

FINAL EXAM PROJECT

MIE 1624 Introduction to Data Science and Analytics

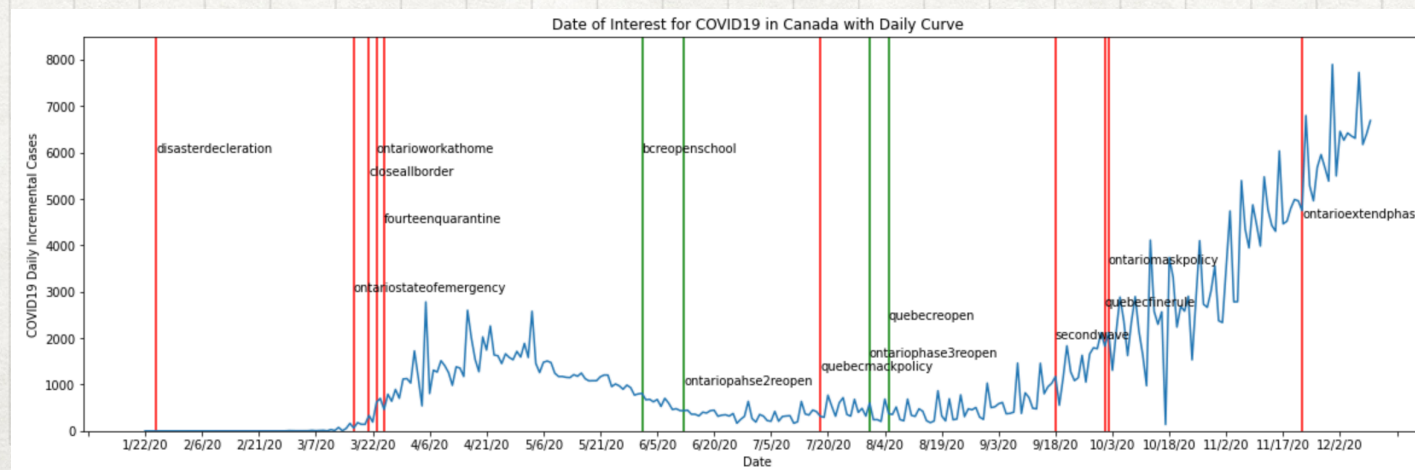
