

**Capstone Project –Visualizing Areas of High Risk for Community Acquired SARS-COV-2 Transmission in
Hamilton, Ontario, Canada.**

Jennifer Sutton

DP0-701

IBM Applied Data Science Program

Section 1.0 – Introduction

1.1 Background

Controlling the spread of the **SARS-CoV-2 virus**, which causes **COVID-19 Disease**, is a top priority for Canada's Healthcare System. SARS-COV-2 is a highly contagious, droplet spread coronavirus from the same family as SARS-CoV-1 which caused the SARS (Severe Acute Respiratory Syndrome) Outbreak in Canada in 2003 (Government of Canada, 2020). To avoid resources being overwhelmed, citizens are being asked to physically distance themselves in order to reduce the spread of the virus. Various businesses and services have been locked down across the country to reduce points of close contact in the public. Ontario is currently one of the major areas of COVID-19 activity according to the provinces' digital disease surveillance system iPHIS (Integrated Public Health Information System) second only to Québec (Integrated Public Health Information System, 2020). As of April 17, 2020, 3.3% of report Ontario COVID-19 cases are in Hamilton. Hamilton Public Health (Hamilton Public Health, 2020a) has reported that **24% of Hamilton's COVID-19 cases have been community acquired** (individuals became infected after being in a public space).

1.2 Problem

Public municipal statistics do not give us the exact location of outbreak cases. For privacy reasons, only basic identifiers are reported (age, sex) in accordance with the Personal Health Information Protection Act (PHIPA) (Personal Health Information Protection Act, 2004). The only location-specific municipality numbers are for COVID-19 critical care patients and Long-Term Care residents. I have several family and friends who are considered part of the "high risk" population. They are curious about which areas are potentially high risk and should be avoided and are asking for my opinion.

Data science can be used to visually identify possible high-risk areas in the community and assist with making these types of decisions. Can we estimate high-risk areas for exposure in Hamilton? For this project I will be taking a closer look at Hamilton's population distribution across the city, location data for two main types of essential services and the location of Hamilton's COVID-19 clinics.

1.3. Interest

Potential key stakeholders for my project would be Hamilton Public Health, local healthcare agencies (Hamilton Health Sciences and St. Joseph's Healthcare), City of Hamilton council, and interested Healthcare Professionals and citizens.

Section 2.0 – Data Methodology

2.1 Data Sources

- **Research Data**
 - Research Data from government institutions and credible healthcare sources will be used and referenced in the Appendix. The main sources include: iPHIS via The

Government of Ontario, the Government of Ontario, The Government of Canada and Hamilton Public Health.

- **Statistical Data**

- Data for this project (Venue Data, Population Numbers per Ward, Hamilton Postal Codes) will be sourced from Statistics Canada and The City of Hamilton by either scraping it using BeautifulSoup 4.0 or by manually inputting it into Excel Files.

- **Data Analysis Software**

- Foursquare will be used to find Grocery and Drugstores in Hamilton including major chain stores and small businesses. Variety Stores will not be included in the analysis but box stores with notable Grocery and Drug Departments will (ex. Wal-Mart). Costco is listed as a 'Big Box Store' and 'Warehouse' in Foursquare so it will not be included at this time.
- Most of Hamilton's data is difficult or impossible to scrape. Microsoft Excel Spreadsheets will be downloaded and/or created based on official Government data.
- Python will be used via the Data Science platform Anaconda to code, visualize and manipulate all data. All Python coding will be done using the Jupyter Notebook web application. The PANDAS module will be used to create the data sets and structures needed for analysis. Folium will be used to visualize and mark the Foursquare venue location co-ordinates. A choropleth map will be made to visualize the population density of each ward. Matplotlib will be used to show bar graphs of data collected to further assess the findings shown in the Folium Maps. GitHub will be used as a data repository.

2.2 Data Cleaning

Hamilton Data was collected and combined into an Excel Document from various government sources. BeautifulSoup was not required. Quality statistical and research data collection took time to complete but all missing data was eventually compiled from the Ontario Government and The Canadian Government's websites. Ward data had to be manually extracted from each, individual Ward Profile on the City of Hamilton's Website. Hamilton Ward Boundaries were retrieved via the Open Hamilton Map API (https://opendata.arcgis.com/datasets/8b0b1f2bf8bb4e1da3a1bf567b17b77f_7.geojson). Calculations will all be completed in Python. Pandas Dataframes will be used for data manipulation.

