Analyzing boroughs in London for Starting a Restaurant

Applied Data Science Capstone Project

Boris Gorodilov

May 4, 2021

Introduction

- London is the capital and largest city of England and the United Kingdom. It is one of the world's most important financial, commerce and educational centers. London has a diverse range of people and cultures, and more than 300 languages are spoken in the region. Its estimated population is roughly 9 million, which made it the third-most populous city in Europe. If we are looking to open a new restaurant, this is one of the best cities to consider possible locations. This project can be useful for business owners and entrepreneurs who are looking to invest in a restaurant. The main objective of this project is to carefully analyze appropriate data and find recommendations for the stakeholders.

Data Collection

- **Borough geo coordinates data**

The data of the boroughs in London was scraped from https://en.wikipedia.org/wiki/List_of_London_boroughs.

- **Borough earnings data**

Information on the income of the population of the borough is collected on the basis of two sources: data on the income of taxpayers living in the borough https://data.london.gov.uk/dataset/average-income-tax-payers-borough, and data on the income of people working in the borough https://data.london.gov.uk/dataset/earnings-workplace-borough.

- **Geographical Coordinates**

The geographical coordinates for London data has been obtained from the GeoPy library in python.

- **Venue Data**

The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all boroughs in London and is used to study the popular venues of different boroughs.

Methodology

- The data on the venues will be used with K-Means clustering model to analyze different clusters of boroughs and determine the best location to start a restaurant business. Depending on the level of income of the working and living population, an adjustment function will be added to the cluster label of borough, to clarifying the attractiveness of opening a restaurant in this location.

Borough and Income data

Sample Borough data

	Borough	Area_sq_mi	Latitude	Longitude
0	Barking and Dagenham	13.93	51.5607	0.1557
1	Barnet	33.49	51.6252	-0.1510
2	Bexley	23.38	51.4549	0.1505
3	Brent	16.70	51.5588	-0.2810
4	Bromley	57.97	51.4039	0.0198
5	Camden	8.40	51.5290	-0.1250
6	Croydon	33.41	51.3714	-0.0970
7	Ealing	21.44	51.5130	-0.3080
8	Enfield	31.74	51.6538	-0.0790
9	Greenwich	18.28	51.4892	0.0648

Sample Income data

	Borough	Tax_payers	Workplace
0	Barking and Dagenham	23900	28553
1	Barnet	28700	32143
2	Bexley	26900	30733
3	Brent	24700	30134
4	Bromley	32000	29819
5	Camden	37300	38147
6	Croydon	27500	32109
7	Ealing	26700	30259
8	Enfield	26300	29134
9	Greenwich	27600	32635

