Data Science Project

Maria-Lida Kounadi

Exploring London Neighborhoods - Restaurant

Contents

Introduction / Business Problem

Introduction

The objectives of the IBM Data science specialisation - Capstone Project are:

- Clearly define a problem or an idea, where Foursquare location data is leveraged to solve or execute
- To remember that data science problems:
  - always target an audience
  - are meant to help a group of stakeholders solve a problem

Problem Statement

"Exploring business idea for a new cafe/restaurant/co-working space in London through Foursquare data"

London, where I live, is a business hub.

According to the Global Power City Index (GPCI)* London ranks first on its power to attract people, capital and enterprises from our the world. London maintained its position for the 8th consecutive year.

Using data science methodology and machine learning techniques like clustering, this project aims to help decision making for property developers to open a new cafe/co-working space.

**http://mori-m-foundation.or.jp/english/ius2/gpci2/index.shtml*

Problem Statement

"Exploring business idea for a new cafe/restaurant/co-working space in London through Foursquare data"

During the daytime, especially in the morning and lunch hours, office areas provide huge opportunities for restaurants and coffee shops.

Also, coworking spaces will become a need to provide people with a space to work and at the same time enjoy their coffee and lunch. They don't have to worry about renting an office and they can use our area for meetings, working environments that can provide amenities.

So our case study is to understand which is the best location to open our new cafe/restaurant/co-working space based on Foursquare data.

Target Audience

“Which audience this project targets?”

1. Property developers
2. Investors
3. People that want to open a new cafe/restaurant/coworking space and understand which areas in London would be ideal for this. This can be also their side business.
4. Data scientists that can use our steps for their own projects, if they want to explore neighbourhoods.

Data

Data

- List of neighborhoods in London from Wikipedia: https://en.wikipedia.org/wiki/List_of_areas_of_London
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data. GeoPy
- Land value data to inform the decision making based on the capital investment required.
- Foursquare data to explore boroughs and its venues, point of interests.
- Map https://joshuaboyd1.carto.com/tables/london_boroughs_proper/public

*source: London Map Art Print by Clair Rossiter

Data Preparation

2. Data Preparation

2.1 Scraping London Neighborhoods from wiki

Using Beautiful Soup the data from the wikipedia page is scraped

	Location	London borough	Post town	Postcode district
0	Abbey Wood	Bexley	LONDON	SE2
1	Acton	Ealing	LONDON	W3
2	Addington	Croydon	CROYDON	CR0
3	Addiscombe	Croydon	CROYDON	CR0
4	Albany Park	Bexley	BEXLEY	DA5

2.3 Average Land Price

Data from https://propertydata.co.uk/cities/london is used and added to the table

	Location	London borough	Post town	Postcode district	Avg price	£/sqm
0	Abbey Wood	Bexley	LONDON	SE2	368814	4133.4
1	Crossness	Bexley	LONDON	SE2	368814	4133.4
2	West Heath	Bexley	LONDON	SE2	368814	4133.4
3	Acton	Ealing	LONDON	W3	547488	7330.3
4	Addington	Croydon	CROYDON	CR0	347577	4757.7

2. Data Preparation

2.2 Getting coordinates through GeoPy.

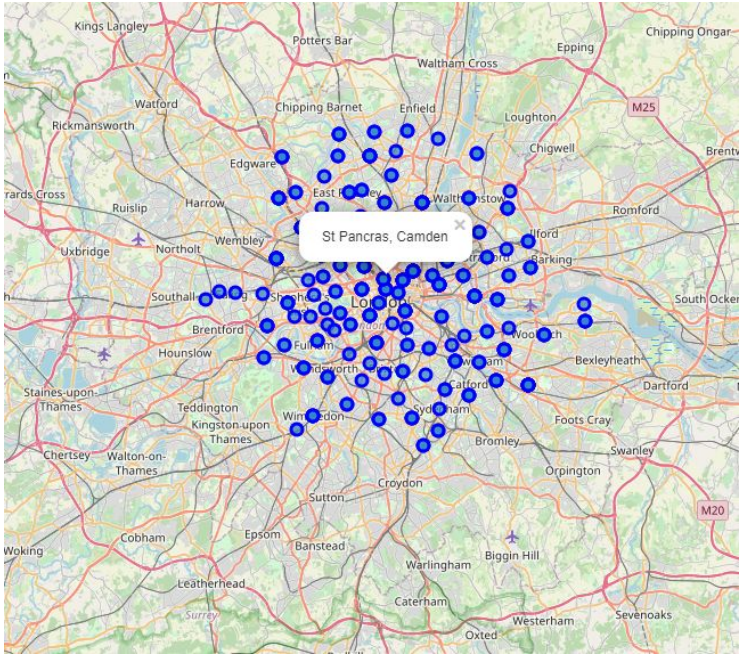
This study explores only “London” post town.