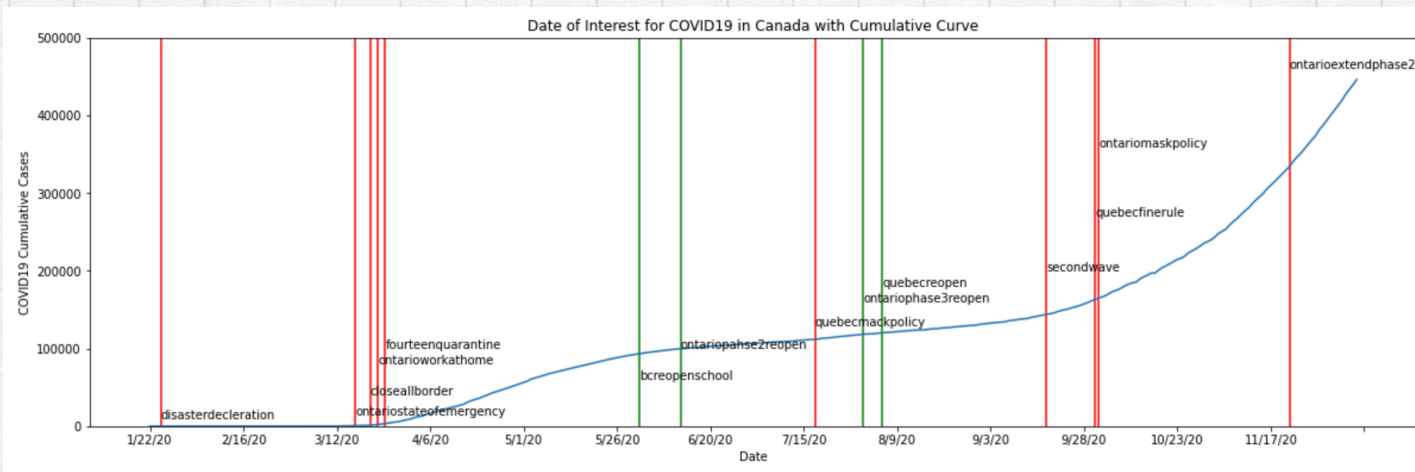
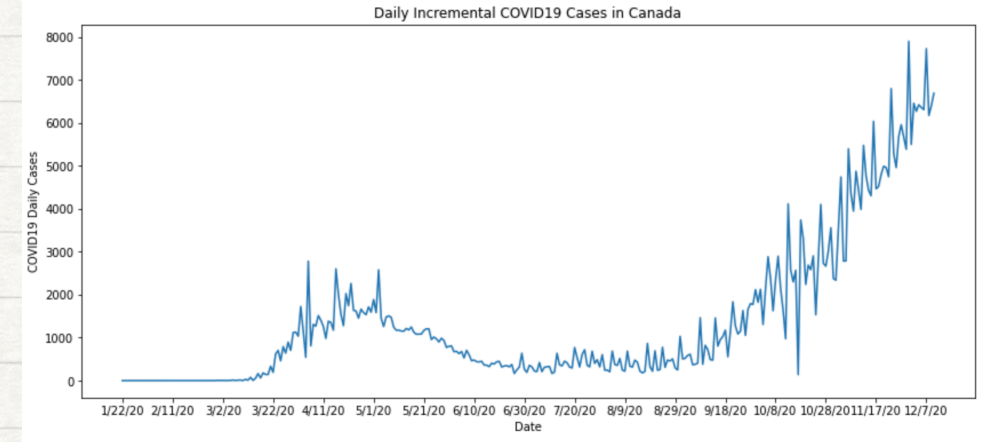
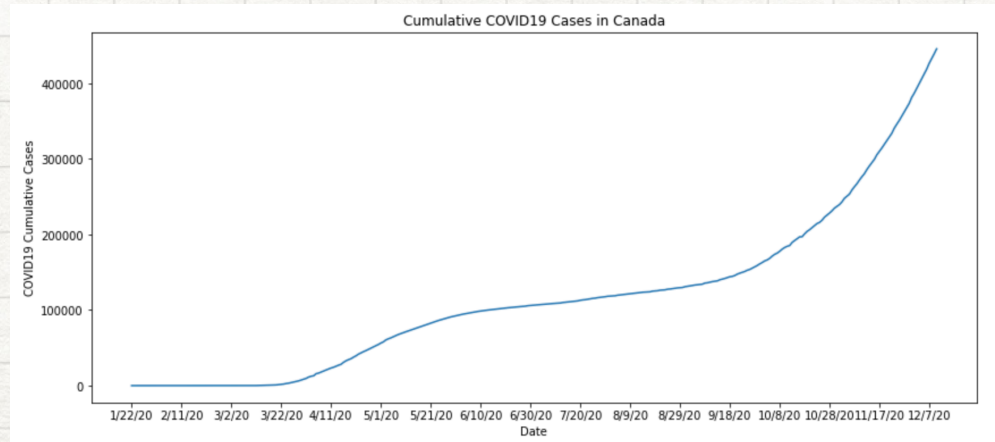
Name: Jiani Jia

