation matrix is created to select the optimal boroughs for the new Uyghur Restaurant.

Each Borough is assessed against the promoting factors and hindering factors. Based on the available data (Section 2), promoting factors are Public Transport Accessibility, Job Density, Household Median Income, Number of Active Businesses, Target Ethnic Group Population. It means that the higher these factors are, the more likely the restaurant will be successful. However, Crime Rates, Median House Price are considered as hindering factors as high values of these parameters may lead to less customers and higher lost for the new restaurant.

Therefore, following rules are implemented to in the evaluation model:

1. Boroughs which are within the top 7 borough group in terms of the promoting factors, get 1 score under individual evaluation factors.
2. Boroughs which are NOT within the top 7 borough group in terms of the hindering factors, get 1 score, because the hindering factors would probably not impact the success of new restaurant significantly.

Figure 5 is a heatmap created using Python to demonstrate the evaluation process and results. The final winner is City of London and Tower Hamlets with 5 score.

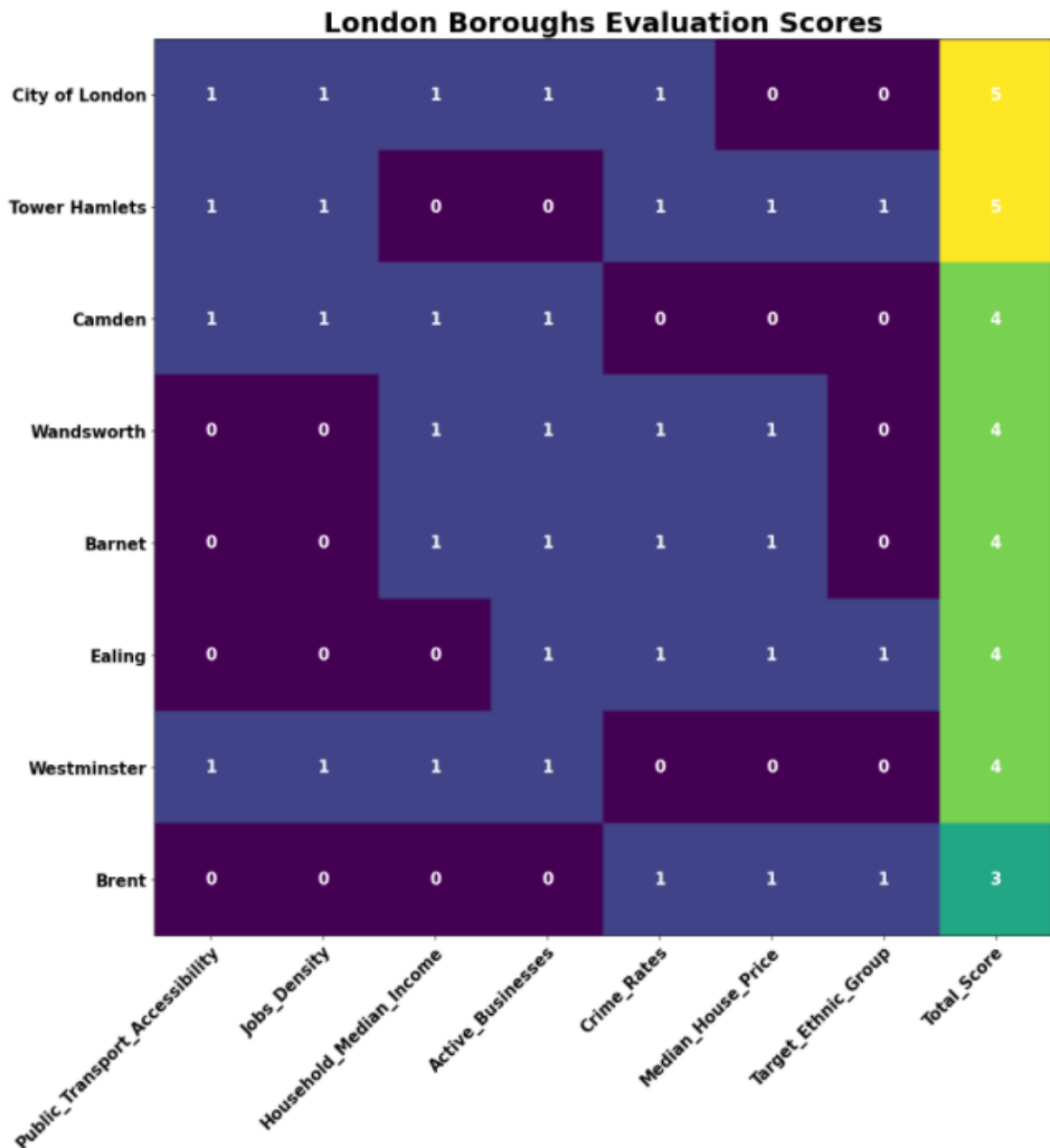


Figure 5: Evaluation Matrix to select the optimal Borough for the Uyghur Restaurant