Using the arcgis the latitude and longitude is added to the table

[30]:

	Location	London borough	Postcode district	Avg price	£/sqm	Latitude	Longitude
0	Abbey Wood	Bexley	SE2	368814	4133.4	51.49245	0.12127
1	Crossness	Bexley	SE2	368814	4133.4	51.49245	0.12127
2	West Heath	Bexley	SE2	368814	4133.4	51.49245	0.12127
3	Acton	Ealing	W3	547488	7330.3	51.51324	-0.26746
4	Aldwych	Westminster	WC2	1662350	17599.1	51.51651	-0.11968
...	...	...	...	...	...	...	...
295	West Ealing	Ealing	W13	579287	7319.5	51.51453	-0.31951
296	West Kensington	Hammersmith and Fulham	W14	874941	10720.9	51.49568	-0.20993
297	West Norwood	Lambeth	SE27	471617	6167.8	51.43407	-0.10375
298	Woodford	Redbridge	IG8	559158	5414.3	51.50642	-0.12721
299	Woodford Green	Redbridge	IG8	559158	5414.3	51.50642	-0.12721

The map of London showing the locations and boroughs is visualised



2. Data Preparation

2.4 Using Foursquare Location Data

Using the Foursqare API, different locations are explored and the Top5 venues (most frequent) of each location are identified. The average price also of each location is added.

	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Price
0	Abbey Wood	Supermarket	Coffee Shop	Platform	Convenience Store	Train Station	0.044347
1	Acton	Grocery Store	Train Station	Park	Indian Restaurant	Breakfast Spot	0.044347
2	Aldwych	Sandwich Place	Pub	Theater	Café	Coffee Shop	0.044347
3	Anerley	Supermarket	Hotel	Fast Food Restaurant	Convenience Store	Grocery Store	0.128359
4	Angel	Food Truck	Coffee Shop	Pub	Italian Restaurant	Hotel	0.652565

Methodology

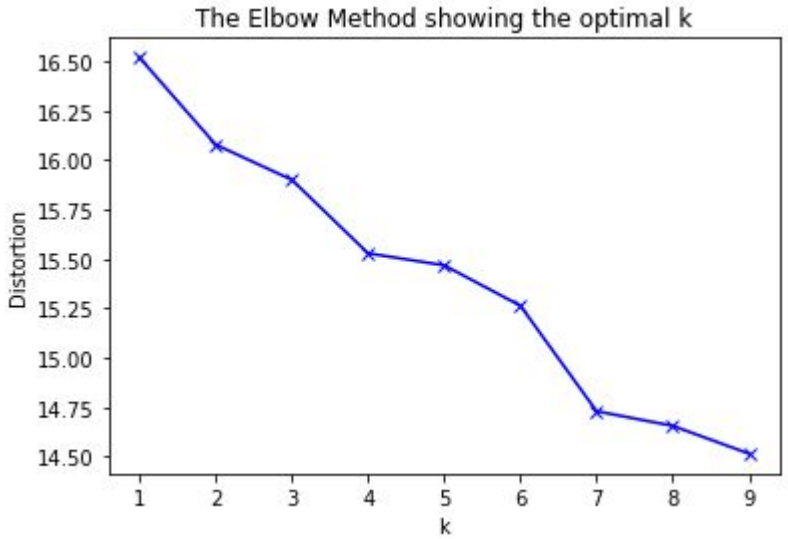
3. Clustering Locations (Neighborhoods)

There are some common venue categories in locations and unsupervised machine learning K-means algorithm is used to cluster the locations.

Initially the dataset is normalised.

Elbow method (figure) and silhouette method are used to identify the best k.

