Wesley Layton
April 6, 2020

Identifying the Best Area in Toronto for a New Pharmacy

1. Introduction

1.1 Background
In the city of Toronto, with an approximate population of 6,341,935, there is a need for advanced and expanding health infrastructure. Within healthcare, pharmacy and prescription drugs are at the forefront of discussion with rising costs.

1.2 Problem
With barriers already in the form of costs, it is important for the health of a population to be able to access these drugs conveniently and adhere to the guidelines given. Because of this, it is important that new pharmacies are placed where they can be most effective for the most people, but this can be difficult to identify for potential pharmacy owners.

1.3 Interest
This information could be key for any investors that are looking to open an independent pharmacy, or for recent pharmacy graduates that are also looking to open and run their own independent pharmacy. By contextualizing the neighborhoods in terms of the density of pharmacies and neighborhood population, they will be able to make decisions to best serve the greater Toronto area and its citizens.

2. Data Acquisition and Cleaning

2.1 Data Sources
I will use Foursquare location data and web scraped postal codes from Wikipedia to be able to analyze neighborhoods and their pharmacy density. This data will be drawn through Foursquare's API by defining the search parameters for each neighborhood as nearby pharmacies and then joining this to our postal codes for the Toronto area.

I will also be using population data from the 2016 census that is available through the Statistics Canada government website in CSV format (https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&S=22&O=A). This data contains raw population numbers for each postal code and will also be joined to our data frame so that we are able to contextualize the need for pharmacies as it is compared to the total population of the postal code.

2.2 Data Cleaning
Data used was joined to one table that represented each Toronto neighborhood, its geospatial coordinates, population data, and the count of pharmacies within it. There were several steps to get the data to this combined table. First, the web-scraped Wikipedia table contained all Canada postal codes. Because of this, any that were not assigned or not within Toronto had to be dropped. Also, any with a blank neighborhood value were filled to contain the name of the borough. The Foursquare venue data came cleaned and ready through the API, although ultimately most columns returned were not used. Neighborhoods were then grouped into clusters through K-Means. Lastly, a new column that divided the number of pharmacies by the population was used to be able to identify the most misrepresented clusters.

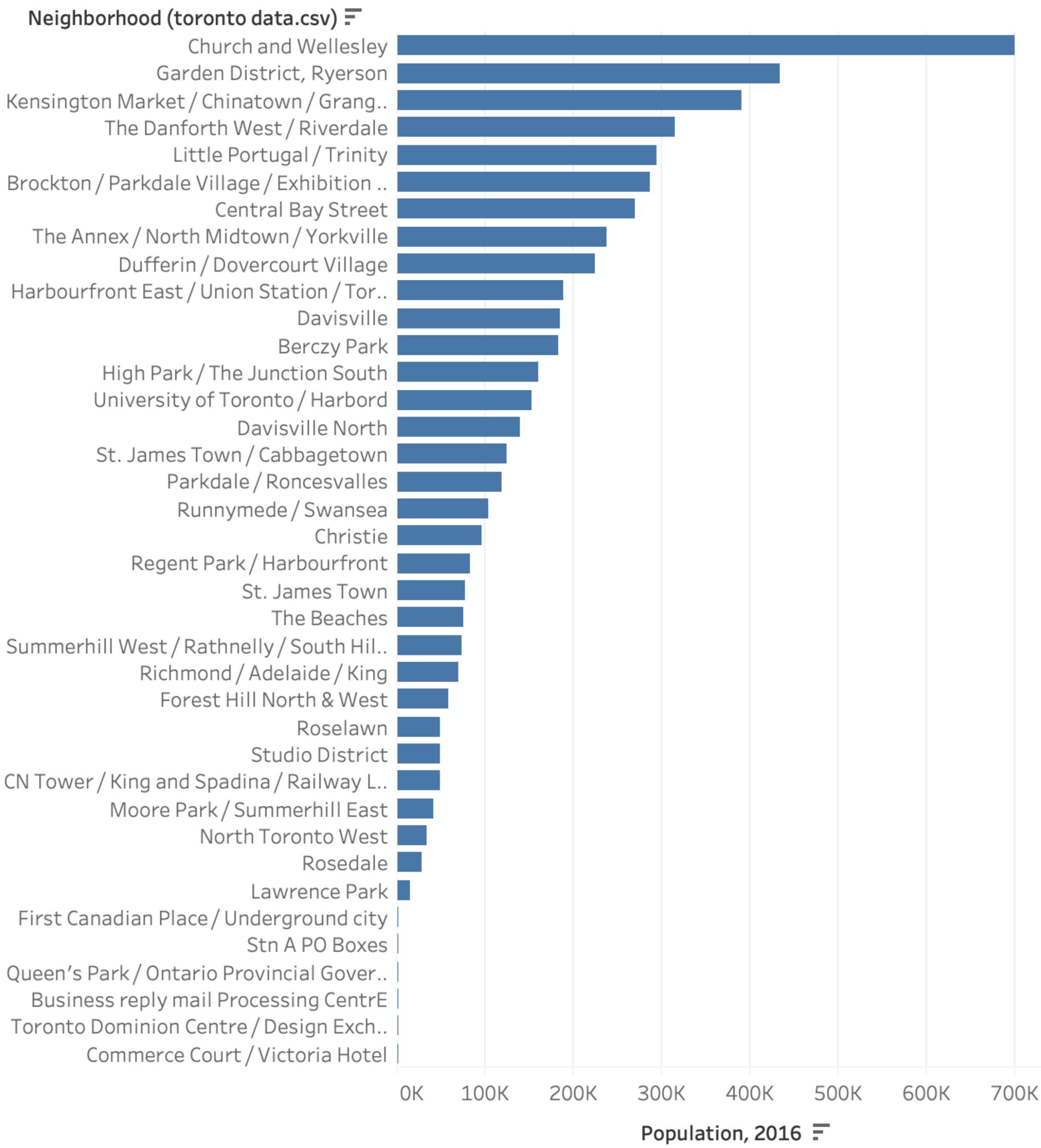
2.3 Feature Selection
After data cleaning, there were 38 neighborhoods with a count of pharmacies in them. Since these ranged greatly in size, joining them to clusters was able to give a more fair comparison for the areas. Also, it was clear that a simple count of pharmacies was not going to be the best indicator as the clusters still varied greatly in population sizes. To combat this, a new feature of pharmacies per population was created by dividing the number of pharmacies in a cluster by the total population. These were ultimately the final two features that were used in our analysis.