3.0. - Exploratory Data Analysis

I first used Foursquare's Free Account to get a list of Grocery Store venues in Hamilton using the co-ordinates 43.26 °N, 79.87°W. I sent a get request and received a JSON file with **44** results. I processed the data and transformed it into a Dataframe. The results I obtained from Foursquare included the major chains (ex. Fortinos, Metro, FreshCo) and several specialty stores. The results are not comprehensive of all the Grocery Stores in Hamilton due to the limitations on Foursquare searches (radius limit, location specific and entries).

name	lat	lng
Nations Fresh Foods	43.258794	-79.873704
the mustard seed	43.266149	-79.882283
Fortinos	43.220846	-79.859429
Tan Thanh Supermarket	43.261117	-79.870609
Fortinos	43.255337	-79.928868
Keon's Super Convenience	43.256003	-79.867122
B&T Food Centre	43.262428	-79.877555

Figure 1- Hamilton Grocery Store Data

I then mapped the location of each of the results with Folium. I used a green marker. I decided to use “Stamen Toner” as my tiling to allow the markers to stand out more and to display major street names.

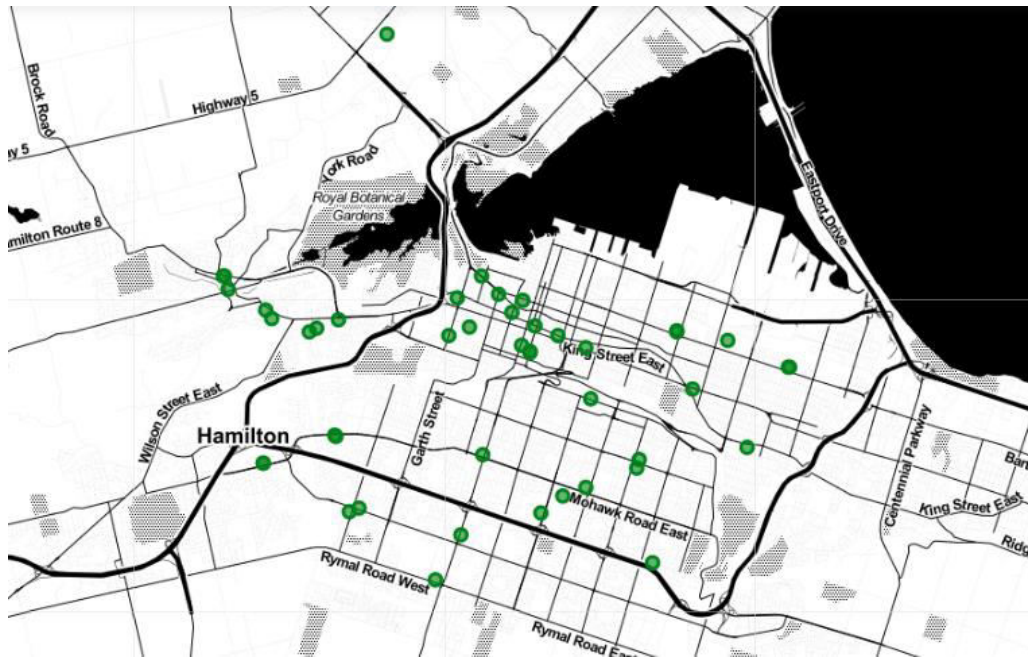


Figure 2- Map of Grocery Stores in Hamilton, ON

I then sent a new get request to obtain a list of Drugstores in Hamilton. I received a JSON file back with **63** results. The results for this venue category included Shoppers Drugmart, Rexall and several smaller chains (Guardian and I.D.A). I altered the radius several times and was unable to get a complete list of results with the limits placed on the free account. The results are not comprehensive of all the Drugstores in Hamilton due to the limitations on Foursquare searches.

name	lat	lng
Shoppers Drug Mart	43.257928	-79.857067
Shoppers Drug Mart	43.256980	-79.877739
Shoppers Drug Mart	43.262307	-79.887374
Shoppers Drug Mart	43.250290	-79.878558
Shoppers Drug Mart	43.247689	-79.839449
Shoppers Drug Mart	43.236946	-79.877695
I.D.A. - City Centre Pharmacy	43.258423	-79.869233
Rexall	43.258269	-79.873060

Figure 3 - Hamilton Drugstore Data

I mapped the location of each of the results with Folium with a blue marker.