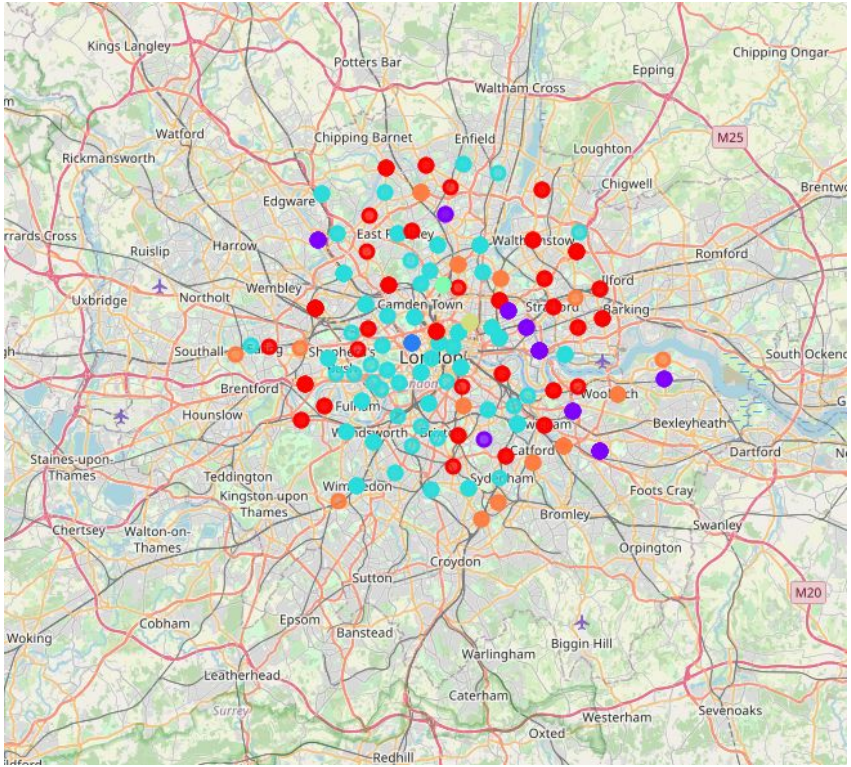
Then the K-Means algorithm was run to cluster the boroughs into 7 clusters because, as shown in the figure, there is a 7 degree for optimum k of the K-Means.



3. Clustering Locations (Neighborhoods)

Then the K-Means algorithm was run to cluster the boroughs into 7 clusters because, as shown in the figure in the previous slide, there is a 7 degree for optimum k of the K-Means.

And the results were visualised in the map in the next figure.



Results

4. Results

4.1 Clusters

By examining the clusters, the following labels were resulted.

Clusters	Labels
Cluster 1	Restaurants
Cluster 2	Mixed Social venues
Cluster 3	Touristic places (cafe - hotel)
Cluster 4	Light bites
Cluster 5	Cafe & Sports events
Cluster 6	All day social venues
Cluster 7	Stores and fast foods

4.2 Average prices tags

By examining the prices, the following labels were resulted.

Labels	Average Price
Low level 1	<£630k
Low level 2	£630k - £985k
Average level	£985k - £1340k
Average level 2	£1340k - £1700k
High level 1	£1700k - £2400k
High level 2	> £2400k

4. Results

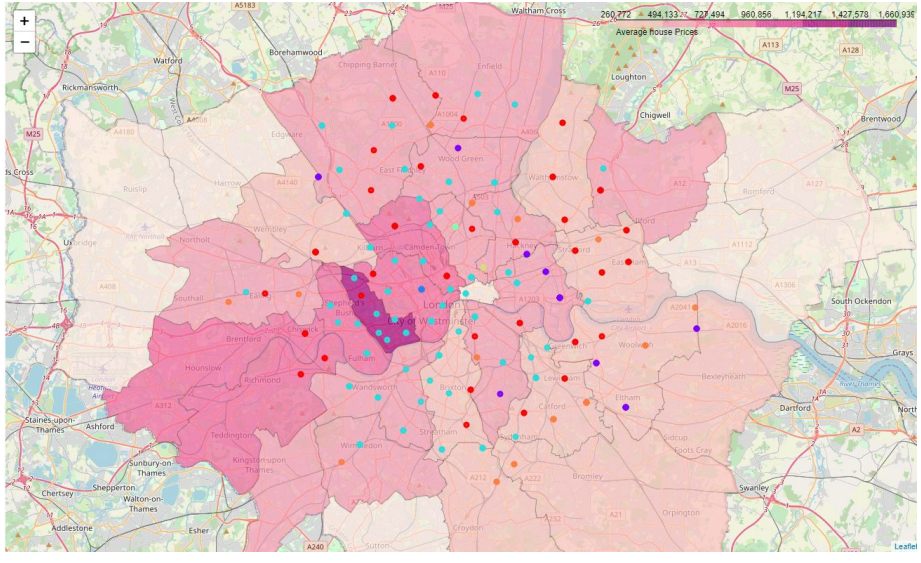
The resulted table is as below (the first 10 locations)

