The Battle of Neighborhoods
Location for fast food restaurants in London

Coursera - IBM Data Science Specialization
Applied Data Science Capstone

The Ask/Problem Statement

Location

Location is decided a London, also considering how big the city is also Narrowed down to Inner London region. But need to find a optimal location in a 319 km²

Type of Restaurants

We are planning to a Fast Food restaurant.

Problem statement

In a Metropolitan like London there are many places restaurant to open a Restaurants, We want to make a Data driven decision to finalize the location of pick to Borough to start the Restaurant

Steps taken/Approach – High level

Collecting Data

Request and Pandas

Data is extracted from London Data Store and Wikipedia, then the indepth Location data from FourSquare.

Cleaning Data

Pandas

Unwanted columns are removed, columns are renamed to match the correct set and sorted to make it easy to process

Processing Data

SciPy and Folium

Required columns are merged, if needed additional data is added, required columns is grouped and weighted mean is calcualted

Result Conclusion

Pandas and Folium

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Implementation Details
Pre-Requirement

1. Need a Python 3.x environment
2. With the below mentioned python package
   a. Jupyter notebook
   b. numpy
   c. pandas
   d. folium
   e. matplotlib
   f. Scipy
   g. Requests
3. Working Internet connection as more than half the data will be pulled from Internet
4. wget in local machine

Implementation Details
Collecting Data

1. Started with pulling 'london-borough-profiles.csv' from London Data Store
2. Pulled the List of London Borough and Local Authorities from wikipedia - List_of_London_boroughs
3. Clean the unwanted columns and retain the required ones and merge both the Tables
4. Filter the data with only the Inner London Boroughs
5. Cleanup the Coordinates to correct Latitude and Longitudes
6. Collect below venues of interest from from FourSquare API
   - Two-year business survival rates
   - Population
   - Tech Startup
   - Supermarket
   - Soccer Field
   - School
   - Pub
   - Supermarket
   - Office
   - Historic Sites
   - Grocery Store
   - Factory
   - College Classroom
   - Coffee Shop
   - Café
   - Bus Stop
   - Fast Food Restaurant
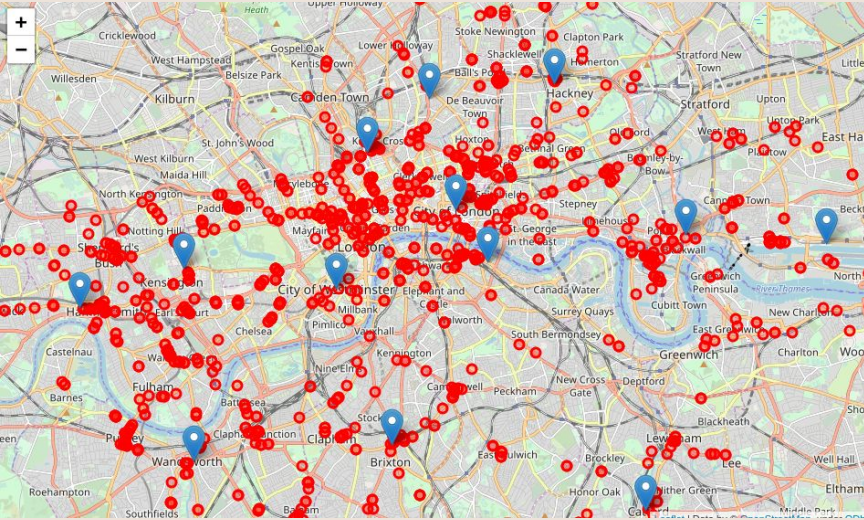   - Breakfast Spot
   - Sandwich Place
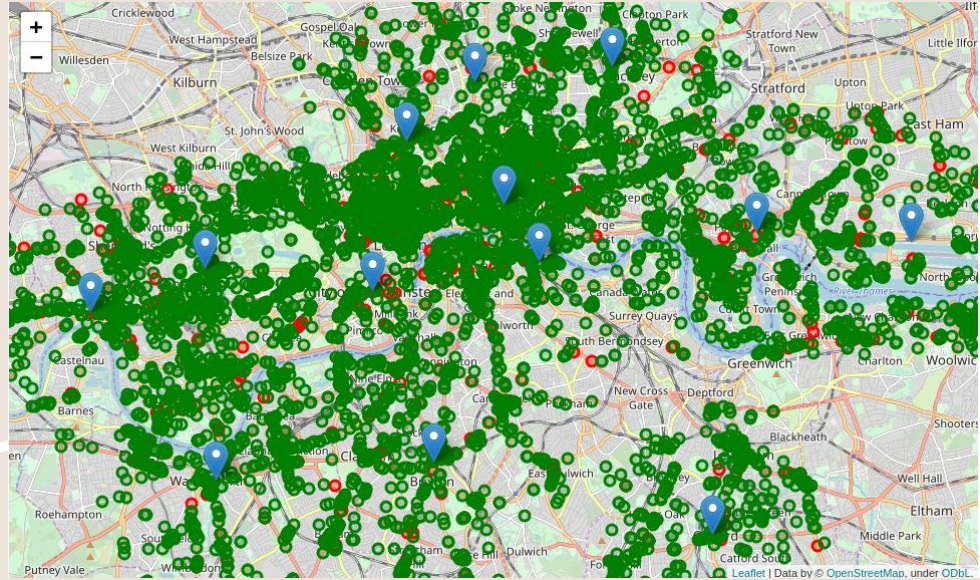

