

Capstone Project - The Battle of Neighbourhoods

Introduction

In the city of London, England, identify the most likely location for a new restaurant.

Choosing a location for any new restaurant is very important and, once chosen, can be expensive and time consuming to change later on. Tips on how and where to start a new restaurant can be found on line (see example sites below), citing various factors to take into consideration when choosing the location of a new restaurant. One of these factors is other venues in the area. The consensus being that siting a new restaurant in an untried location is a gamble but, equally, choosing a location with too many restaurants with the same concept can over saturate an area, leading to potential business failure.

How to Open a Restaurant, Part 2: Location, Location, Location

(<https://openforbusiness.opentable.com>)

8 Factors for Choosing a New Restaurant Location (<https://www.foodnewsfeed.com>)

Tips on Where to Locate Your Restaurant (<https://www.thebalancesmb.com>)

It is, therefore, very important to analyse an area for the number and type of different restaurants, to assist a new restaurant owner in determining a good location for their restaurant and avoiding those areas where their restaurant would be more likely to fail.

Equally, any type of new business enterprise would benefit from analysing where to locate their premises, in order to determine the best location for their customers to find them.

Business Problem

Using cluster analysis on Foursquare location data can we assist a new restaurant owner to identify potential 'goldilocks' neighbourhoods that are both not untried nor over-saturated and, therefore, potentially viable locations for their new venture.

For the purposes of this project, the value of oversaturation may need to be adjusted but will be initiated at 6 restaurants already in the same area. Additional specialisation may be required to differentiate between different types of restaurant within a neighbourhood.

Data Sources

It will be necessary to obtain a list of neighbourhood names and associated postcodes. These can be scraped from the Wikipedia page https://en.wikipedia.org/wiki/G_postcode_area. The data will need to be cleaned to remove un-necessary columns and extra text within the neighbourhood's field. There are also a few postcodes that are labelled 'non-geographic', which will also need to be removed from the dataset. Finally, the data will be restricted to those postcodes that are within the local authority area of 'Glasgow City' as the other local authority areas would give too great a physical area in which to locate the new restaurant. Resulting in: Postcode Neighbourhood G1 Merchant City G2 Blythswood Hill, Anderston G3 Anderston, Finnieston, Garnethill, Park, Woodlands, Yorkhill G4 Calton, Cowcaddens, Drygate, Kelvinbridge, Townhead, Woodlands, Woodside G5 Gorbals

Due to the buggy nature of geocoder, a CSV file for the longitudes and latitudes for each postcode will be manually generated and uploaded for use in this project:

The Foursquare location data will then be leveraged to identify and locate other venues in each of the neighbourhoods. This will be restricted to other restaurants and cafes as the objective is to recommend the location of a new restaurant. This data can then be scraped to give the restaurant information for a particular area:

Using the venue ID, a second call to Foursquare can retrieve the rating for that venue, for example, Jamie's Italian is 7.3 and The ASK Italian in Port Dundas Place is 6.6 This could be useful additional information for deciding the suitability of a location for a new restaurant, as if there is a highly popular venue of the same category in an area, their customer's may not be tempted away to a new restaurant. However, these are premium calls to Foursquare, which are limited per day. Once a suitable 'goldilocks' neighbourhood has been identified, this additional step may be useful to confirm its viability.