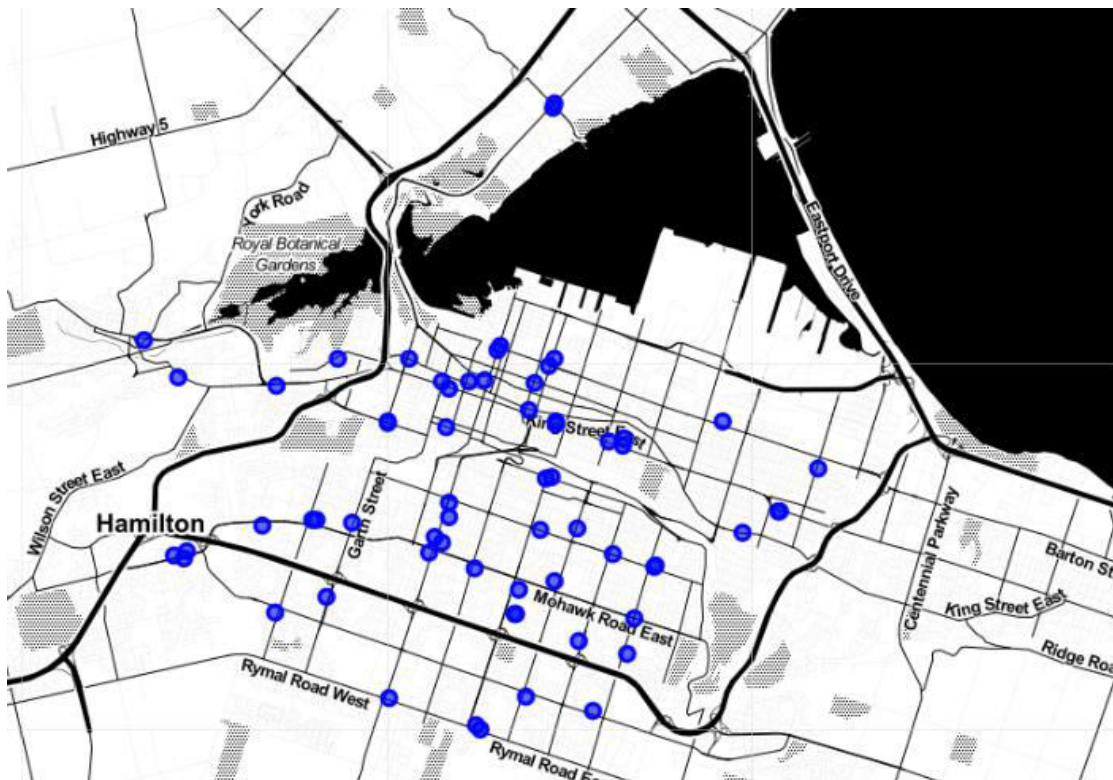


Figure 4- Map of Drugstores in Hamilton, ON

I then took both sets of data and combined them into one map. This allowed me to visually see where Grocery Stores and Drugstores were in relation to one another. Grocery stores and Drugstores are not always located near each other. The results show there are more Drugstores than Grocery Stores in Hamilton.

