 Big Data Analytics Project - F2021

**CIND 820 Section D10**

**Steven Connolly**

**Student Id #501073413**

**Supervisor: Ceni Babaoglu, PhD**

**Submitted: October 18, 2021**

Table of Contents

[Abstract 3](#_Toc452457337)

[Literature Reviews 4](#_Toc452457337)-10

Research Question [4](#_Toc452457339)

Approach [5](#_Toc452457340)

[Searching Literature 5](#_Toc452457341)

Analyzing Material [6](#_Toc452457341)-8

Managing Results [10](#_Toc452457341)

Description of Dataset [1](#_Toc452457337)1

Graph of Tentative Methodology [12](#_Toc452457337)

GitHub Repository [12](#_Toc452457337)

References [13](#_Toc452457337)

# Abstract

The emergence of COVID 19 into our daily lives has had a dramatic impact on us all. In fact, no other factor has contributed to a worldly discussion more than this global pandemic has since World War 2. It had been approximately 100 years since the last global viral outbreak and needles to say the world was unprepared. In Canada, no place was harder hit than our Long Term Care Homes (LTCH). At the beginning of the pandemic is seemed as though the only people affected were our most vulnerable who happened to reside in LTCH. COVID 19 had a devasting affect on our LTCH. The mortality rates were/are significantly higher than the rest of the country. This devastating virus has helped to shed a light on the underfunding/lack of attention our most vulnerable peoples are provided.

The problem that I am trying to solve is if deaths per confirmed cases were mitigated in any way after the very start of the pandemic and there afterward decreasing over time.

The dataset I will be using will be the Long Term Care Home Covid Summary (<https://data.ontario.ca/en/dataset/long-term-care-home-covid-19-data>) provided by the Government of Ontario. This dataset shows the cases of COVID 19 over time along with deaths and recoveries. I will be using classification in order to identify why there were so many deaths, and if the cause was specific to only a few care homes or all care homes. COVID 19 has shed light on the poor standards at LTCH and it would be disappointing to discover that deaths/cases haven’t improved over the course of the pandemic.

I will be using Python as my programming language of choice. It is the one I’m most familiar with and allows me to use a module that I find to be the most straightforward (Sci-kit Learn (module)). Along with Python I will be using several classifying algorithms (naïve bayes, decision trees or random forest) in order to find the correlation between cases and deaths. Is this a system wide problem? Some data cleaning will be necessary on the dataset. My hope is that the deaths/cases will have dropped from the beginning of the pandemic (due to better treatments and vaccinations). The COVID 19 pandemic has forever changed the world. I am hopeful that the mortality rate at the long term care homes in Ontario has changed as well; for the better.

# Literature Reviews

## Step 1: Research Question

The purpose of this literature review and the problem I am trying to solve is the cause of the high deaths in Long Term Care Homes in Ontario. Why did LTCH in Ontario get hit so badly by COVID 19. Is this a global problem? Was anything done to mitigate the damage done by this terrible virus on societies most vulnerable people? This is what I would like to discover.

I decided to do a broad search on deaths in Long Term Care homes. There appeared to be some research and analysis done on the impact of Covid 19 in LTCH in Ontario and elsewhere. Problems addressed looked at deaths, vaccination rates, outbreaks by region among others. The topic that interested me most is deaths in LTCH.

My search was narrowed to Long Term Care Homes and specifically deaths related to COVID 19. I tried to look for data that occurred past the very beginning of the pandemic – when the data was less available. I also search for broader implications of Covid 19 death rates in LTCH. I considered the socio-economic factors that would result in someone being placed in a care home versus a private facility or having in-home care. Were the LTCH ignored because society deems their lives as less important? Or was a class system also in play here.

## Step 2: Approach

After you have determined your research question, it is important to discern what method you will use to research, and which specific terms you will use to conduct your search. Determine what kind of literature you want to look at, whether it be journal articles, books, electronic resources, newspapers, or even other literature reviews on similar topics.

You will also need to determine what databases you want to use based on your topic. Ryerson University has over 300 databases available to students, visit the [A-Z List.](http://learn.library.ryerson.ca/az.php)

However, it would be beneficial to narrow down your search in your topic according to the relevant discipline. For example, if you were writing a literature review for a class in social work, you would benefit from using the Social Work databases.

## Step 3: Searching Literature

In searching for literature on the subject of mortality rates in Long Term Care Homes I made sure to concentrate on specific key words and string during my search.

