# Big Data Analytics Project - F2021

**CIND 820 Section D10**

**Steven Connolly**

**Student Id #501073413**

**Supervisor: Tamer Abdou, PhD**

**Submitted: September 27, 2021**

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.