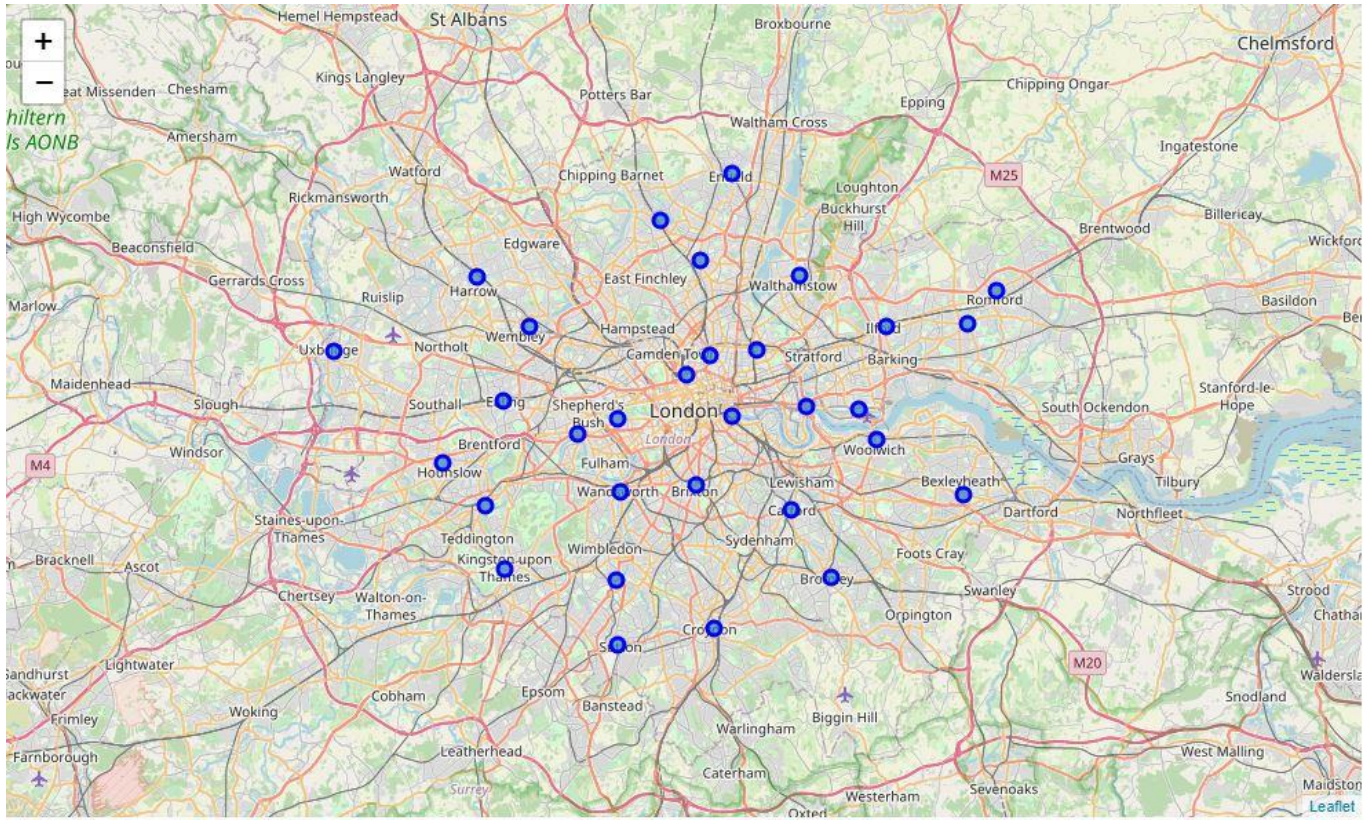
Venue data

Sample Venue data

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking and Dagenham	51.5607	0.1557	Central Park	51.559560	0.161981	Park
1	Barking and Dagenham	51.5607	0.1557	Lara Grill	51.562445	0.147178	Turkish Restaurant
2	Barking and Dagenham	51.5607	0.1557	The Eva Hart (Wetherspoon)	51.570460	0.130342	Pub
3	Barking and Dagenham	51.5607	0.1557	Costa Coffee	51.576890	0.179497	Coffee Shop
4	Barking and Dagenham	51.5607	0.1557	Harrow Lodge Park	51.555648	0.197926	Park
5	Barking and Dagenham	51.5607	0.1557	The Range	51.575550	0.180254	Furniture / Home Store
6	Barking and Dagenham	51.5607	0.1557	Hoo Hing	51.567561	0.135999	Grocery Store
7	Barking and Dagenham	51.5607	0.1557	Debenhams	51.579097	0.182720	Department Store
8	Barking and Dagenham	51.5607	0.1557	Ciao Bella	51.576103	0.182819	Italian Restaurant
9	Barking and Dagenham	51.5607	0.1557	Pets at Home	51.569605	0.183878	Pet Store

Map of London

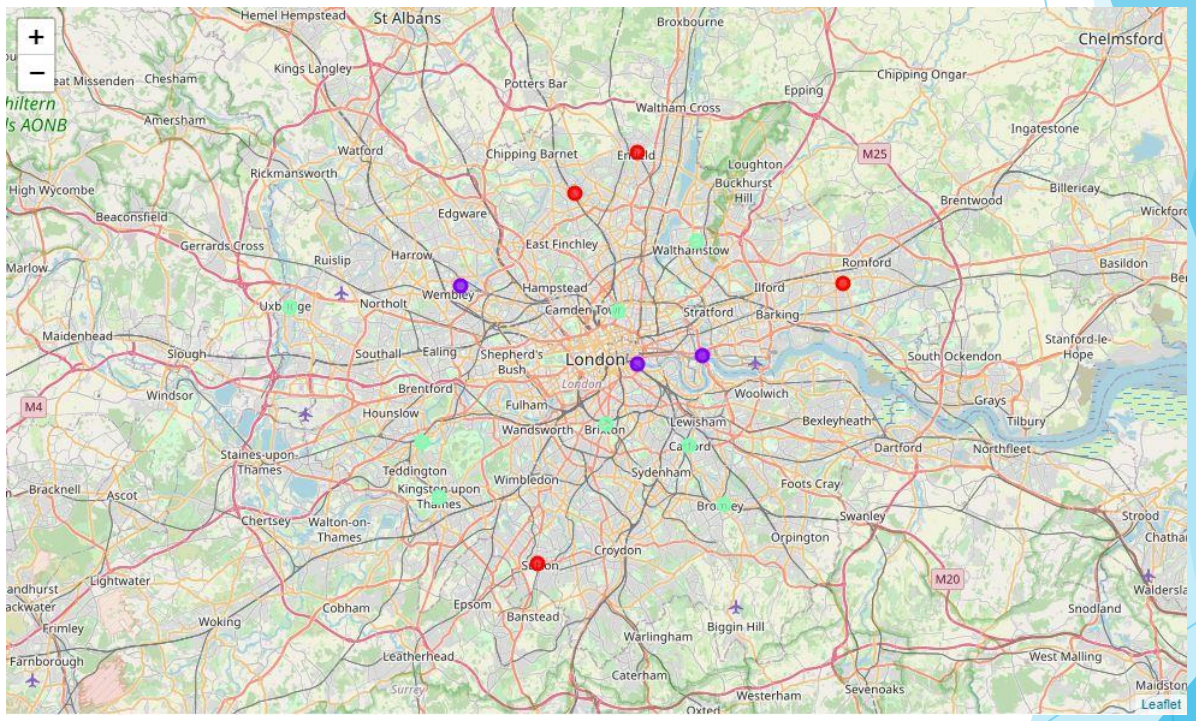
Depicting the boroughs spread across London



