Introduction:

Substance usage has becoming a dangerous problem among the nations. According to the National Survey on Drug Use and Health, around 40.3 million Americans were struggling on substance usage in 2020 (drugabuse.com). Similarly, approximately 6 million Canadians (21% of the overall population) struggle with substance usage and addiction according to Canadian Mental Health Association (ontario.cmha). When the problem controls over a big portion of the population, the negative externalities on other people due to substance users’ unconscious behaviours as well as the impact on the country’s economy, health, crime rates, and legal issues -causing approximately 600billion every year(addictioncenter.com) - are unavoidable.

Therefore, in order to prevent an increase on the substance usage or even to try decreasing it among the people, treatment centers and treatment availability are extremely important. According to Addiction Center statistics, only 10% of addicted Americans receive treatment. (addictioncenter.com) So, it is important to further investigate and assess the availability of substance treatment.

In this paper, I will use substance use treatment locations in Toronto to further analyze whether the availability of treatment centers correlate with any other factors or not. For this analysis, I will pull treatment centers’ and neighbour characteristics datasets in Toronto from Toronto Open Data portal to look at their frequency and distribution depending on the neighborhoods. I will then look into neighbourhood characteristics which have treatment centers in the region and investigate more on potential factors which might affect the availability of treatment centers within the city. I will process my analysis in R, using mainly tidyverse (Wickham et al. 2019) and dplyr (Wickham et al. 2021) packages and will generate the output on R Markdown. The paper will proceed Data and References sections that I will be talking about data sources, data characteristics, data analysis, and further analysis on other potential contributing factors.

Data:

To investigate the accessibility and frequency of treatment centers in Toronto, I used Wellbeing Youth – Substance Use Treatment data (source) from the Toronto Open Data Portal. The data was last updated on March 16, 2020 and published by Social Development, Finance & Administration. The raw data includes 56 different agent and agent-related information out of 29 neighbourhoods in Toronto. Agent-related information consists of agency name and address, neighbourhood name, phone number, e-mail, website, eligibility, description and application for the treatment centers, languages used, accessibility, working hours, legal status, geographical location, postal code, and municipality. I extracted, cleaned, analyzed, and graphed my data using R (citation), tidyverse (citation), dplyr (citation), janitor (citation), tidyr (citation), knitr (citation), ggplot2 (citation), sf (citation).

Before selecting the columns I want to use, I cleaned my data and removed any “n/a” parts. In order to see the frequency of treatment centers in different neighbourhoods, I only used neighbourhood column to see the total number of treatment centers counting them with the count function according to the neighbourhood they belong to. In order to continue my analysis neatly, I also extracted the numerical part attached to each neighbourhood name through select function. That way I only included the name of each neighbourhood. In order to better see the distribution, I graphed my counted neighbourhood data using ggplot.

(BURAYA İLK GRAPH)

Although most of the neighbourhoods have 1-2 treatment centers on average, the neighbourhoods Church - Yonge Corridor, Niagara, and Trinity-Bellwoods have significantly higher number of treatment centers.

In order to visualize this difference I also included a data table of the counts in my analysis.

(BURAYA DATA TABLE)

Before moving to further investigation on the neighbourhood characteristics which might be explanatory on such difference, I also used ggplot2 to draw a map of Torontonian neighbourhoods with treatment centers and see if there are any clusters of treatment centers. To do that, I first extracted the geographical location column of each neighbourhoods in the raw data. I then labeled the locations of each treatment center in Toronto on a map including the neighbourhood names to see the distribution of the centers in the city.

(BURAYA MAP)

As it can be seen from the map, most of the treatment centers are clustered in downtown Toronto and gets significantly less frequent when moved out of the city center. Although it might be because of factors such as higher population, therefore higher demand, I also wanted to compare these findings with other neighbourhood characteristics.

To do that, I used Neighbourhood Profiles data (source) from the Toronto Open Data Portal. The data was last updated on October 7, 2019 and published by Social Development, Finance & Administration. The raw data includes detailed information on all 140 neighbourhoods of Toronto. Among over 100 variables, in this paper I focus mainly on population and components of population such as working-age percentage and senior percentage, immigration percentage, percentage of people who obtained bachelor’s degree education, average income of a household, and unemployment rate within the neighbourhood. The main reason for this focus within the dataset is to see if different components of a population will be correlated or explanatory of the cluster of treatment centers in the city. Additionally, because income and level of education might be other potential factors on willing to receive treatment on substance usage, I include these variables in my analysis as well. I extracted, cleaned, and analyzed the data again using R (citation), tidyverse (citation), dplyr (citation), janitor (citation), tidyr (citation), knitr (citation), and ggplot2 (citation).