Student ID: 1002226245

Data Cleaning

1. Extract the cumulative confirm, death, recovery COVID 19 cases in Canada and Quebec
2. Extract the daily confirm COVID 19 cases in Canada and Quebec
3. Extract the total population in Canada and Quebec

Exploratory Data Analysis

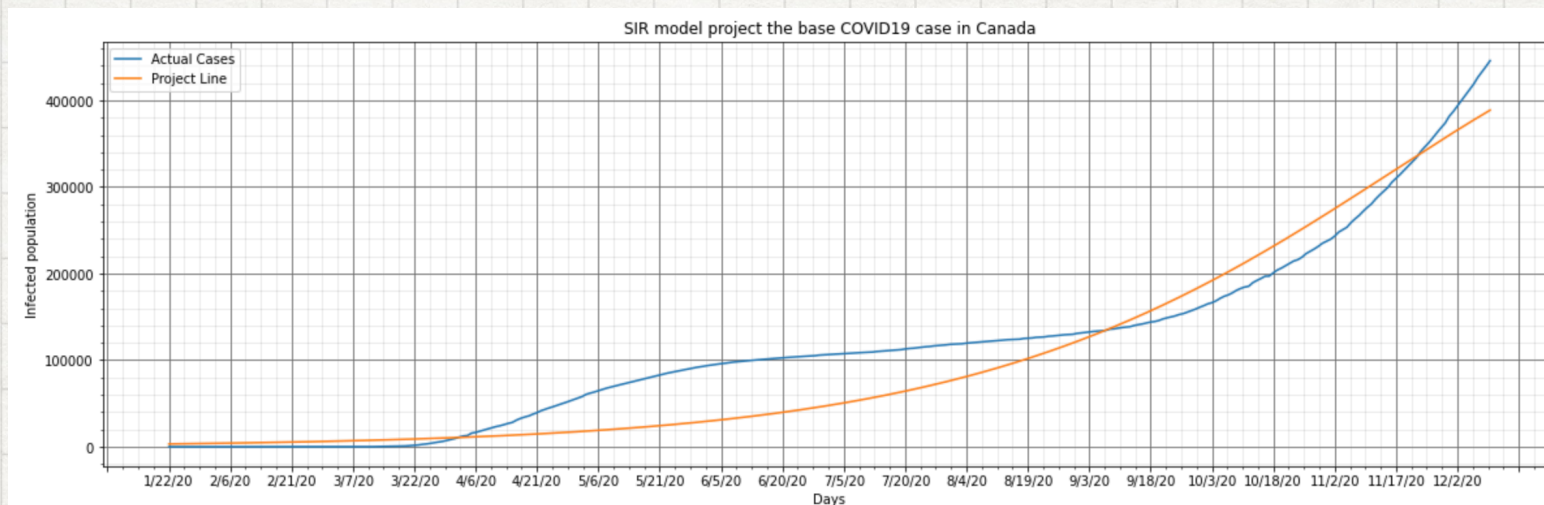


- It is clear to see that between 2020-3-12 to 2020-7-31, there is a peak and then goes down. And after that, we are facing the second peak, the data shows until now the Canada is not reach to its peak value, the daily confirm cases are still shown an increase trend.
- At the beginning of COVID19, lots of policies promulgated include close the border, school and university shutdown, work at home, etc. After that, the incremental confirm cases reach the peak within a month and continuously decrease. When the daily incremental number keep in a low level, the reopen and back to school policies promulgated. Then the second wave come, although more protective rules are promulgated, the second wave still growth fast and much higher than the first peak.
- From the experience of facing the first peak, the corresponding policies should be announced at the beginning of the increase trend occur to fast the speed to reach the peak value.