Key areas of focus were:

* Covid 19 in Ontario
* Long Term Care Homes in Ontario
* Covid 19 Death Rates – Long Term Care Homes
* Covid 19 Death Rates – Global – Long Term Care Homes
* Covid 19 in GTA

There seemed to be plentiful information found with those key string searches. There was no need to go any more in depth or search for any individual or author of interest.

Remember to keep track of what keyword searches you have conducted so you can save time and determine how to improve your search and continue with the searching process.

### Refined Search

By the end of my searching I concentrated mostly on Covid 19 and effects on Long Term Care homes/ by demographic.

## Step 4: Analyzing Material

During my search for literature matching the topic I am analyzing I came across a lot of material. I made sure to only search for materials that I deemed to be accurate, credible and authentic.

Purpose of the work:

The purpose is to publish relevant findings. It is to analyze data in order to better help people living in Long Term Care Homes in Ontario. Perhaps this data analysis can be used to help future generations from overcoming a similar situation. This analysis can also be referred to by other data scientists to help aid in their own discoveries.

### Reference 1

The authors are Mancini, Pedersen and Common. They are credible journalists who work for the CBC. The authors cite their sources – who are themselves credible. The journalists consider themselves to be unbiased when reporting on the news.

The work is based on a scholarly work but is done by journalists so it can be considered peer-reviewed in the slightest of ways.

The information is accurate and is supported by evidence that can be fact-checked. Most data discovered in this literature review is considered credible. Similar information can be found to back up what is being said in this article. This material has been made in the last year and a half, so time has not diminished the value of this information. The data mentioned in this example falls in the same time range as the dataset I will be using. The information was last updated in December of 2020.

### Reference 2

The author is Howlett. They are too a credible journalist who work for the Globe and Mail. The author cites their sources as do most journalists at a high level. The journalist, like at any news organization consider themselves to be unbiased when reporting on the news.

The work is done by a credible journalist and published by what is considered to be one of the most credible newspapers in Canada.

The information is accurate and is supported by evidence that can be fact-checked. The information gathered by the journalist can be easily referenced and be found credible. This material has been made in the 2 years, so time has diminished slightly the value of this information but not to a noticeable effect as Covid 19 is still very relevant. The data mentioned in this example falls at the beginning of the range in time of the dataset I will be using. The information was last updated in August 11, 2020.

### Reference 3

The author is Chudasama. They are using the International Journal of Epidemiology to publish the results of their findings. The data is quite profound and detailed. This data is similar to the dataset I am using with the exception of it dealing with cases in the UK. The is a scholarly and peer reviewed piece of work.

This information can be considered very accurate and is sure to cite and bevy of sources for reference. The information was last updated in September 2021.

### Reference 4

The author is Ibrahim. They are a scholar who works for Health Law and Ageing Research Unit, Department of Forensic Medicine, Monash University, Victorian Institute Forensic Medicine, Southbank, Victoria, Australia. This is a worldly article written for Nature.

It is highly detailed and peer-reviewed so can be considered credible. The analysis made by this scholar used over 10 references that too can be deemed credible. The material being used here doesn’t pertain to Ontario, but the information still falls in the same family of analysis. The information is accurate and is supported by evidence that can be fact-checked. Most data discovered in this literature review is considered credible. Similar information can be found to back up what is being said in this article. This material has been made in the last 6 months, so can be considered quite timely. The data mentioned in this example falls in the same time range as the dataset I will be using. The information was last updated in July of 2021.

### Reference 5

The authors Public Health Ontario. They are public health officials tasked with relaying information to the general public of Ontario. Responsible government can be considered to be credible in the province of Ontario, so the information can be deemed credible.

The information being used in this resource is cited by using information gathered by Ontario Agency for Health Protection and Promotion (Public Health Ontario). This material has been made in the last year, so it is quite relevant. The data mentioned in this example falls in the same time range as the dataset I will be using. The information was last updated in March of 2021.

### Reference 6

The authors are Canadian Institute For Health Information (CIHI). They are a non profit health organization designed to give information to the general public in Canada regarding health information.

This work can be considered peer reviewed based on the resources they are citing. The information and analysis gathered matches that of other datasets and analysis found regarding the topic of Covid 19 in Long Term Care Homes. This data refers to all of Canada so is a bit beyond the scope of my analysis but remains very pertinent. This material has been made in the last year and a half, so time has not diminished the value of this information. The data mentioned in this example falls in the same time range as the dataset I will be using. The information was last updated in December of 2020.