3. Methodology

3.1 Exploratory Analysis

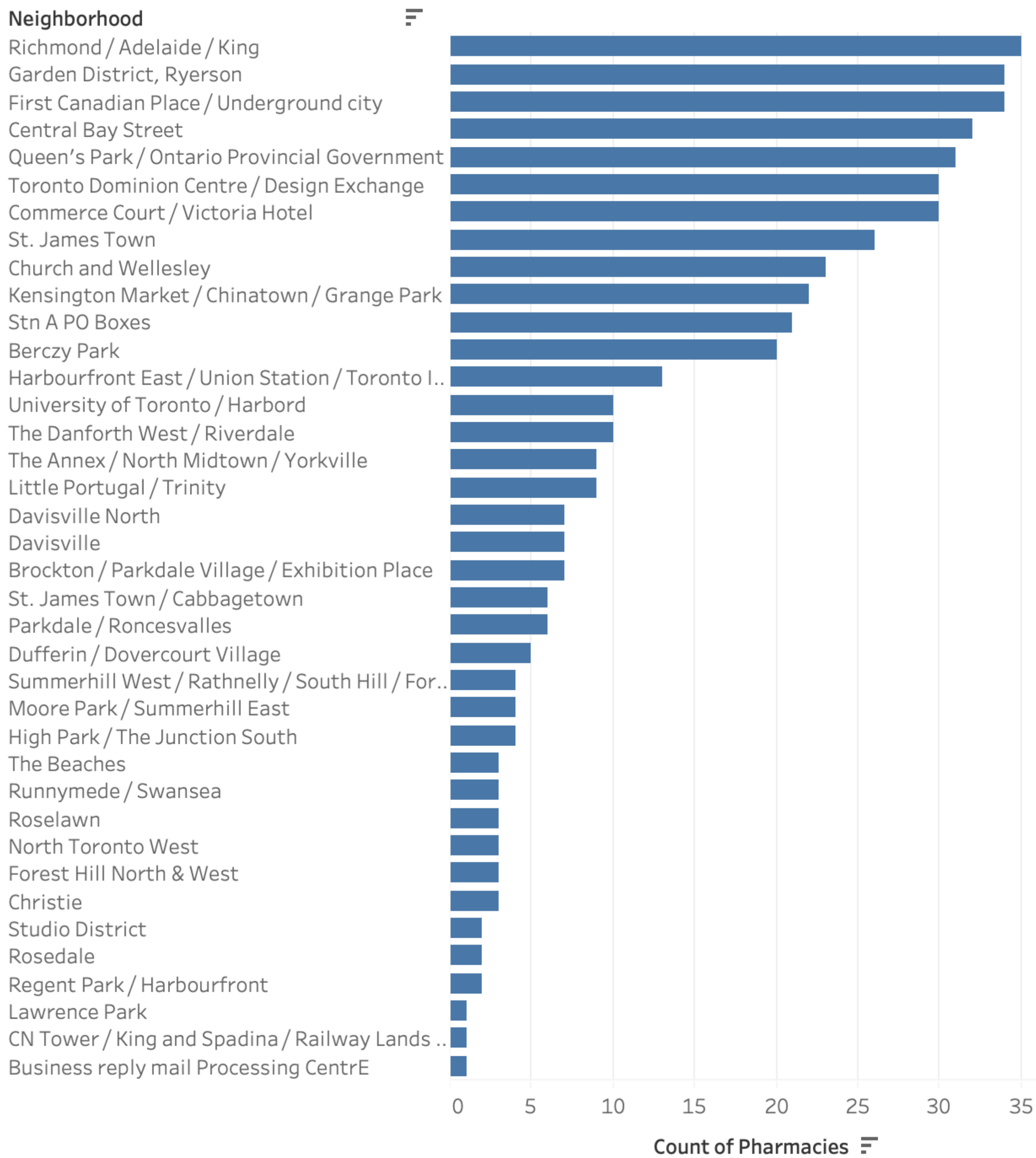
3.1.1 Population of Neighborhoods

Once we had our data loaded in, I began by examining the population of our neighborhoods. In the bar chart below we can see where our most heavily populated areas are.



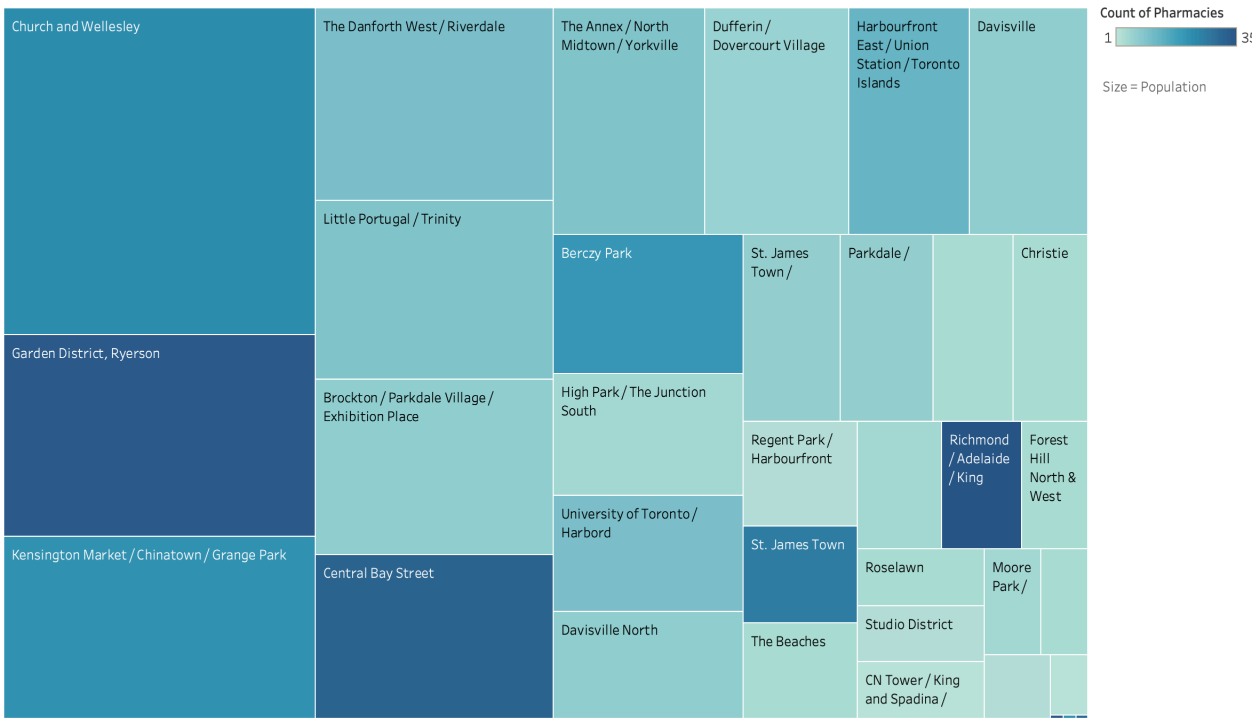
3.1.2 Pharmacy Counts in Neighborhoods

Another important aspect I looked at as I explored my data was the amount of pharmacy venues in each neighborhood. Below we can see those that are most heavily filled with pharmacies. Notice that this does not necessarily mean that these are our most populated neighborhoods. These will be neighborhoods that we would want to avoid starting a new pharmacy in as the population will already have a regular pharmacy they use, and the efforts could better serve an underserved area.