Methodology

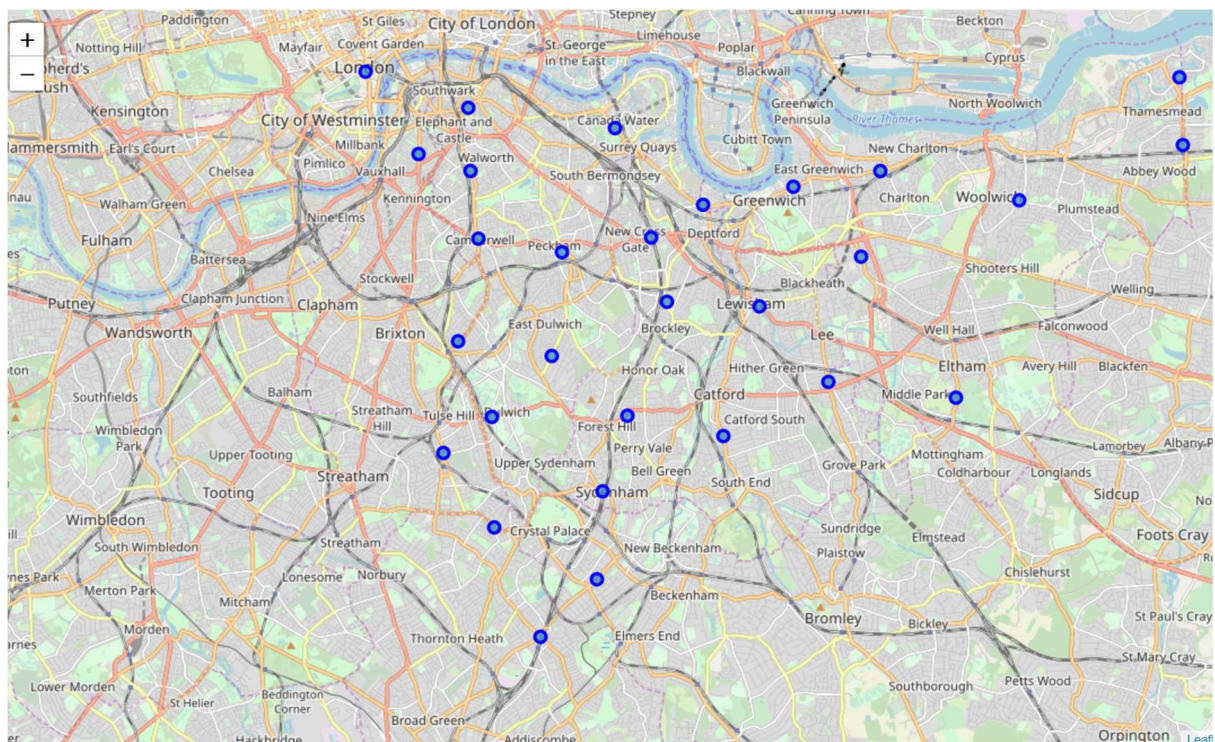
The Methodology is separated into 3 main sections:

1. Retrieving and Cleaning Neighbourhood data for Glasgow
2. Retrieving Venue Details for Glasgow
3. Distribution Analysis of Glasgow Venues

Retrieving and Cleaning Neighbourhood data for Glasgow

Postcode data for the Glasgow area was retrieved from Wikipedia¹ and scraped into a Pandas dataframe using the Beautiful Soup Python library.

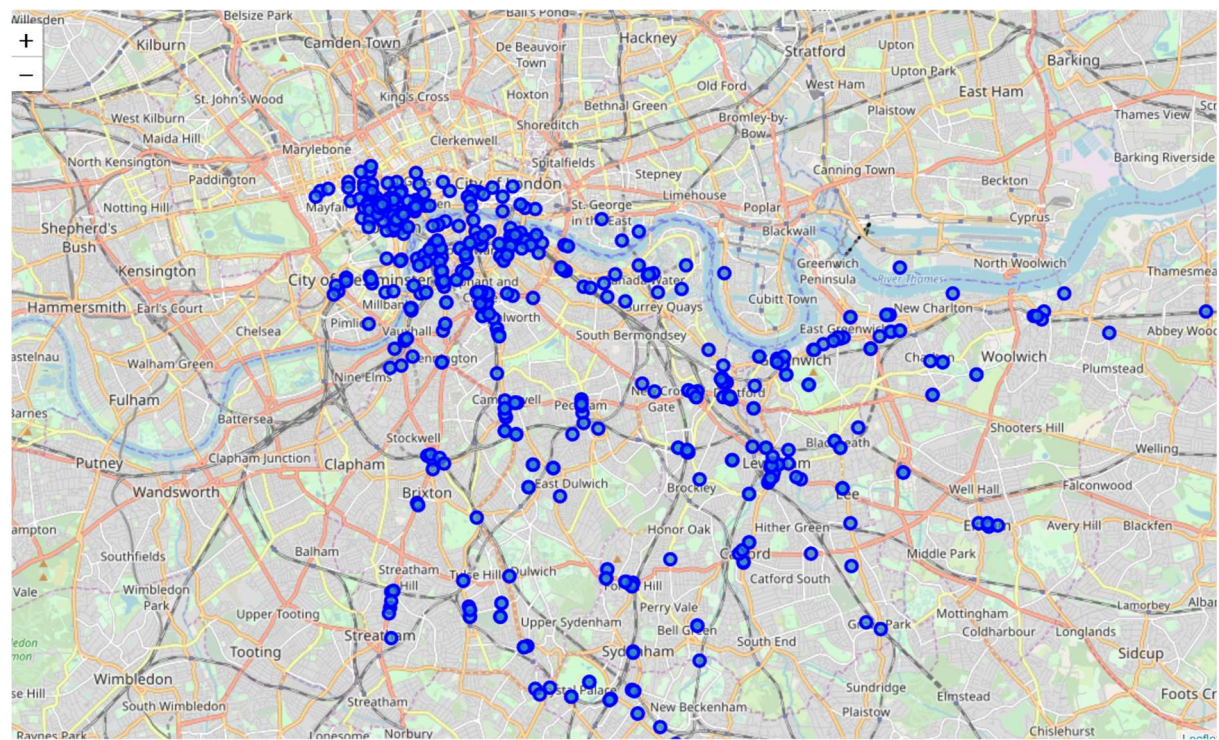
Postcode	Neighbourhood	Latitude	Longitude
SE1	Bankside, South Bank, Lambeth (part), Southwar...	51.49960	-0.09613
SE1P	NaN	51.50642	-0.12721
SE2	Abbey Wood, West Heath, Crossness, Thamesmead ...	51.49245	0.12127
SE3	Blackheath, Kidbrooke, Westcombe Park	51.47138	0.02338
SE4	Brockley, Crofton Park	51.46268	-0.03558