4. Modelling

4.1 Selection Process

City of London and Tower Hamlets are selected as optimal boroughs for the new restaurant as shown in Figure 5. Next step is to select the optimal postcode areas within these two boroughs. The following criteria are implemented in order to narrow down target postcode areas.

Table 5: Selection Process to Narrow Down the Target Postcode Areas

Step Nr.	Number of Postcode Areas before the selection	Selection Criterium	Number of Postcode Area after selection
1	323306	select postcode in use	179130
2	179130	select “Urban Major Conurbation”, “Urban City and Town” under column Rural_Urban	178488
3	178488	select all postcode areas within the two optimal boroughs (City of London, Tower Hamlets)	7343
4	7343	select postcode areas which have distance to station less than 0.3 km and average income less than £80000	3418
5	3418	Machine Learning Model clustering	30

It is worth mentioning that in the last step unsupervised machine learning algorithm – clustering is used to narrow down the target location. Clustering is the task of dividing data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters. Clustering is used to create cluster center of each optimal postcode area group and effectively reduced the target areas from 3418 to 30.

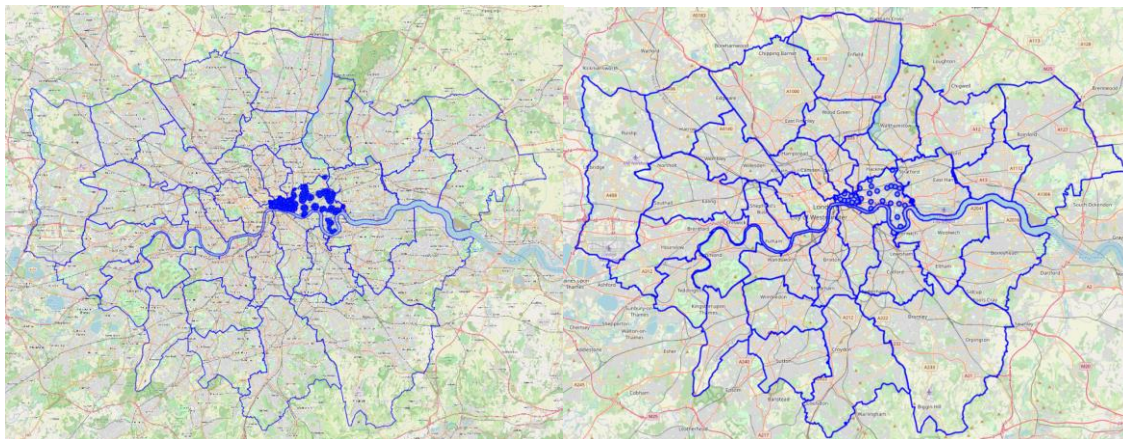


Figure 6: Target Areas before Selection Process

Figure 7: Target Locations after Selection Process

4.2 Competitor Analysis

When choosing a location, it is prudent to make sure that the market is not over-saturated. Some competition is healthy, but if there are multiple restaurants with the same cuisine and concept as the New Restaurant, the location needs to be removed from the target area list. In order to carry out the competitor analysis, Foursquare location data are used to analyse the venues in the radius of

500m of each target location. In addition, it helps us to locate the New Restaurant near to complimentary businesses e.g. substation, museum, theatre etc.

The following criteria are used to complete competitor analysis:

1. Target location must have more than 5 venues (complementary businesses) in the radius of 500m;
2. Target location cannot have Chinese or Turkish Restaurant (direct competitor businesses) in the radius of 500m;
3. Target location cannot have more than 1 restaurant (indirect competitor businesses) in the radius of 500m.

In the end, 10 optimal location for a new Uyghur Restaurant are selected and are shown in the map below. There are four optimal locations in City of London and six optimal locations in Tower Hamlets.