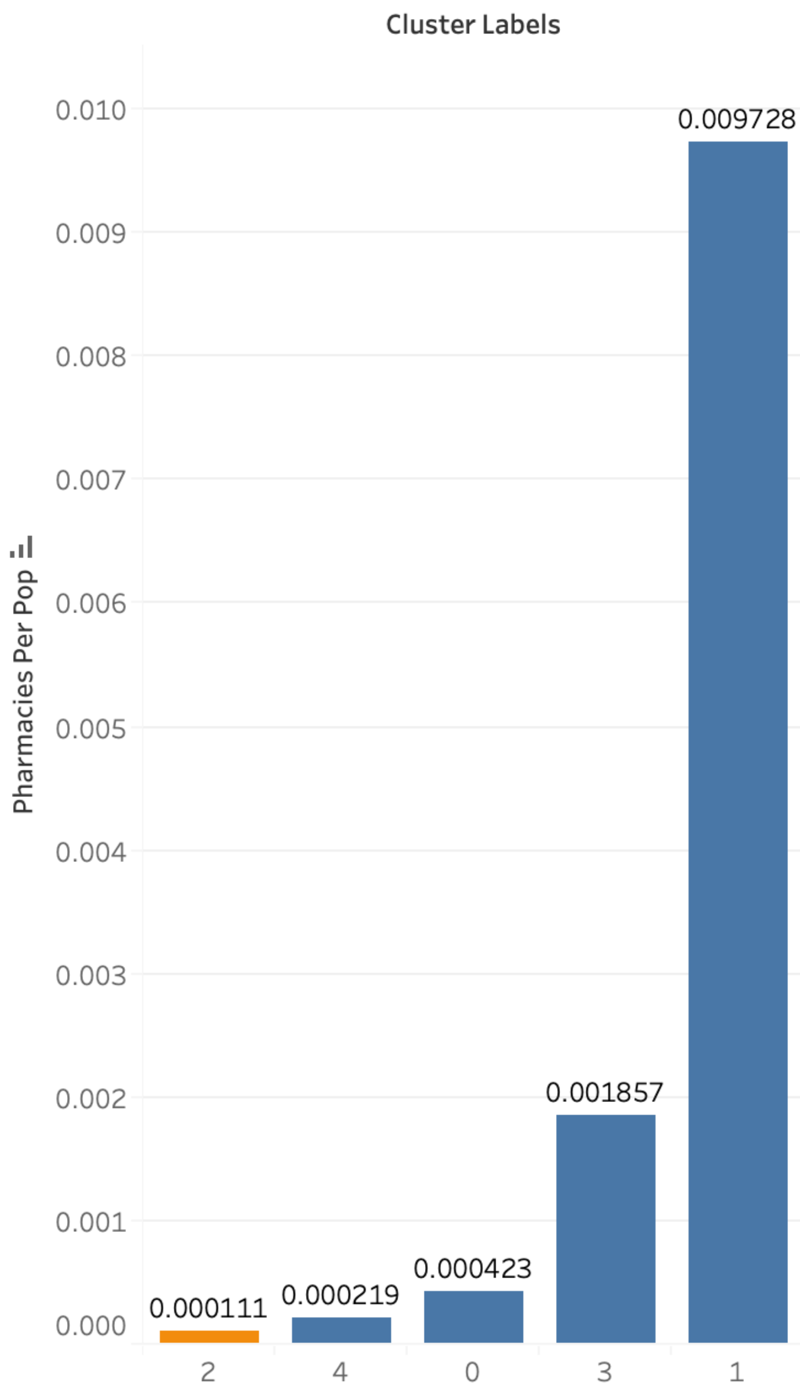
3.1.3 Pharmacy Density and Population Treemap

Ultimately, combining the two previous measures is where any insight can be drawn. By comparing the population size with the density of pharmacies, we can find where there is an area that is misrepresented based on its size. In the below chart, size is relevant to the population and the color scale corresponds to the number of pharmacies in the neighborhood. The neighborhoods that would be best served by new pharmacies are those where the size is larger but the color scale is lighter, which indicates a low rate of pharmacies per person.