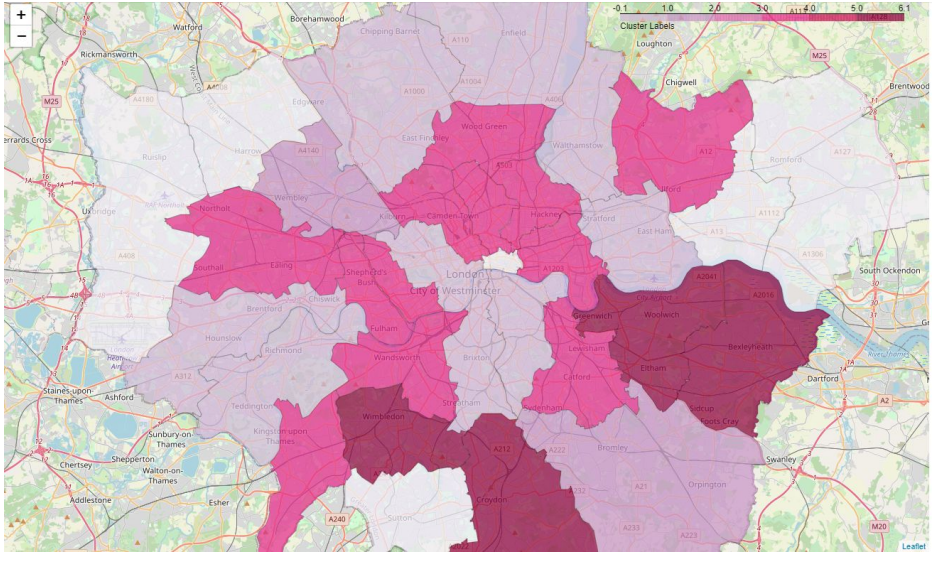
	Location	London borough	Post town	Postcode district	Avg price	£/sqm	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Price	Price-Categories	Cluster-Category
0	Abbey Wood	Bexley	LONDON	SE2	368814	4133.4	51.49245	0.12127	1	Supermarket	Coffee Shop	Platform	Convenience Store	Train Station	0.044347	Low level 1	Mixed Social Venues
1	Crossness	Bexley	LONDON	SE2	368814	4133.4	51.49245	0.12127	1	Supermarket	Coffee Shop	Platform	Convenience Store	Train Station	0.160450	Low level 1	Mixed Social Venues
2	West Heath	Bexley	LONDON	SE2	368814	4133.4	51.49245	0.12127	1	Supermarket	Coffee Shop	Platform	Convenience Store	Train Station	0.228340	Low level 1	Mixed Social Venues
3	Acton	Ealing	LONDON	W3	547488	7330.3	51.51324	-0.26746	6	Grocery Store	Train Station	Park	Indian Restaurant	Breakfast Spot	0.044347	Low level 1	Stores and fast foods
4	Aldwych	Westminster	LONDON	WC2	1662350	17599.1	51.51651	-0.11968	3	Sandwich Place	Pub	Theater	Café	Coffee Shop	0.044347	Average level 2	Light bites
5	Charing Cross	Westminster	LONDON	WC2	1662350	17599.1	51.51651	-0.11968	3	Sandwich Place	Pub	Theater	Café	Coffee Shop	0.218559	Average level 2	Light bites
6	Covent Garden	Westminster	LONDON	WC2	1662350	17599.1	51.51651	-0.11968	3	Sandwich Place	Pub	Theater	Café	Coffee Shop	0.361649	Average level 2	Light bites
7	St Giles	Camden	LONDON	WC2	1662350	17599.1	51.51651	-0.11968	3	Sandwich Place	Pub	Theater	Café	Coffee Shop	0.048912	Average level 2	Light bites
8	Anerley	Bromley	LONDON	SE20	348970	5500.4	51.41009	-0.05683	6	Supermarket	Hotel	Fast Food Restaurant	Convenience Store	Grocery Store	0.128359	Low level 1	Stores and fast foods
9	Penge	Bromley	LONDON	SE20	348970	5500.4	51.41009	-0.05683	6	Supermarket	Hotel	Fast Food Restaurant	Convenience Store	Grocery Store	0.183618	Low level 1	Stores and fast foods

Discussion

Discussion

As closer to the centre a location is, the higher price is required for a property. Also, however boroughs with frequent restaurants are scattered, areas can be identified with “all day venues”. Also as more out of the centre we are located the lower priced stores and fast foods are located.





Conclusion

The method and results of this study can be used to identify potential areas to open a new restaurant.

The resulted maps can visualise and help the decision making of the potential business that can be open to fit within the context of the neighborhood, taking into account the required initial capital expenditure.