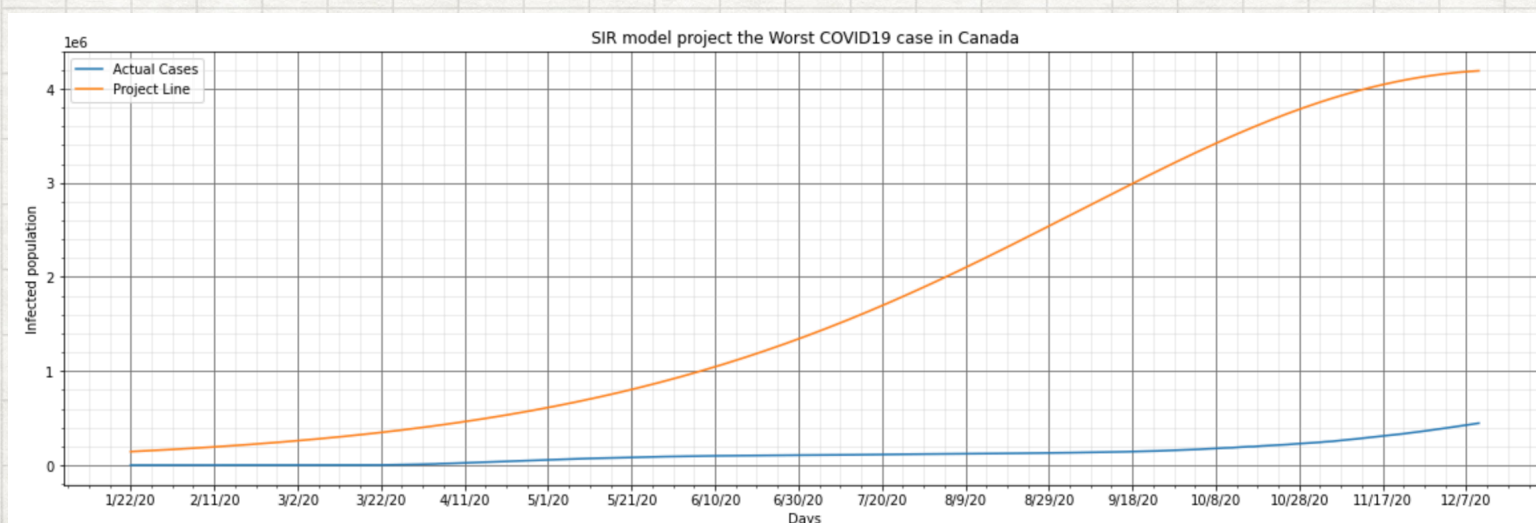
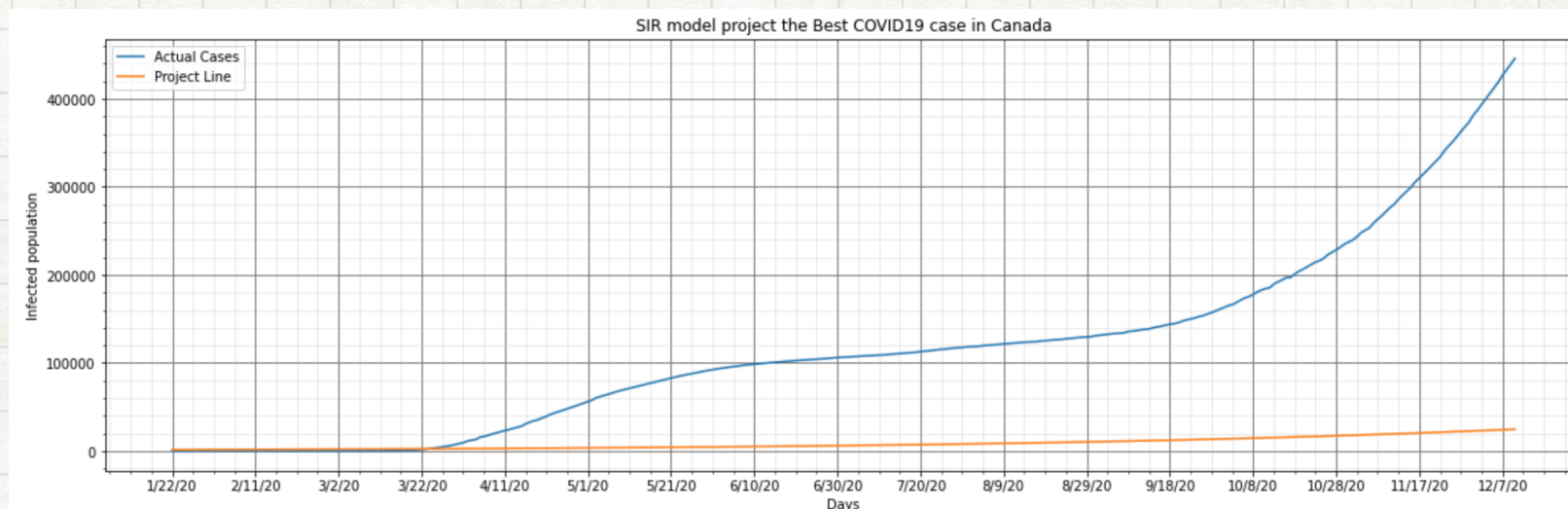
Model Selection and Fitting to Data

- Model Selection: SIR
- The SIR model aims to predict the number of individuals who are susceptible to infection, are actively infected, or have recovered from infection at any given time.
- Important parameters in SIR:
 - β is the transmission rate
 - γ is called the removal or recovery rate
 - $S(0)$ is the initial proportion of the population that are susceptible
 - $I(0)$ is the initial proportion of the population that are infectious
 - $R(0) = \beta / \gamma$, it is the effective transmission rate
- Assumption in SIR:
 - There are no inbound or outbound infected travellers during this period
 - Recovery people will not infect again
 - R_0 is constant

Parameters	Best Projection	Worst Projection	Basic Projection
β	0.109	0.035	0.1096
γ	0.1	0.02	0.0926
$R(0)$	1	3.0	2.4
$I(0)$	1469	144053	3251