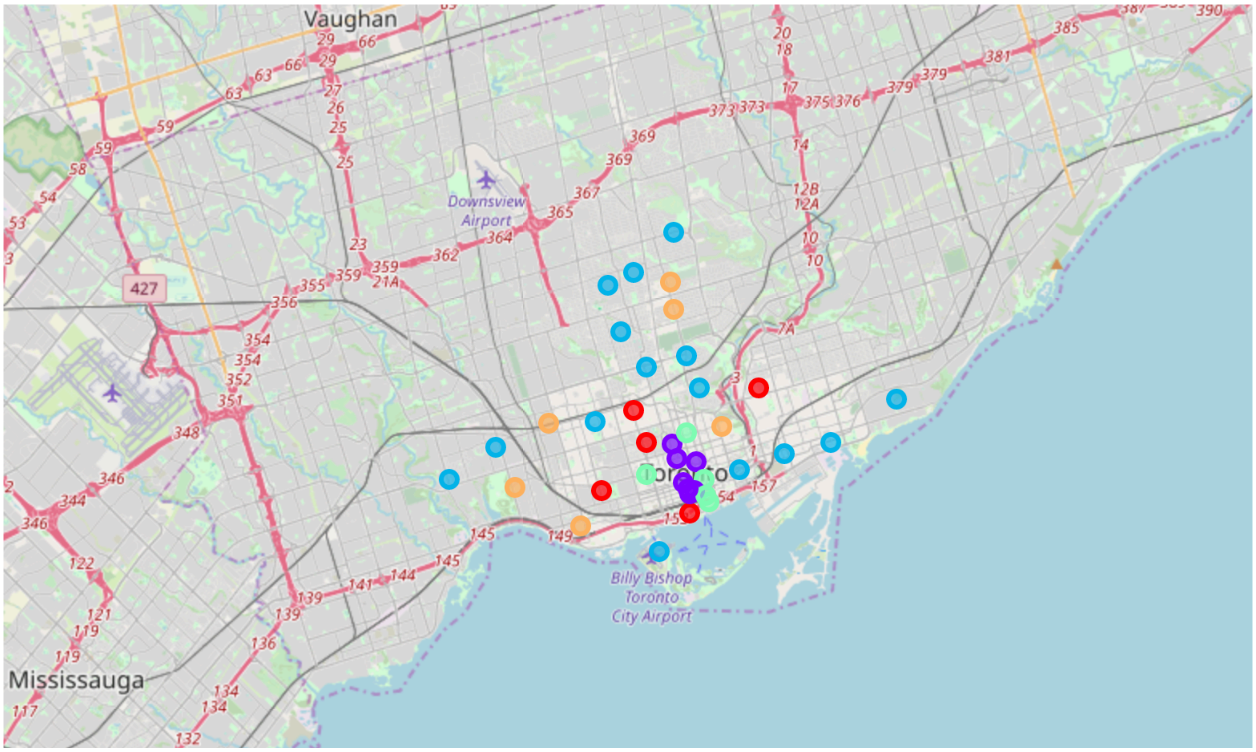
3.1.4 Pharmacy per Population by Cluster

This is where we finally land for our meaningful conclusion. After K-Means clustering, we have grouped the neighborhoods into 5 clusters based on similarity. From here we can look at the rate of pharmacies per population to determine which of these would be best served by a new pharmacy.



3.2 K-Means Clustering

To better generalize our neighborhoods, we used K-Means clustering as a tactic of grouping together similar neighborhoods. By doing this we can provide an at-large or overhead view for stakeholders on the best area to place a new pharmacy. The below map shows our clusters by color grouped out and how they lie in relation to each other.



4. Results

The results from our analysis show that cluster 2 has the lowest rate of pharmacies per population and would have the greatest need in the area for a new pharmacy. Cluster 1 is heavily populated with pharmacies per population in comparison to the rest of the areas, and Cluster 4 and 0 are relatively low as well and would also be a good location for any possible store-owner.

5. Discussion

As you can tell from the map, the most-dense area of pharmacies is located in the heart of downtown Toronto. However, this area also has the lowest population. Clusters 2, 4, and 0 are those that would be most advantageous for a new store-owner or investor to fund for a pharmacy. It is important to note that these clusters span a large area on the outskirts of downtown Toronto, and there are potentially several locations that would be effective. In fact, many of the neighborhoods analyzed in these clusters have nearly identical rates for pharmacies per population.

6. Conclusion

In this project, we have gone through the process of identifying the business problem, extracting and cleaning the data, performing deep level exploratory analytics, and deploying machine learning methods for clustering neighborhoods. Through our efforts we can answer the root question for any stakeholders, which is where in Toronto would be the most effective location for a new pharmacy to best serve the population. As we found in our final results, a neighborhood in Cluster 2 would be most effective for serving the population. This cluster represents the lowest rate of pharmacies per population and would better increase the area's access to timely and quality medications to help better increase the ease and compliance of patients in the area.