Retrieving Venue Details for Glasgow

To retrieve restaurant information within each neighbourhood, the Foursquare5 search API was employed. The radius was determined by measuring the distance between adjacent postcodes6 and taking half of the average. This was found to be 1300m, rounded to 1500m. Also, a limit of 10,000 venues was set to ensure the maximum number of venues was captured. For each venue, its Neighbourhood name, latitude and longitude and the Venue name, category, latitude and longitude were recorded. Some neighbourhoods did not return any results, but an if statement was used to check if the 'name' feature was present in the Foursquare results. In the event that it was not the Neighbourhood name, latitude and longitude were recorded a before but a value of 'none' was recorded for Venue name and category and 0 for Venue latitude and longitude.

Neighbourhood	Latitude	Longitude	Venue_Name	Venue_Cat	Venue_Lat	Venue_Lng
Abbey Wood, West Heath, Crossness, Thamesmead ...	51.49245	0.12127	Taj Mahal Indian Restaurant	Indian Restaurant	51.491146	0.120691
Abbey Wood, West Heath, Crossness, Thamesmead ...	51.49245	0.12127	none	none	0.000000	0.000000
Anerley, Crystal Palace (part), Penge, Beckenh...	51.41009	-0.05683	Costa Coffee	Coffee Shop	51.420814	-0.067452
Anerley, Crystal Palace (part), Penge, Beckenh...	51.41009	-0.05683	High Peak Coffee Co	Café	51.419183	-0.053644
Anerley, Crystal Palace (part), Penge, Beckenh...	51.41009	-0.05683	Coffee Craft	Coffee Shop	51.400116	-0.075263