Boroughs picked for our Analysis

Implementation Details
Analysing Data

Let's begin analysis with visualizing the data and having a look at the data

Boroughs picked for our Analysis, with existing Fast Food restaurant

Boroughs picked for our Analysis with competition and potential neighbours

Implementation Details
Analysing Data

1. Let's start with consolidating all the existing data mapped to each of the Boroughs
2. Group them based on the name of locations based on the Borough
3. Then lets normalise the data so that the counts will overshoot when we normalise the data
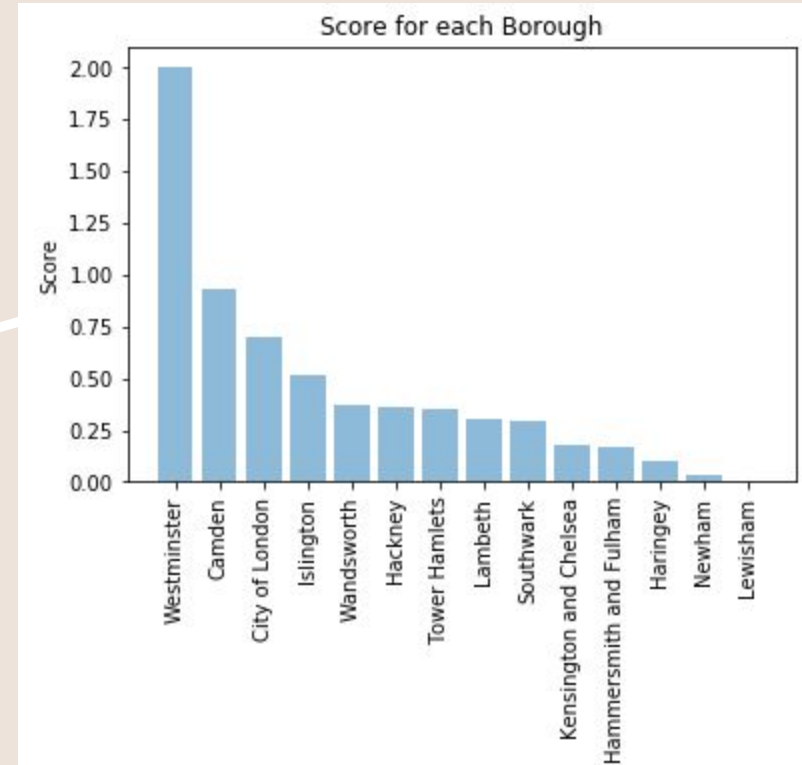
Normalised data