Visualizing the clustering of boroughs in London

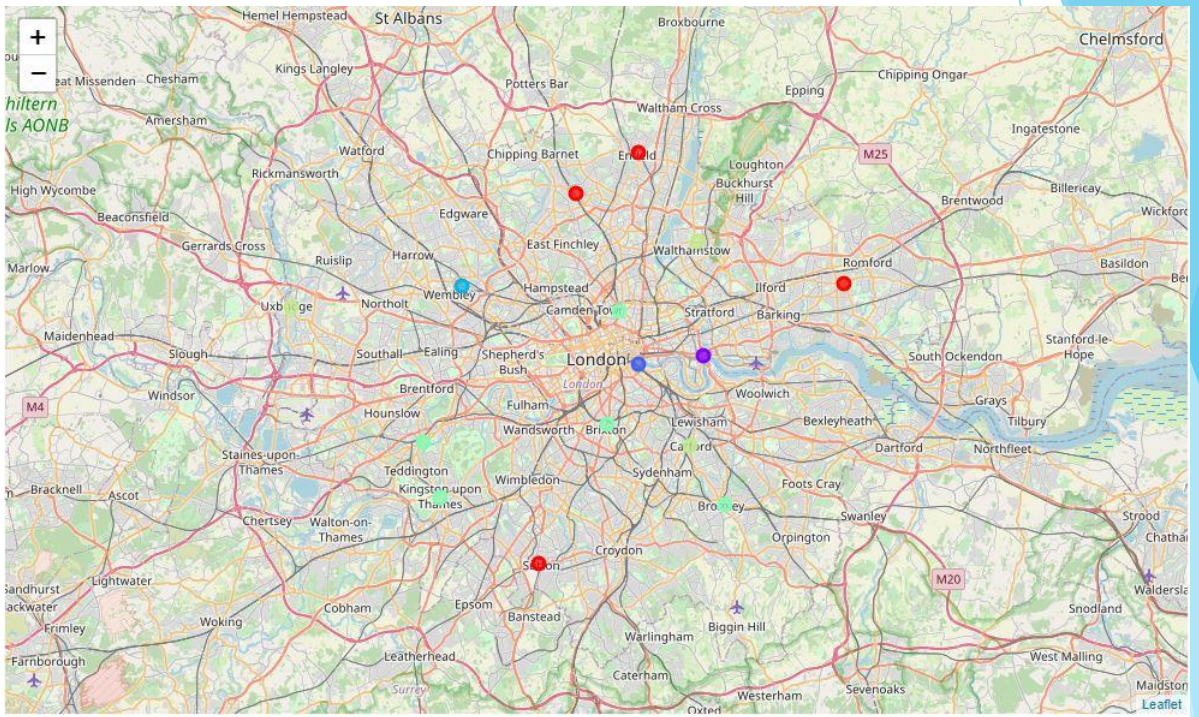
Boroughs from the second cluster (purple marks) have the greatest potential for opening a new restaurant, because in the top 3 categories of venues there are: no parks (dense development) and there are hotels (tourists and business travelers can significantly increase attendance).

On opposite side top-10 venues of boroughs from the first cluster (red marks) are made up of parks, supermarkets, furniture stores, grocery store and etc. This set of venues characterizes these boroughs as outskirts. Most of the venues are likely to be used by locals and suburbanites who drive through the area. This makes the opening of a new restaurant in these boroughs the least attractive.



Conclusion: Visualizing the ratings of boroughs in London

The results of the clustering processes based on venue and income data was combined under a single value for the final letter based rating ("A +" - most preferred, "C-" - least preferred).



Borough rating	Description
A+ A A-	Boroughs rated "A" are the best option to open a new restaurant, especially those rated "A +".
B B-	Boroughs rated 'B' may be considered for open a new restaurant, but they are not the best options.
C-	Boroughs rated 'C' should not be considered for a new restaurant opening. However, with a more detailed exploring and identification of points of attraction of foot-traffic, they can be a good option.