After cleaning my data, I only included the neighbourhoods that has treatment centers and neighbourhood specific variables that I talked above. In order to make it visually more appealing, I changed rows to columns and columns to rows. So that the neighbourhood names will be rows and neighbourhood specific variables will be columns. Then finally, I composed my data table for the characteristics of each neighbourhood.

(DATA TABLE)

Considering the cluster of the treatment center in downtown Toronto, population can be one of the explanations for it since the total number of population around the cluster-area is higher than the neighbourhoods far from downtown. However, I cannot conclude that population is the reason for individual neighbourhoods to have more treatment centers. Although Niagara, Church-Yonge Corridor and Trinity-Bellwoods have the highest number of treatment centers, other neighbourhoods that have 1-2 treatments centers on average have similar number of population. Therefore, population might not be the main cause for individual neighbourhoods to have more treatment centers but might be one of the reasons for the cluster around the downtown area. Due to increased population and potentially increased demand, there might me more treatment centers clustered. Additionally, although working-age population is the biggest portion for most of these 29 neighbourhoods, it is important to highlight the top-3 neighbourhoods with the largest number of treatment centers, the working-age population is significantly higher (80.9%, 71.6%, 65.6% respectively) than the average of the whole city of Toronto which is 51.2% (citation).

Moving onto the immigration rates, although there is not clear relationship between neighbourhoods having more treatment canters and having less immigration rate, when the number of treatment center that each neighbourhoods has decreased, the immigration rate increases compared to the ones with larger number of treatment centers. Similarly, neighbourhoods with less treatment centers’ have lower bachelor’s degree percentage, whereas the ones with larger number of treatment centers have higher rates.

There are no clear patterns on neither unemployment rates nor average income in neighbourhoods with them having more/fewer treatment centers. However, downtown Toronto includes most of the business centers. Therefore, different than any potential relationship between individual neighbourhoods and income, having lots of business centers might be an explanation for the cluster of treatment centers in the downtown Toronto. It might be because business centers want to protect their business from substance users so that have exclusive deals with either provincial government or treatment centers. However, further investigation and research on that will answer to that question.

In addition, the question of willing to get the treatment might be another research topic for further investigation. The underlying reason for having a cluster within a city or individual neighbourhoods having more treatment centers than the average might be because people in those places are more willing to receive a substance treatment therefore have a high demand for them which will cause to have more treatment centers in specific parts of the city.

When looking both of the datasets, I extracted them from a reliable resource of Toronto Open Data Catalogue. Both of the datasets include very few missing values and have detailed components within, which can be used for other analysis. However, there are 2 major challenges with the datasets. First one is the sample size of the data is small which might lead to have a statistical bias in my analysis. For instance, we only have data for 29 neighbourhoods out of 140 neighbourhoods for the first dataset. Among 29 of these neighbourhoods there are 3 of them which has larger number of treatment centers. Such issue might cause sampling bias due to small sample size, therefore wrong assumptions and finding on the population, for instance treatment center distribution in Canada overall. Further investigation on substance treatment center frequencies as well as any potential correlation with neighbourhood characteristics might be needed from other provinces. Additionally, the first dataset is more recent-2020, whereas the second one is from-2016. Not only the difference might create a mismatch between the datasets and lead to a wrong conclusion, the true Neighbourhood characteristics might be different in 2020 since it is very likely those datapoints (percentages, rates, population, etc.) to change quickly. For further investigation, it will be better to use an updated dataset for neighbourhood characteristics to have a better comparison.

Abstract:

While substance usage has becoming a significant problem, this report analyzes the frequency and distribution of treatment centers in different neighbourhoods in the city of Toronto. Then it compares treatment center locations with specific neighbourhood characteristics using two datasets from Toronto Open Data Portal. Though the dataset used in this report might have a small sample size, therefore leading to have a statistical bias, further investigation on updated and different neighbourhood characteristics will be needed to explain any potential pattern on the substance treatment center distributions.