CIND 820 Big Data Project – Summer 2021

William Murphy

Project abstract:

Problem & Theme

The theme of this project is to study and identify factors that led to the mortality of residents of Ontario who contracted COIVD 19. This will be achieved by analysis of compiled daily reported data from public health units on confirmed positive cases of COVID-19 in Ontario.

This project will identify how the following factors contribute to mortality of ontario residents who tested positive for covid 19:

* Accurate Episode Date – if available an appoxamation of when symptoms began
* Test Reported Date – The date indicated on the laboratory slip of the Covid 19 test
* Age Group – Age by decade
* Gender
* Case Acquisition Info
* Outcome – Resolved or Fatal
* Outbreak Related
* The Public health unit that reported the outbreak

The data set I will be using is the Confirmed positive cases of COVID19 in Ontario :

<https://data.ontario.ca/dataset/f4112442-bdc8-45d2-be3c-12efae72fb27/resource/455fd63b-603d-4608-8216-7d8647f43350/download/conposcovidloc.csv>

Compiled daily reported data from public health units on confirmed positive cases of COVID-19 in Ontario.

I will be using classification analysis to find the degree to which location, age, gender, outbreaks and test date affect the mortality rate.

I’ll be using python to do this analysis.