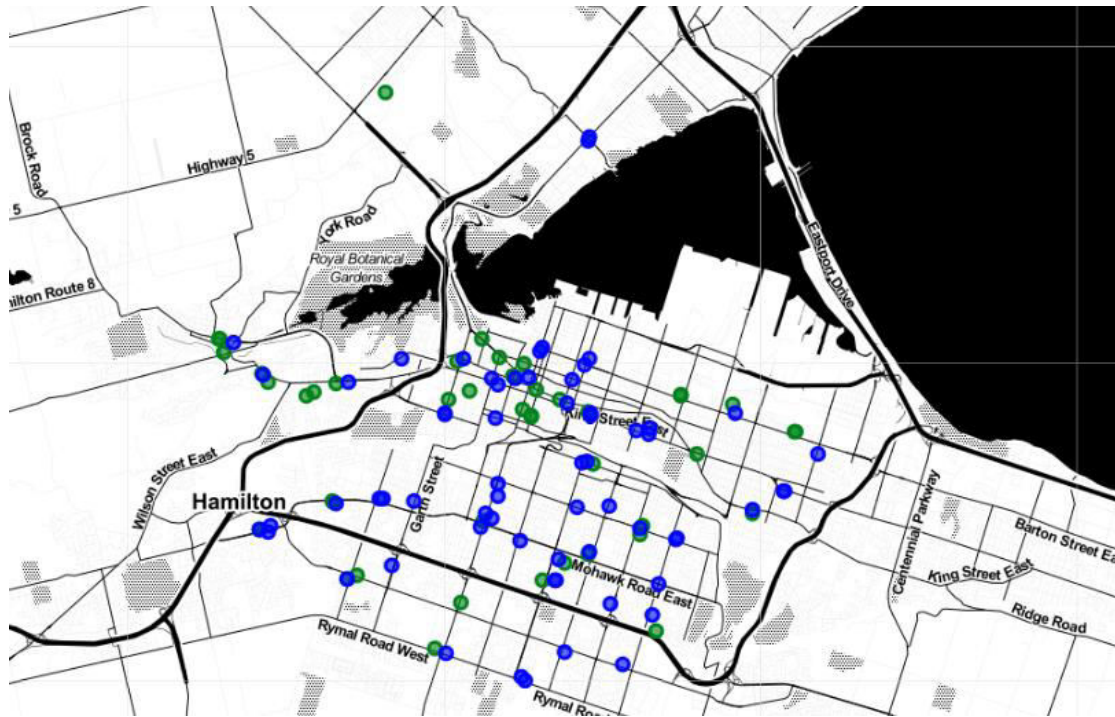


Figure 5- Map showing Grocery (Green) and Drug (Blue) stores in Hamilton, ON

With the Foursquare data mapping completed, I went to work on the Population Data. The City of Hamilton has **15 Wards** and is comprised of the original city and several of its surrounding municipalities. Population statistics were taken from the 2016 Census Excel File located on the Statistics Canada website. I took what I needed from the Statistics Canada Spreadsheet and combined it with another Excel File I created containing the information on each of the Hamilton Wards. I converted it into a Dataframe and processed and cleaned the columns. These census numbers are likely to be low as they are four years old and Hamilton has had a large, noticeable population increase since 2016.

	Ward	Name	Latitude	Longitude	Population
1	1	Chedoke-Cootes	43.261	-79.916	29850.0
2	2	Downtown Hamilton	43.402	-80.017	33600.0
3	3	Hamilton Centre	43.254	-79.841	41205.0
4	4	East Hamilton	43.224	-79.699	38595.0
5	5	Red Hill	43.238	-79.805	41855.0
6	6	East Mountain	43.196	-79.851	38650.0
7	7	Central Mountain	43.232	-79.845	47460.0
8	8	West-Central Mountain	43.231	-79.905	34485.0
9	9	Upper Stoney Creek	43.184	-79.721	28760.0
10	10	Lower Stoney Creek	43.216	-79.742	37215.0
11	11	Glanbrook	43.28	-79.853	25415.0
12	12	Ancaster	43.218	-79.984	42560.0
13	13	Dundas	43.362	-80.306	35365.0
14	14	West Mountain	43.206	-79.901	34230.0
15	15	Flamborough	43.267	-79.959	27675.0