Figure 8: Optimal Locations for a new Uyghur Restaurant in London

The address of these ten locations are found using Python Geopy package and is outlined as below:

	Latitude	Longitude	Address
0	51.511482	-0.085488	(Gracechurch Street / Monument Station, Lombard Court, Bishopsgate, City of London, Greater London, England, EC3V 0BT, United Kingdom, (51.5113694, -0.0855789))
1	51.513069	-0.097199	(The Young Lovers, Saint Paul's Church Yard, Blackfriars, City of London, Greater London, England, EC4V 5AN, United Kingdom, (51.5132037, -0.0971893))
2	51.513221	-0.091905	(HSBC UK, Pancras Lane, Barbican, City of London, Greater London, England, EC4N 1SE, United Kingdom, (51.5130324, -0.09195332008496107))
3	51.519429	-0.097174	(Aldersgate Street, Smithfield, City of London, Greater London, England, EC1A 4AB, United Kingdom, (51.51943492229131, -0.09715411213765121))
4	51.519994	-0.059132	(Kempton Court, Dunward Street, Whitechapel, London Borough of Tower Hamlets, London, Greater London, England, E1 5BB, United Kingdom, (51.5201377, -0.05958108194432182))
5	51.526666	-0.056334	(Old Street Brewery & Taproom, Arch 11, Gales Garden Mews, Bethnal Green, London Borough of Tower Hamlets, London, Greater London, England, E2 0EJ, United Kingdom, (51.5269805, -0.0563595))
6	51.524364	-0.012912	(24, Hannaford Walk, St Andrews development, Bow, London Borough of Tower Hamlets, London, Greater London, England, E3 3FA, United Kingdom, (51.52422336, -0.012851860000000001))
7	51.526222	-0.019783	(Webb House, 3, Trevithick Way, Bromley-by-Bow, Bow, London Borough of Tower Hamlets, London, Greater London, England, E3 3GD, United Kingdom, (51.52621425, -0.019894890459650698))
8	51.525079	-0.032139	(435, Mile End Road, Mile End, London Borough of Tower Hamlets, London, Greater London, England, E3 4PA, United Kingdom, (51.52590470909091, -0.032285663636363635))
9	51.519085	-0.016375	(21, Balladier Walk, Bromley-by-Bow, Poplar, London Borough of Tower Hamlets, London, Greater London, England, E14 6BG, United Kingdom, (51.5182704, -0.0166563))

5. Conclusions

This data science project used London Borough demographic data and Foursquare Location data as well as Machine Learning Algorithm Clustering to find optimal locations for a new Uyghur Restaurant in London.

The following conclusions have been reached during the analysis.

1. Greater London has 33 Boroughs in total. 14 boroughs are located in Inner London. They are City of London, Camden, Hackney, Hammersmith and Fulham, Haringey, Islington, Kensington and Chelsea, Lambeth, Lewisham, Newham, Southwark, Tower Hamlets, Wandsworth, Westminster. 19 boroughs are located in Outer London, they are Barking and Dagenham, Barnet, Bexley, Brent, Bromley, Croydon, Ealing, Enfield, Greenwich, Harrow, Havering, Hillingdon, Hounslow, Kingston upon Thames, Merton, Redbridge, Richmond upon Thames, Sutton, Waltham Forest.
2. Top 7 Boroughs with the best Public Transport Accessibility are: City of London, Westminster, Kensington and Chelsea, Camden, Islington, Lambeth and Tower Hamlets.
3. Job density of City of London is ten times as much as in all other Boroughs.
4. The Top 7 richest boroughs are City of London, Kensington and Chelsea, Richmond upon Thames, Westminster, Wandsworth, Kingston upon Thames and Hammersmith and Fulham.
5. The most prosperous areas are City of London, Westminster, Camden, Islington Tower hamlets, Kensington and Chelsea and Southwark
6. City of London and Tower Hamlets are two optimal boroughs to open a new Uyghur restaurant.
7. After geodata analysis (venues analysis) and competitor analysis, 10 optimal locations are selected for a new Uyghur Restaurant. four of which are in City of London and 6 of which are located in Tower Hamlets.

6. Future directions

This analysis has two limitations:

1. Due to the lack of data of ethnic group population, public transport accessibility, number of active businesses in each postcode area, the accuracy of the optimal location is limited. It is recommended that these ten locations with the areas in the radius of 500m be regarded as the optimal location.
2. Due to limited number of free API calls of Foursquare location data, the target locations have to be significantly reduced using clustering method under the assumption that the centre of each cluster (target area group) still fulfil all the criteria proposed in this analysis.