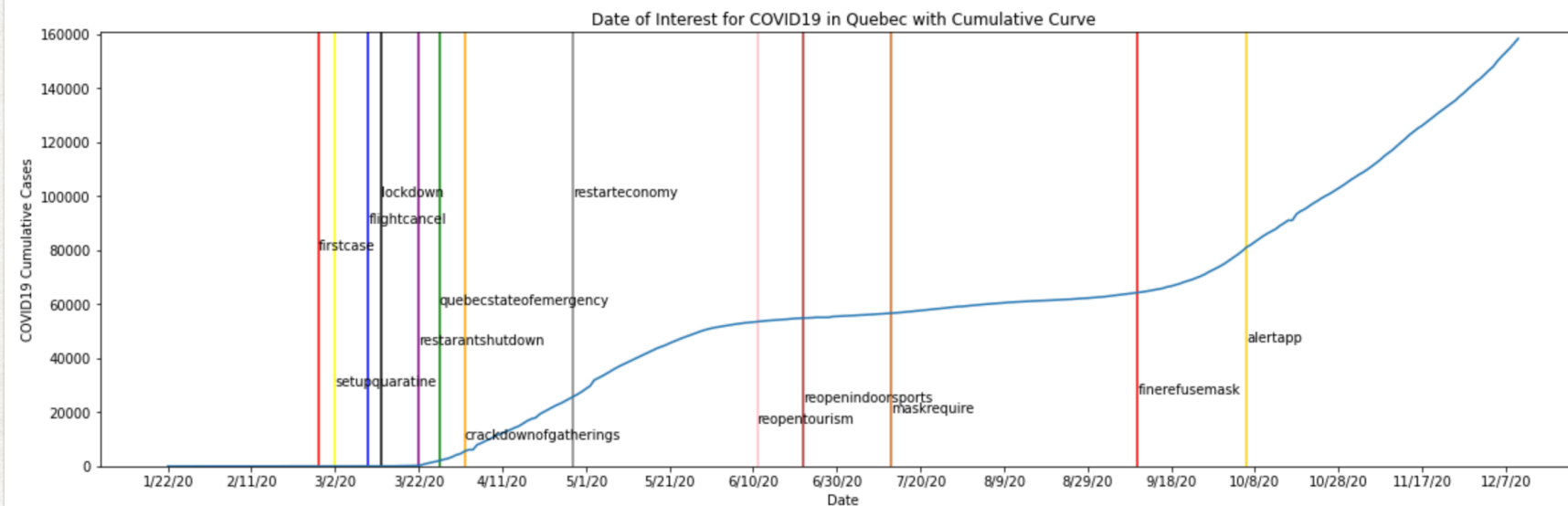
- Projection line is much lower than the actual line, and shows a slow-growth trend, after almost one year, the cumulative COVID19 cases in Canada is not over than 50000 people.
- The best case happen for government promulgates effective policies in correct time slot, and all people follow the rules, new vaccine is developed to effective cure the disease.



- Similar to actual cumulative curve but not shown an obvious fluctuate
- The validity of the results relies on the associated assumptions of the SIR model which may be violated, unreported cases as well as the inbound and outbound travellers would make the model fitting flawed

- At that time, the confirm cases show a fast increase trend, some people do not wear mask, most than one parade per every week, shopping malls are filled with people. The protective policies are not playing their roles.
- The projection line is much higher than the actual line, and shows a fast-growth trend, after almost one year, almost everyone in Canada will be infected

Relating COVID-19 Projections to Quebec



- There are total thirteen policies list in the Quebec COVID19 cumulative curve
- The three plots below cut the a period of time that behind some policies promulgate.

- The trend of COVID19 cumulative cases after lockdown in Quebec is similar to the best projection curve. It tells that the lockdown policy is useful to control the COVID19 spread. Within one month, the total infection is under 15000 cases. This situation is not only depend on the lockdown policy, at the same time period, the flight cancellation begin and large event cancelled, etc. These policies also help to slow down the growth confirm cases.
- The trend of COVID19 cumulative cases after mask requirement in Quebec is similar to the best projection curve. It tells that the mask requirement policy is useful to control the COVID19 spread. Within one month, the total infection within 15000 cases. But during this time, at Jun 25, CBC reported a News that "Quebec stops publishing daily COVID-19 data despite leading country in number of cases", so that the dataset in these period may not accuracy, part of confirm cases are not recorded.
- The trend of COVID19 cumulative cases after fine for refuse mask in Quebec is similar to the worst projection curve. It tells that the fine for refuse mask policy is not useful to control the COVID19 spread compared to other policies. Within one month, the total infection close to 20000 cases and show an exponential increase trend. The social situations during these time, gather a crowd over 200 people is allowed, public entertainments are reopened. The policy published time is not at the correct point and other reopen policies effect the infection number.