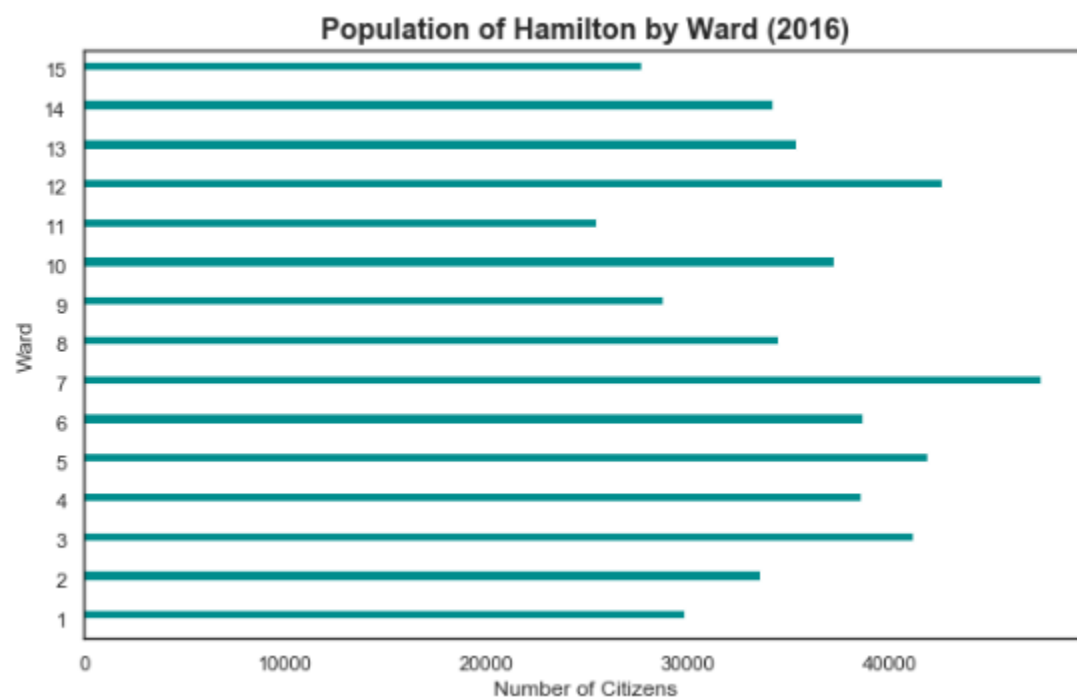
Figure 6- Hamilton Ward Data

I did a quick calculation of the Population Percentage in each of the wards using Python. The percentages are remarkably close to one another, so I decided against charting them. I kept the column for information purposes only.

Ward	Name	Latitude	Longitude	Population	Percentage of Population
1	Chedoke-Cootes	43.261	-79.916000	29850.000000	5.559487
2	Downtown Hamilton	43.402	-80.017000	33600.000000	6.257916
3	Hamilton Centre	43.254	-79.841000	41205.000000	7.674328
4	East Hamilton	43.224	-79.699000	38595.000000	7.188222
5	Red Hill	43.238	-79.805000	41855.000000	7.795389
6	East Mountain	43.196	-79.851000	38650.000000	7.198465
7	Central Mountain	43.232	-79.845000	47460.000000	8.839306
8	West-Central Mountain	43.231	-79.905000	34485.000000	6.422745
9	Upper Stoney Creek	43.184	-79.721000	28760.000000	5.356478
10	Lower Stoney Creek	43.216000	-79.742000	37215.000000	6.931200
11	Glanbrook	43.280000	-79.853000	25415.000000	4.733480
12	Ancaster	43.218000	-79.984000	42560.000000	7.926693
13	Dundas	43.362000	-80.306000	35365.000000	6.586642
14	West Mountain	43.206000	-79.901000	34230.000000	6.375251
15	Flamborough	43.267000	-79.959000	27675.000000	5.154399

Figure 7- Percentage of the Population of Hamilton in each ward

I then graphed the Population of each ward using Matplotlib and Seaborn to easily visualize how they compared to each other. Ward 7 Central Mountain shows the highest population of citizens in Hamilton.