Tech Startup	Supermarket	Soccer Field	School	Residential Building (Apartment / Condo)	Pub	Professional & Other Places
0.870370	0.750000	0.8750	0.000000	0.555556	0.8	0.0
0.740741	1.000000	1.0000	0.222222	0.444444	0.6	0.0
1.000000	0.583333	0.1875	0.777778	0.666667	1.0	0.0
0.685185	0.333333	0.6250	0.666667	0.000000	0.2	0.0
0.000000	0.208333	0.1250	0.444444	0.111111	0.5	1.0
0.759259	0.875000	0.8125	0.444444	0.444444	0.8	0.0
0.685185	0.833333	0.3750	0.666667	0.444444	0.8	0.0
0.666667	0.375000	0.0000	0.888889	1.000000	0.1	0.0
0.018519	0.416667	0.0625	0.222222	0.111111	0.4	0.0
0.037037	0.250000	0.0625	0.777778	0.555556	0.0	0.0
0.851852	0.666667	0.6250	0.333333	0.777778	0.4	0.0
0.685185	0.291667	0.0625	1.000000	0.555556	0.3	0.0

	Borough	Tech Startup	Supermarket	Soccer Field	School	Residential Building (Apartment / Condo)	Pub	Professional & Other Places	Office	Historic Site	Grocery Store	Factory	College Classroom	Coffee Shop	Café	Bus Stop
0	City of London	56.0	30.0	17.0	6.0	48.0	50.0	0.0	69.0	30.0	57.0	78.0	42.0	65.0	32.0	50.0
1	Camden	49.0	36.0	19.0	8.0	47.0	48.0	0.0	77.0	31.0	52.0	82.0	43.0	51.0	43.0	50.0
2	Hackney	63.0	26.0	6.0	13.0	49.0	52.0	0.0	47.0	19.0	50.0	72.0	48.0	58.0	75.0	50.0
3	Hammersmith and Fulham	46.0	20.0	13.0	12.0	43.0	44.0	0.0	84.0	22.0	49.0	12.0	29.0	55.0	57.0	49.0
4	Haringey	9.0	17.0	5.0	10.0	44.0	47.0	2.0	61.0	16.0	44.0	17.0	5.0	51.0	49.0	50.0
5	Islington	50.0	33.0	16.0	10.0	47.0	50.0	0.0	76.0	36.0	53.0	80.0	48.0	55.0	56.0	50.0
6	Kensington and Chelsea	46.0	32.0	9.0	12.0	47.0	50.0	0.0	91.0	47.0	46.0	10.0	40.0	44.0	48.0	50.0
7	Lambeth	45.0	21.0	3.0	14.0	52.0	43.0	0.0	52.0	9.0	47.0	14.0	3.0	69.0	46.0	50.0
8	Lewisham	10.0	22.0	4.0	8.0	44.0	46.0	0.0	48.0	3.0	44.0	8.0	0.0	45.0	53.0	49.0
9	Newham	11.0	18.0	4.0	13.0	48.0	42.0	0.0	35.0	5.0	28.0	47.0	28.0	41.0	52.0	50.0
10	Southwark	55.0	28.0	13.0	9.0	50.0	46.0	0.0	76.0	35.0	55.0	70.0	43.0	61.0	38.0	50.0
11	Tower Hamlets	46.0	19.0	4.0	15.0	48.0	45.0	0.0	99.0	13.0	43.0	54.0	49.0	48.0	41.0	50.0
12	Wandsworth	61.0	12.0	10.0	15.0	46.0	48.0	0.0	64.0	11.0	48.0	14.0	15.0	48.0	59.0	50.0
13	Westminster	59.0	27.0	17.0	10.0	48.0	46.0	0.0	73.0	33.0	50.0	32.0	47.0	51.0	47.0	50.0

Implementation Details
Analysing Data

1. Then a weighing amount is applied on the each category based on the importance that we feel
2. Weighing amount for Competition in that same are also applied but are given negative score
3. Normalised the weighted mean some for all the categories are grouped by Borough are sorted and plotted in a Bar-chart
4. We can see a clear trend that the prefered location is Westminster

Conclusion

As the Data is so obvious when we do a weighted score of the Boroughs we have not proceeded with K-Mean algorithm. When we add more parameters like, Restriction and benefits in each of Borough Council, Fresh data on the transport and tourists stops, Safety in each of the council , Cost of running in each Council.

Borough of Westminster is the best place to start a Fast Food Restaurant