**Step 5: Managing Results**

After analyzing the material and determining what information and sources you want to use, it's important to keep track of what you have looked through. Keep a list of the following:

* Several Searches were made for Long Term Care Homes.
  + This was too broad and only became successful when searching for COVID 19 as well
* Databases used: Ontario Data Catalogue, The Canada Government Dataset, Google
* The sources I want to use for literature review are:
  + “ These nursing home chains have the highest COVID-19 death rates in Ontario, data analysis. *APA Style Blog”*
  + “More than 1,000 long-term care residents died of COVID-19 in older, multibed homes, analysis shows.”
  + “Penetration and impact of COVID-19 in long term care facilities in England: population surveillance study.”
  + “An equation to predict deaths of nursing home residents during a pandemic”
  + “COVID-19 in Long-Term Care Homes in Ontario: January 15, 2020 to February 28, 2021”
  + “The Impact of COVID-19 on Long-Term Care in Canada.”
* I would like to research topic including Covid 19 death rates by region and demographic globally. Africa has the lowest recorded immunization rates and is therefore of particular interest to me.

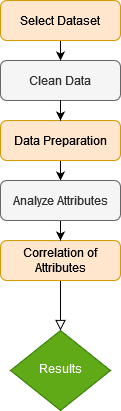
# Description of Dataset

The data used for this project can be found at <https://data.ontario.ca/en/dataset/long-term-care-home-covid-19-data>).

This dataset has cases of Covid 19 in Long Term Care homes in Ontario. It shows total cases, recovered cases and deaths. The analysis of this dataset is to determine the causality or pattern in the high level of Covid 19 deaths. This dataset will look at the attributes and determine if there is any high correlation between any 2 attributes.

The attributes used include: Report\_Data\_Extracted (Date), LTC\_Homes\_with\_Active\_Outbreak (Int), LTC\_Homes\_with\_Resolved\_Outbreak (Int), Confirmed\_Active\_LTC\_Resident\_Cases (Int), Total\_LTC\_Resident\_Deaths (Int), Total\_LTC\_Resident\_Deaths (Int), Total\_LTC\_HCW\_Deaths (Int), Active\_Outbreaks\_with\_No\_Resident\_Cases (Int).

# Graph of Tentative Methodology



# GitHub Repository

## Link

https://github.com/StevenConnolly/LTCHOntarioCovid

# References

1. Mancini, M., Pedersen, K., Common, D. (2020, December 20). These nursing home chains have the highest COVID-19 death rates in Ontario, data analysis. *APA Style Blog*. <https://www.cbc.ca/news/canada/nursing-homes-covid-19-death-rates-ontario-1.5846080>
2. Howlett, K. (2020, August 11). More than 1,000 long-term care residents died of COVID-19 in older, multibed homes, analysis shows. <https://www.theglobeandmail.com/canada/article-nursing-homes-with-multibed-wards-left-residents-vulnerable-to/>
3. Chudasama, D.Y. (2021, September 01). Penetration and impact of COVID-19 in long term care facilities in England: population surveillance study. <https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyab176/6361027>
4. Ibrahim, J.E. (2021, July 15). An equation to predict deaths of nursing home residents during a pandemic. <https://www.nature.com/articles/s43587-021-00083-x>
5. Public Health Ontario. (2021, March 03). COVID-19 in Long-Term Care Homes in Ontario: January 15, 2020 to February 28, 2021. <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjl7rOz49TzAhUsn-AKHaynBgUQFnoECAIQAQ&url=https%3A%2F%2Fwww.publichealthontario.ca%2F-%2Fmedia%2Fdocuments%2Fncov%2Fepi%2F2020%2F06%2Fcovid-19-epi-ltch-residents.pdf%3Fla%3Den&usg=AOvVaw3wUPyPxciK4FEDjw0gE3X_>
6. Canadian Institute For Health Information (CIHI). (2020, December 20). The Impact of COVID-19 on Long-Term Care in Canada. <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjl7rOz49TzAhUsn-AKHaynBgUQFnoECAgQAQ&url=https%3A%2F%2Fwww.cihi.ca%2Fsites%2Fdefault%2Ffiles%2Fdocument%2Fimpact-covid-19-long-term-care-canada-first-6-months-report-en.pdf&usg=AOvVaw3dEI_Q_hjk60tDepaYcU-t>