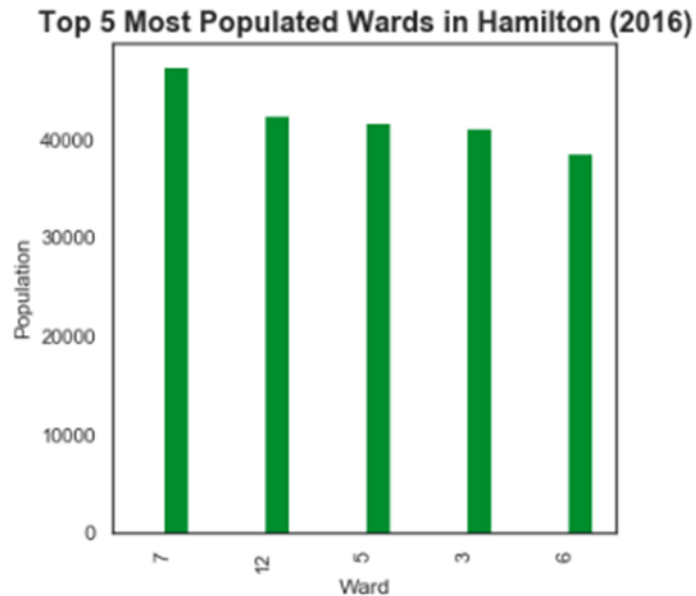


A new Dataframe was created from the data to include only the Top 5 most populated wards in

Ward	Name	Latitude	Longitude	Population	Percentage of Population
7	Central Mountain	43.232	-79.845000	47460.000000	8.839306
12	Ancaster	43.218000	-79.984000	42560.000000	7.926693
5	Red Hill	43.238	-79.805000	41855.000000	7.795389
3	Hamilton Centre	43.254	-79.841000	41205.000000	7.674328
6	East Mountain	43.196	-79.851000	38650.000000	7.198465

Figure 8- Top 5 Most Populated Wards in Hamilton

Hamilton in order of largest to smallest. The top 5 were then graphed using Matplotlib and Seaborn to



compare them. Python was used to calculate the Percentage of the Hamilton population living in the top 5 wards. **39.4%** of the total population of Hamilton live in one of these wards. The Top 5 populations were then mapped to compare where the largest number of citizens were living in comparison to where



the venues were located.

I then combined all the data into one map for analysis against the Ward Boundary lines. The Foursquare results do not extend to cover all the wards (see the Appendix for a better resolution of the results). I counted the total number of venues in the top 5 wards and added the values to an excel spreadsheet.

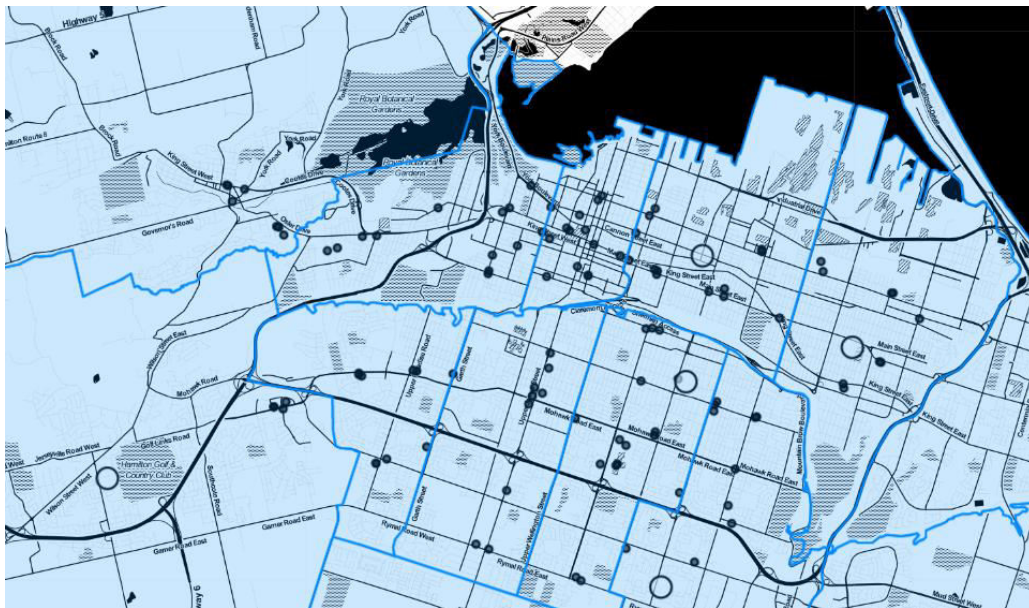
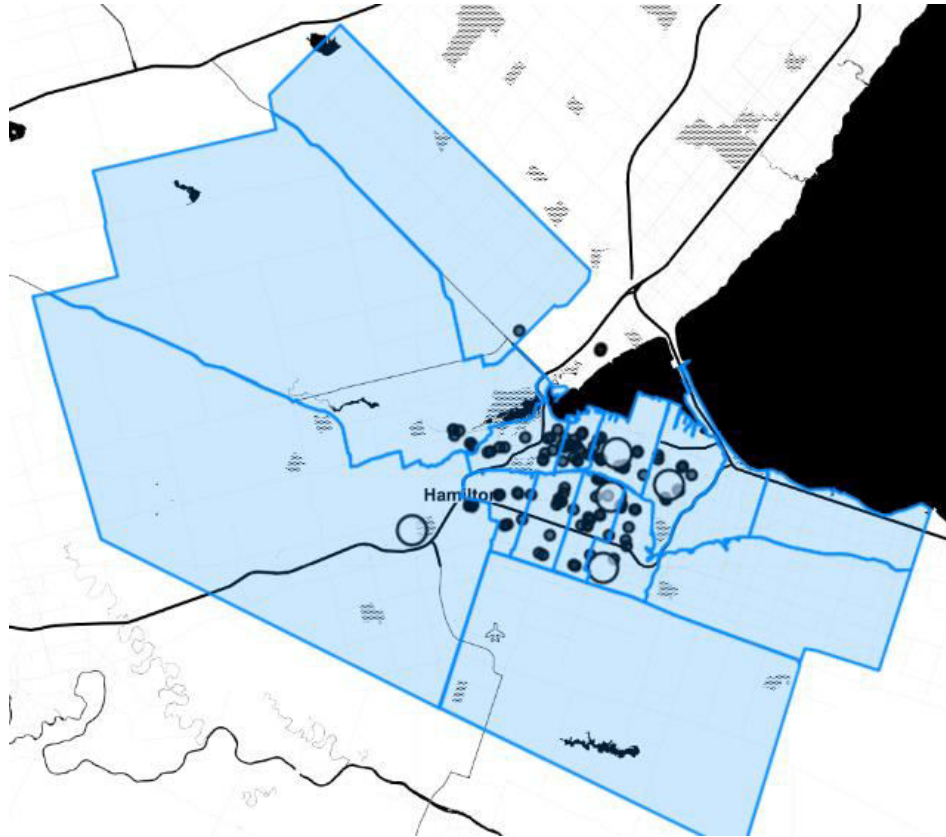


