Section 1.0 – Introduction and Business Problem

Controlling the spread of the SARS-CoV-2 virus, which causes COVID-19 Disease, is vital to the health and safety of Canada's citizens. To avoid resources being overwhelmed, citizens are being asked to physically distance to reduce the spread of the infection and various businesses and locations have been locked down across the country. Ontario is currently one of the major areas of COVID-19 activity according to the provinces' digital surveillance system iPHIS (Integrated Public Health Information System) (Government of Ontario, 2020). As of April 17, 2020, 3.3% of COVID-2019 Ontario cases have been reported in Hamilton. According to Hamilton Public Health (HPH, 2020), 24% of these cases have been community acquired. Public statistics do not give us the exact location of outbreak cases, only basic identifiers for health data privacy as per the Personal Health Information Protection Act (PHIPA) (HPH, 2020). The only location-specific numbers are for patients in each of the major hospitals. Is there a way we can use data science to identify high exposure risk areas?

As a Registered Nurse in Informatics studying Data Science, I want to take a closer look at two of the essential services being heavily utilized during the pandemic. Their current popularity places them as one of the most likely points of transmission outside the home. The 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 staff and interested Healthcare Professionals. I am looking to determine possible strategic locations for intervention implementation (such as Health and Safety spot checks and client education initiatives) based on exposure risk.

Section 2.0 – Data

- **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 and Hamilton Public Health, the Government of Ontario 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 scraping it using Beautiful Soup 4.0 and by downloading publicly available 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 (ex. Big Bee Mart) but box stores with notable Grocery and Drug Departments will (ex. Wal-Mart).
        - Beautiful Soup 4.0. is the web scraper that will be used to scrape data from available websites for use in Python. In areas where scraping is difficult or impossible, Microsoft Excel Spreadsheets will be downloaded and/or created for use to be imported into Python.
        - 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.