Analyzing Locations for a New Bank in London, Ontario, Canada

Wacey Schell

April 14, 2020

1. Introduction

For my capstone project I will be looking at different neighbourhoods in the city of London, Ontario and deciding the best location to open a new bank. I will be looking at the geographical layout of the city, splitting the neighbourhood up into clusters and comparing each cluster's current need for a new bank. Identifying the best place in a community to open a new bank can be mutually beneficial for both a bank scouting out potentially successful locations and a community lacking access to more banking options.

It's easy to understand the importance of a bank's location to the bank itself. Clearly, the customers and thus a bank's profits can be greatly affected by the location of a bank. What can be a bit less obvious is how important banks can be to a community. A bank and all of its services can be incredibly important to the structure of a healthy community economy. Even during a time where day to day banking needs are being taken care of digitally, a brick and

mortar location can still serve its own purposes for customers, and give options to those who may not get them through other mediums. Increasing community access to brick and mortar banking locations increases the access households have to credit products that can be used to build wealth or savings accounts to invest in the future. Alternatively, without access to basic financial services, communities can become a target for more predatory, high-cost credit options.

1.1 Interest: Obviously, financial institutions looking to open new locations, which could include both new banks or credit unions looking to open brick and mortar locations in communities, as well as existing banks that are looking to add locations in a city or town in which they already reside. Along with banks, governments-local and beyond-may be interested in studying neighbourhoods that are under-served in terms of banking and how that impacts these neighbourhoods and whether there is a trend in neighbourhoods that banks under-serve.

2. Data

To solve the problem we will use the following data:

- list of neighbourhoods in London, Ontario. We chose to look specifically at the city of London, Ontario to ask the question of where would be the best location for a new bank. London is a medium sized city located in southwestern Ontario, surrounded by rural communities, both of which can be really affected by the (lack of) presence of banking locations.
- latitude and longitude coordinates of those neighbourhoods which will allow us to plot the map and obtain venue data, such as number of banks in each neighbourhood

- venue data, such as the number of banks in each neighbourhood, which we will use for clustering the neighbourhoods and ultimately deciding on the neighbourhoods in most need of another bank-and thus the best location for a bank

This project will make use of several data science skills, including web scraping, data cleaning, data wrangling, machine learning techniques such as K-means clustering, Foursquare API, and map visualization with Folium. We will take a look at the neighbourhoods in London, Ontario and group them together using k-means clustering. After they are grouped together into k groups, we will analyze these clusters and find which neighbourhood clusters have the least amount of banks, signally an opportunity for a new location.

The Wikipedia page:

https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_London,_Ontario contains a list of neighbourhoods in London Ontario, with a total of 9 neighbourhoods. I will use web scraping techniques from python and beautiful soup packages to extract the data from the wikipedia page containing all of the London neighbourhoods. I will then get the latitude and longitude coordinates of the neighbourhoods with the Python Geocoder package.

Once I've got the geographical coordinates of the neighbourhoods I will use the Foursquare API to get the venue data for the London neighbourhoods, specifically the number of banks in each neighbourhood, which will allow me to identify which area has the least amount of banks, and thus the greatest need for another brick-and-mortar banking location.

3. Methodology

We will start out by getting a list of neighbourhoods in London, Ontario. We will obtain this list from the Wikipedia page noted above in the data section. We will scrape the data, which in this case is the list of neighbourhoods, from the Wikipedia page using Python and Beautiful Soup packages. In

addition to the list of neighbourhoods in London, Ontario, we will also need the latitude and longitude of the neighbourhoods in order to use Foursquare API. To get these geographical coordinates we will use the Geocoder package that converts addresses into latitude and longitude. Once we have gathered this data we will then put it all together into a Pandas Dataframe. From there we will be able to visualize this data on maps using a Folium package. We can use this step to do a quality check on our data, ensuring that everything checks out and our previous steps were done correctly. At this point we will have a map of London with each of its neighbourhoods marked by a dot.

Once we have the map of London with its neighbourhoods we will take advantage of Foursquare API's services to get the top 100 venues within a 2 km radius. After setting up a Foursquare developer account, we can use our own personal ID and secret key to make calls using the geographical coordinates of the neighbourhoods in a Python loop. With the help of Foursquare, we can start to look at each neighbourhood and see what kind of venues each neighbourhood consists of, taking note of banks in particular for this project. This step will set up for later on when we start clustering the neighbourhoods, filtering banks as the clustering category for the neighbourhoods which will aid us in figuring out whether or not neighbourhoods within a given cluster is adequately served or underserved in terms of banking.

Finally, we will perform one of the simplest and most popular unsupervised machine learning algorithms called k-means clustering. K-means clustering partitions data points into k clusters. In which each data point belongs to a cluster with the nearest mean. We will cluster the neighbourhoods of London into three clusters based on their frequency of occurrence of banks. This will allow us to easily visualize and identify which neighbourhoods have a higher concentration of banks and which have a lower concentration. From this we will be able to speculate the best location for a new bank to open up, serving the neighbourhoods with a lower amount of banks available to them.