Figure 10a and 9b- Ward Boundaries (L to R- 15, 13, 12, 1, 2, 3, 4, 5, 10, 14, 8, 7, 6, 9, 11). Peak Population areas are indicated by the large, black outlined circles. Grocery Stores and Drugstores are indicated by black and grey circles.

I uploaded the Excel file into Python and created a new Dataframe with the venue count. I then sorted the data from most to least number of venues. Central Mountain has the most venues out of the top 5 wards and Ancaster has the least.

Ward	Name	Venues
7	Central Mountain	15
3	Hamilton Centre	12
6	East Mountain	9
5	Red Hill	7
12	Ancaster	3

Figure 11- Number of venues in each Top 5 Ward

Using the information and co-ordinates from Hamilton Public Health's website, I marked the COVID-19 clinics onto the map with the venues and population markers. Both clinics are located below the escarpment at Hamilton's two Urgent Care Clinics.



Figure 12 - Location of Hamilton's COVID-19 Assessment and Testing Clinics are shown as yellow circles.

4.0 - Results

Looking at the data, several key observations regarding Hamilton's wards were identified.

- Hamilton has a Population Density of 480.6 per square kilometer (Statistics Canada, 2016). From the generated maps we can see that the % of the population in each ward ranges from 4.7-8.8%.
- Most of the population resides in the middle of the Hamilton Map.
- The Top 5 Largest Wards are Central Mountain, Ancaster, Red Hill, East Mountain Hamilton Central. These 5 wards are home to 39.4% of the total population.

- 3 of these 5 wards are located on the escarpment.
- Foursquare Location Data was obtained for 44 Grocery stores and 63 Drugstores. According to the maps, most of these services are located in Ward 1 Chedoke-Cootes and Ward 2 Downtown Hamilton.
- Ward 3 Hamilton Centre has a large population and a moderate amount of venues. It shares a border with Ward 2 Downtown Hamilton.
- Ward 7 Central Mountain has the largest population in Hamilton and the most venues out of the top 5.
- Both COVID-19 clinics are below the escarpment.

