IBM Data Science Capstone Project

Picking the right location for a new Pretzel store in London

Thomas Rüedi

June 9th, 2020

Introduction

The IBM Data Science Professional certificate concludes with a Capstone Project. This project is about using the data science toolset on a theoretical real-life problem and demonstrating the creation of value by applying the learned skills. The analysis was performed in Python.

Problem Definition

For this capstone project, I chose the following problem statement:

A famous Pretzel chain wants to open a new store in London. Data science shall help to identify where to open the new store.

Audience

The outcome of this analyzes is not only interesting for the hypothetical Pretzel chain management, but also for any restaurant owner, as the analysis will provide valuable insights as I will perform a k-means cluster analysis on food places in London.

Data

The following data sources were used to perform the analysis:

1) List of neighborhoods of London (https://en.wikipedia.org/wiki/List_of_areas_of_London)
2) Geo-coordinates of the neighborhoods of London
3) Foursquare venue data

Methodology

Data preparation & exploration

First the wiki data had to be cleaned and the postcodes had to be split & expanded. Then the geo-coordinates could be added using geocoder.

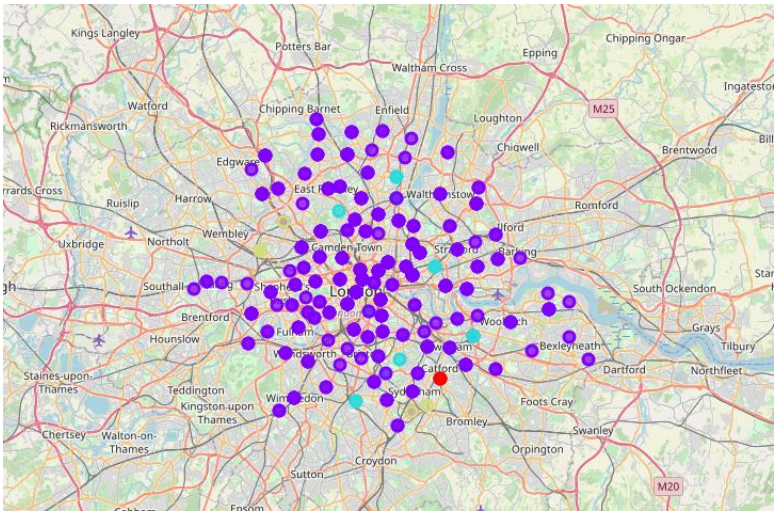
	Neighborhood	London borough	Postcode	Latitude	Longitude
0	Abbey Wood	Bexley, Greenwich	SE2	51.49245	0.12127
1	Acton	Ealing, Hammersmith and Fulham	W3	51.51324	-0.26746
2	Acton	Ealing, Hammersmith and Fulham	W4	51.48944	-0.26194
3	Aldgate	City	EC3	51.51200	-0.08058
4	Aldwych	Westminster	WC2	51.51651	-0.11968

Subsequently the venue data was pulled in from foursquare. As the analysis only focused on food places, the venue data was filtered accordingly. Based on this data I analyzed the top 10 most common venue categories for each of the neighborhoods.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abbey Wood	Coffee Shop	English Restaurant	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop	Dim Sum Restaurant
1	Acton	Pub	Italian Restaurant	Coffee Shop	Bakery	Café	Portuguese Restaurant	Breakfast Spot	Indian Restaurant	English Restaurant	Dessert Shop
2	Aldgate	Restaurant	Salad Place	Coffee Shop	Cocktail Bar	English Restaurant	Italian Restaurant	Pub	Wine Bar	Beer Bar	Asian Restaurant
3	Aldwych	Pub	Sandwich Place	Café	Japanese Restaurant	Restaurant	Coffee Shop	Korean Restaurant	Bakery	Wine Bar	Italian Restaurant
4	Anerley	Fast Food Restaurant	Xinjiang Restaurant	English Restaurant	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop

Clustering

I selected k-means to perform the clustering and used a function to determine the optimal value for k. In a subsequent step the neighborhoods got broken out into 4 clusters. I used folium to visualize these clusters on a map.



Results

Below you can find an a few on each of the four clusters. By looking at this data, we can see that cluster 2 is the one that we are most interested in.

Cluster 0

	London borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
29	Lewisham	0	Turkish Restaurant	Xinjiang Restaurant	Donut Shop	Coffee Shop	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega
66	Lewisham	0	Turkish Restaurant	Xinjiang Restaurant	Donut Shop	Coffee Shop	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega
295	Lewisham	0	Turkish Restaurant	Xinjiang Restaurant	Donut Shop	Coffee Shop	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega

Cluster 1

	London borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Bexley, Greenwich	1	Coffee Shop	English Restaurant	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop
1	Ealing, Hammersmith and Fulham	1	Pub	Italian Restaurant	Coffee Shop	Bakery	Café	Portuguese Restaurant	Breakfast Spot	Indian Restaurant	English Restaurant
2	Ealing, Hammersmith and Fulham	1	Pub	Italian Restaurant	Coffee Shop	Bakery	Café	Portuguese Restaurant	Breakfast Spot	Indian Restaurant	English Restaurant

Cluster 2

	London borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
37	Lewisham	2	Pub	Xinjiang Restaurant	Donut Shop	Coffee Shop	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega
44	Tower Hamlets	2	Pub	Coffee Shop	Burger Joint	Bar	Donut Shop	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint
53	Tower Hamlets	2	Pub	Coffee Shop	Burger Joint	Bar	Donut Shop	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint

Cluster 3

	London borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Bromley	3	Fast Food Restaurant	Xinjiang Restaurant	English Restaurant	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop
22	Bromley	3	Fast Food Restaurant	Xinjiang Restaurant	English Restaurant	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop
23	Bromley	3	Fast Food Restaurant	Xinjiang Restaurant	English Restaurant	Colombian Restaurant	Comfort Food Restaurant	Creperie	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop

Discussion & Recommendations

Cluster 1 & cluster 2 stand out meeting our requirements for the new Pretzel store. Pubs are very dominant in these neighborhoods.

Cluster 0 as well as cluster 3, are lacking on these.

The recommendation will be to go with cluster 2 as it fits most with the client's requirements of the new location.

Conclusion

This paper discussed the process of coming up with an answer for a hypothetical though real-life like business problem. The analysis had been performed based on the toolset of data science and relied heavily on the use of Python and its libraries such as Pandas, Scikit, or Folium.

References

The Jupyter notebook of the analysis can be found on GitHub:

https://github.com/thomasr1979/Coursera_Capstone/blob/master/Coursera_IBM_Data_Science_Capstone.ipynb