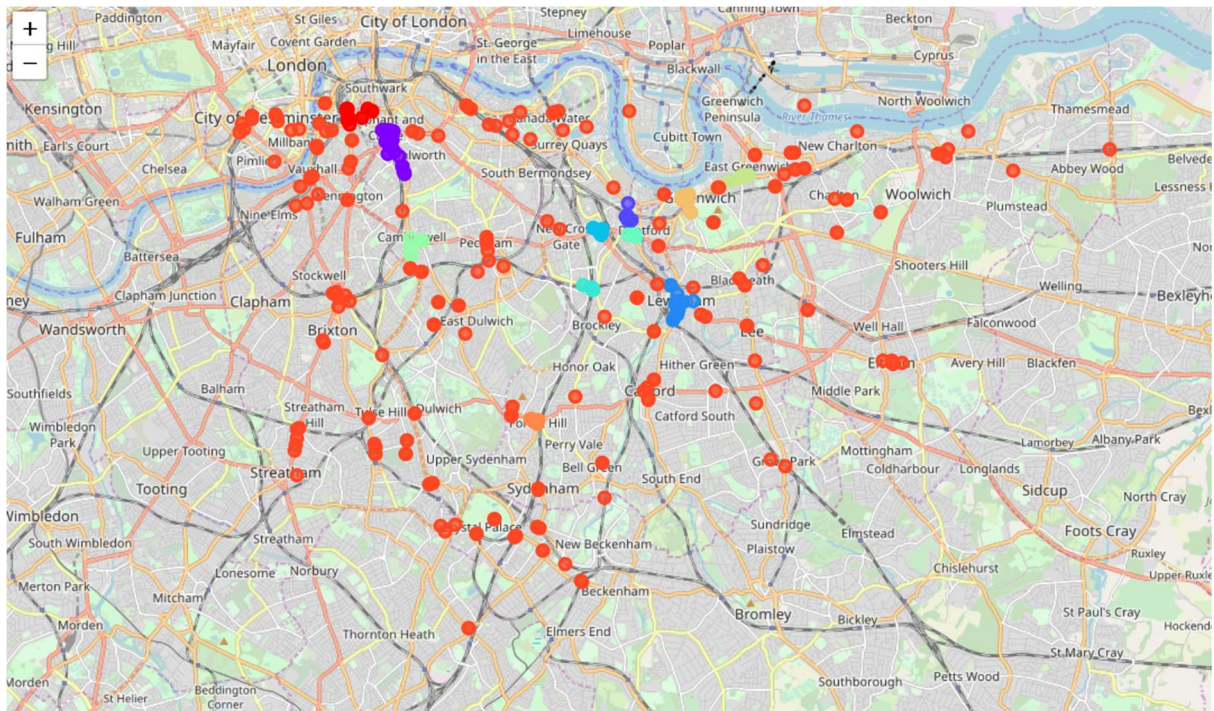


Results and Discussion

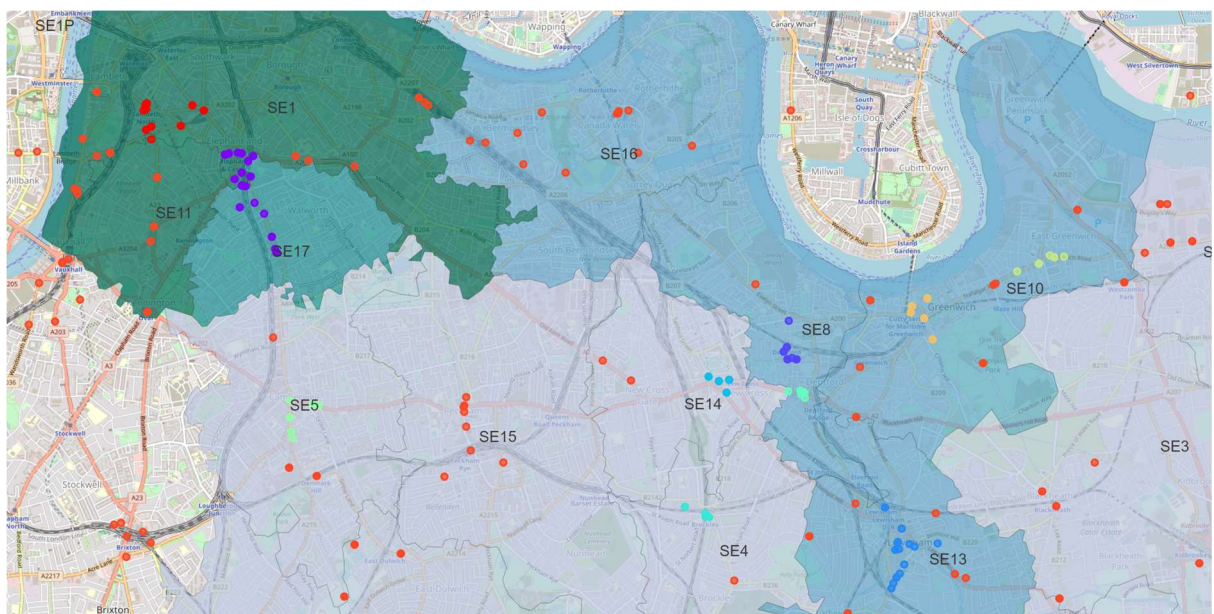
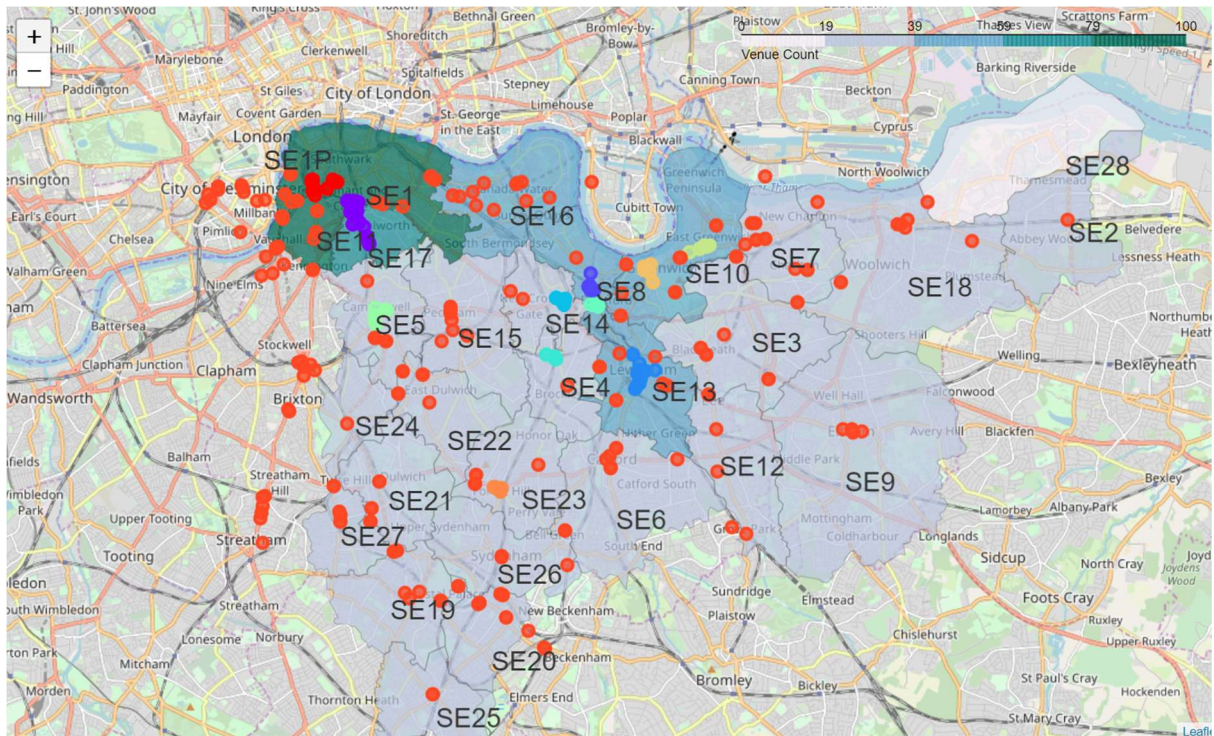
Counts of the number of Venues per Neighbourhood :

