Coursera - Applied Data Science Capstone – Final Assignment Week 1 – Introduction and Business Problem

Chris Blackwood

Introduction

COVID-19 has been a major health and economic issue worldwide since the spring of 2020. The virus was first identified in humans in Wuhan, China in December of 2019, and the World Health Organization recognized a global pandemic in March of 2020 (Wikipedia, 2021). As of late January 2021, nearly 100 million cases have been identified worldwide, and over 2 million people have died (World Health Organization, 2021).

Many organizations and countries initiated vaccine development promptly. The Pfizer-BioNTech vaccine was the first vaccine approved in Canada, and distribution began in late December of 2020 in that country (CBC, 2021). Vaccine production is complex, worldwide demand is extreme, and distribution logistics and vaccine storage is non-trivial. As a result, it will take many months before everyone receives a vaccine. In Canada, the priority recipients of the vaccine are generally the elderly and healthcare workers, and individual provinces determine the exact distribution order (Government of Canada, 2021). In the province of Ontario for example, health care workers in hospitals, long-term care homes, and retirement homes were amongst the first to receive vaccination (Pelly, 2021).

Business Problem

Data scientists can play a valuable role in understanding the distribution and prevalence of COVID-19 infection at a detailed level. Data can be used to identify the characteristics of people who are most prone or least prone to COVID-19 infection. If populations who are most at risk for COVID-19 infection can be identified, then a vaccine distribution strategy to prioritize these populations can be developed. The target audience for my study would be government officials responsible for coordinating COVID-19 vaccination distribution.

I will examine COVID-19 infections in the city of Toronto, the capital of the Canadian province of Ontario. Firstly, I am interested in the rates of COVID-19 infection at a given time, and also the cumulative cases since the start of the pandemic. COVID-19 data is available from the City of Toronto's website. I will sort these data by geographic area within the city to see if certain areas have more confirmed infections than others, or if the distribution is random. I will generate choropleth maps to categorize the distribution of COVID-19 cases. I will also query Foursquare for neighbourhoods of interest to see the types and abundances of venues in these neighbourhoods, and determine if any trends are apparent.

Secondly, I will attempt to understand the different regions of Toronto in terms of demographic characteristics. The City of Toronto website hosts a variety of census

data, including population statistics, rates of home ownership, household income, and education level. If there are regions of the city that have particularly high or low COVID-19 cases, I will attempt to identify any demographic factors of significance in these areas. I will use an assortment of statistical and mapping analysis techniques to determine relevance of demographic factors.

My hope is that trends will become apparent, such as particular geographic regions, common venues, or demographic factors correlate with high COVID-19 case rates. If I am successful in identifying factors that correlate with high COVID-19 infection rates, then these correlations could be used to help prioritize populations for COVID-19 vaccine distributions.

References

CBC, retrieved January 21, 2021. *Here's the COVID-10 vaccine rollout plan, province by province. https://www.cbc.ca/news/canada/covid19-vaccine-rollout-plans-canada-1.5836262*

Government of Canada, retrieved January 21, 2021. *Vaccines and treatments for COVID-19: Vaccine rollout. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/prevention-risks/covid-19-vaccine-treatment/vaccine-rollout.html*

Pelly, Lauren, retrieved January 21, 2021. *Health-care workers lining up for COVID-19 vaccine, but some warn of 'real troubles' with hesitancy. https://www.cbc.ca/news/canada/toronto/health-care-vaccines-covid-19-hesitancy-1.5852119*

World Health Organization, retrieved January 21, 2021. *WHO Coronavirus Disease (COVID-19) Dashboard*. https://covid19.who.int/

Wikipedia, retrieved January 21, 2021. *COVID-19 Pandemic.* https://en.wikipedia.org/wiki/COVID-19_pandemic