4. Results

The results from running k-means clusters with these neighbourhoods shows us that we are able to sort the neighbourhoods in London, Ontario into three distinct clusters based on the number of banks that reside within them. Cluster 0 (red) and cluster 2 (lime green) have a medium to high number of banks, while cluster 1 (purple) has little to no banks in its neighbourhoods (see map below).
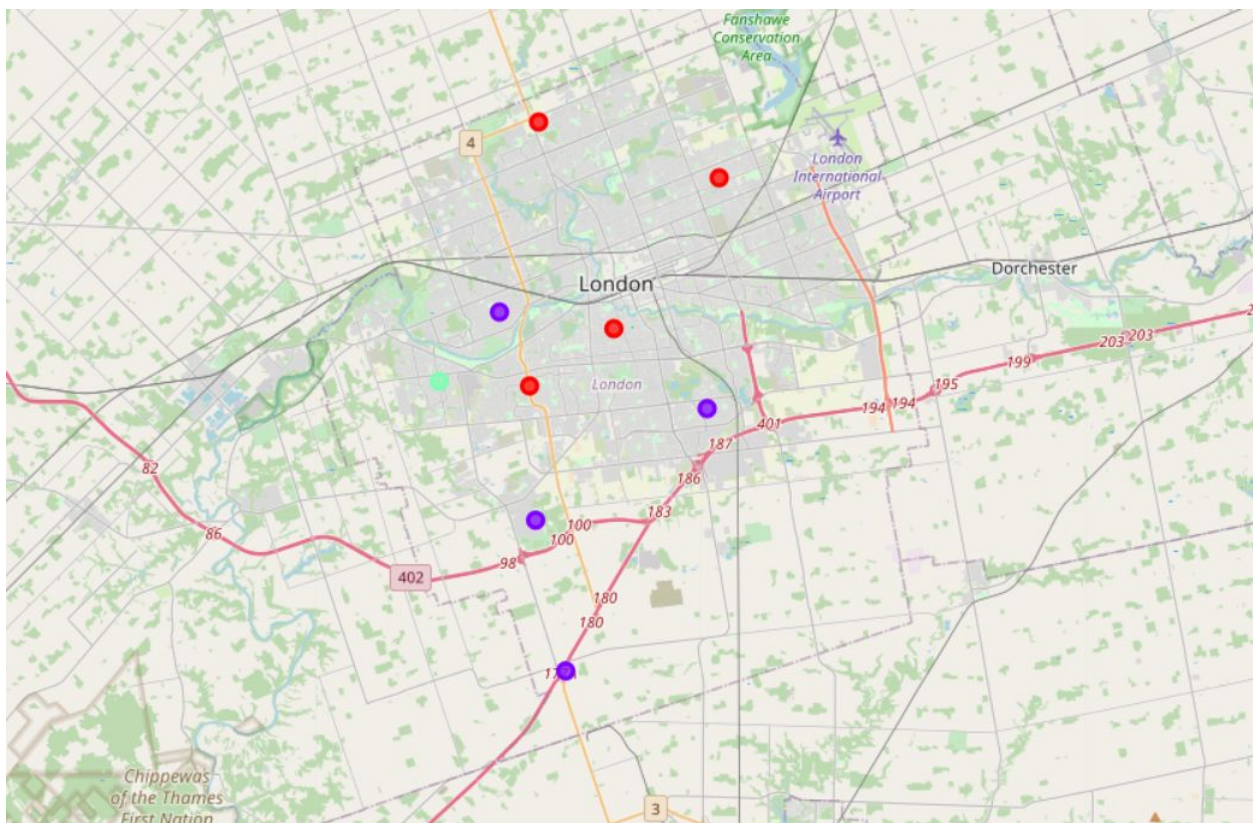
5. Discussion

Most of the banks are concentrated in the central and north neighbourhoods of London, with the highest number in cluster 0 (red) and cluster 2 (lime green). On the other hand, cluster 1 (purple)

has a lower number to totally no banks in its neighborhoods. This represents an opportunity for a new bank to open a branch as our model indicates that the neighbourhoods in this cluster are under-served in terms of banking and there would be minimal competition from existing banks. Of course cluster 1 includes two neighbourhoods on the southern perimeter of London, which may include less dense housing/lower population than inner city neighbourhoods. That being said, adding a location within this cluster may still be a great opportunity for a bank because it could count on some potential business from rural communities outside of the city on their way into London along the 401 and 402 highways, which are two of Ontario's largest highways.

5.1 Recommendations

With only 9 neighbourhoods to cluster, London, Ontario may not be the most ideal city to study for k-means clustering. Of course, doing this type of analysis in a larger city with more neighbourhoods would give us more data to sort and we could start to come up with greater relationships between cities within a given cluster. That being said, it is still important to do these types of analysis on smaller towns and cities to reveal any patterns that may be helping or hindering its neighbourhoods.

5.2 Further Research

We could pair this research with additional data on each neighbourhood to make further insights on systemic inequalities. For example, we could look at data on socio-economic status or mean income by neighbourhood to see whether there's a trend in which neighbourhoods are being under-served in terms of financial institutions. This could help us uncover any systemic inequalities or barriers that may be present within the city, and need to be addressed. We could also take a look at other types of venues like predatory high-interest credit options within lower income neighbourhoods, or low income housing projects and see if there's a (negative correlation) pattern between the occurrence

of these venues and the occurrence of banks. Clearly, this could be incredibly beneficial for a city to look into in an attempt to address issues of inequality.

6. Conclusion

In this study, I analyzed the neighbourhoods of London, Ontario, Canada to determine the answer to the business question of where the best location to open a bank would be. I sorted the neighbourhoods of London into similar groups with respect to the occurrence of banks within them using a simple unsupervised machine learning algorithm, k-means clustering. Once this algorithm was run, it was easy to see which neighbourhoods had little to no banks available to them, which I decided to take as the main indicator for the best place for a new bank to open doors. According to our analysis, cluster 1 along the south part of London (see map above) was found to have the least amount of banks, and thus opening a bank within this region would be the most beneficial.

References:

Category: Neighourhoods in London, Ontario. Wikipedia. Retrieved from

https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_London,_Ontario

Foursquare Developers Documentation. Foursquare. Retrieved from

https://developer.foursquare.com/docs