Postcode	Neighbourhood	Latitude	Longitude	Counts
SE1	Bankside, South Bank, Lambeth (part), Southwar...	51.49960	-0.09613	79.0
SE11	Kennington, Lambeth (part), Vauxhall (part), O...	51.49084	-0.11108	61.0
SE17	Walworth, Kennington (part), Newington	51.48764	-0.09542	41.0
SE8	Deptford, Evelyn, Rotherhithe (part)	51.48114	-0.02467	27.0
SE13	Lewisham, Hither Green, Ladywell	51.46196	-0.00754	25.0
SE16	Rotherhithe (part), Surrey Quays, South Bermon...	51.49574	-0.05157	22.0
SE10	Greenwich, Maze Hill, Greenwich Peninsula	51.48454	0.00275	19.0
SE4	Brockley, Crofton Park	51.46268	-0.03558	17.0
SE26	Sydenham, Crystal Palace (part)	51.42674	-0.05533	16.0
SE14	New Cross	51.47489	-0.04038	16.0
SE7	Charlton	51.48759	0.02912	15.0
SE24	Herne Hill, Tulse Hill (part), Dulwich (part)	51.45529	-0.09928	15.0
SE27	West Norwood, Gipsy Hill (part)	51.43407	-0.10375	14.0
SE20	Anerley, Crystal Palace (part), Penge, Beckenh...	51.41009	-0.05683	14.0
SE15	Peckham, Nunhead, South Bermondsey (part), Old...	51.47218	-0.06779	13.0
SE5	Camberwell, Denmark Hill, Peckham, Brixton (part)	51.47480	-0.09313	13.0
SE23	Forest Hill, Honor Oak, Crofton Park (part)	51.44122	-0.04764	9.0

DBSCAN cluster analysis of the data



Combination of the count choropleth and DBSCAN analysis



Conclusion

The purpose of this project was to suggest a neighbourhood within SE London where a new restaurant owner could potentially site a new venue. A location that was neither untried nor oversaturated with other venues. To this end, London was divided into its district postcodes and cluster analysis was applied to lists of current venues obtained from Foursquare.

Through analysis of the results it was suggested that the most viable neighbourhoods for sustaining and new venture would likely be SE5 or SE13

However, it was also noted that

1. There was some concern that Foursquare was not returning a full list of venues
2. Location is only one factor to consider when siting a new business
3. The district postcode areas are quite large and it may be beneficial to fine tune the analysis by reapplying it to the sector postcodes.

That being said, this is a good first step for any budding restaurant owner to take in setting up their new business.