5.0. – Discussion

Ward 7 has the highest likelihood of being high risk due to its population size and number of open venues and proximity to other largely populated wards.

The largest issue is how outdated the population data is. Hamilton has grown over the past few years with dozens of newly developed houses, condominiums and shopping centres on the escarpment. The areas below the escarpment had some development but not to the same extent. Canada is scheduled for a census in 2021 so there will not be any new data for awhile. Currently only inflation prediction data available for Hamilton. Any decisions need to consider the growth that has occurred in the city.

Hamilton has limited access to health resources in relation to its population distribution. Three of the top 5 populated wards would have to drive a minimum of 20-30 minutes to be the nearest COVID-19 clinic. Hamilton has set up COVID-19 Triage Telephone Hotline to reduce the need for citizens to have to go to the centre unless necessary with does help (HPH, 2020A). I worry that citizens/clients with transportation access limitations might be dissuaded from being tested if the centre takes too long to get to or is impossible to get to. Care access is further compromised by the fact that only one hospital is located on the escarpment – Juravinski Hospital (West 5th is primarily for Psychiatric Services). Hamilton General Hospital, McMaster Hospital and St. Joseph's Hospital are all located below the escarpment. Juravinski does have a General Intensive Care Unit (ICU), but the specialty ICUs are located at the Hamilton General and St. Joseph's. As of April 25, 2020, HPH has opened a drive-through Testing Centre but its location is unknown to the public.

There are a lot of open venues. More than 107 venues (including those not obtained in this project) are operating in the city and all of them have the potential to be areas of contact.

I would have liked to analyze more data. I would like to take an even closer look at the demographic data of each of the wards and would like to K-Means Cluster the wards to determine the probability of exposure. If I could find a more complete Hamilton Geojson file, I would like to create a choropleth map. The data is limited by the results provided by Foursquare for each location point. The data does not include Costco Wholesale which is a popular destination providing bulk groceries and pharmacy services. I would like to add in the Costco locations and other missing locations and visualize the data again.

5.0 - Conclusion

COVID-19 continues to affect the city of Hamilton. Identified cases continue to rise everyday in Ontario and the provincial lockdown remains in effect. Community spread is more likely to occur where social distancing is compromised such as populated Grocery Stores and/or Drugstores and populated neighbourhoods. By better understanding where the risk is highest stakeholders can attempt to mitigate it.

References

- City of Hamilton. (2020). *Ward Boundaries*. [Data Set].
http://open.hamilton.ca/datasets/8b0b1f2bf8bb4e1da3a1bf567b17b77f_7
- City of Hamilton. (2016). *Ward Profiles*. <https://www.hamilton.ca/city-initiatives/strategies-actions/ward-profiles>
- Cybo Global Business Directory, (2020). *Postal Code Directory*. [Data Set]. <https://postal-codes.cybo.com/>
- Government of Canada. (2020). *Coronavirus disease (COVID-19): Summary of assumptions*.
<https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/assumptions.html>
- Government of Ontario. (2020). *The 2019 Novel Coronavirus (COVID-19)*.
<https://www.ontario.ca/page/2019-novel-coronavirus#section-0>
- Hamilton Public Health. (2020a). *Status of Cases*. <https://www.hamilton.ca/coronavirus/status-cases>
- Hamilton Public Health. (2020b). *COVID-19 Information from Public Health*.
<https://www.hamilton.ca/coronavirus>
- Integrated Public Health Information System. (2020). *iPHIS Resources*.
<https://www.publichealthontario.ca/en/diseases-and-conditions/infectious-diseases/ccm/iphis>
- Personal Health Information Protection Act, Revised Statutes of Ontario (2004, S.O. 2004, c. 3, Sched. A).
<https://www.ontario.ca/laws/statute/04p03>
- Statistics Canada. (2016). *Hamilton Census Profile*. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CD&Code1=3525&Geo2=PR&Code2=35&Data=Count&SearchText=hamilton&SearchType=Begins&SearchPR=01&B1=All&TABID=1>
- Statistics Canada. (2016). *Toronto Census Profile*. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?B1=All&Code1=3520&Code2=35&Data=Count&Geo1=CD&Geo2=PR&Lang=E&SearchPR=01&SearchText=Toronto&SearchType=Begins&TABID=1>

Appendix

