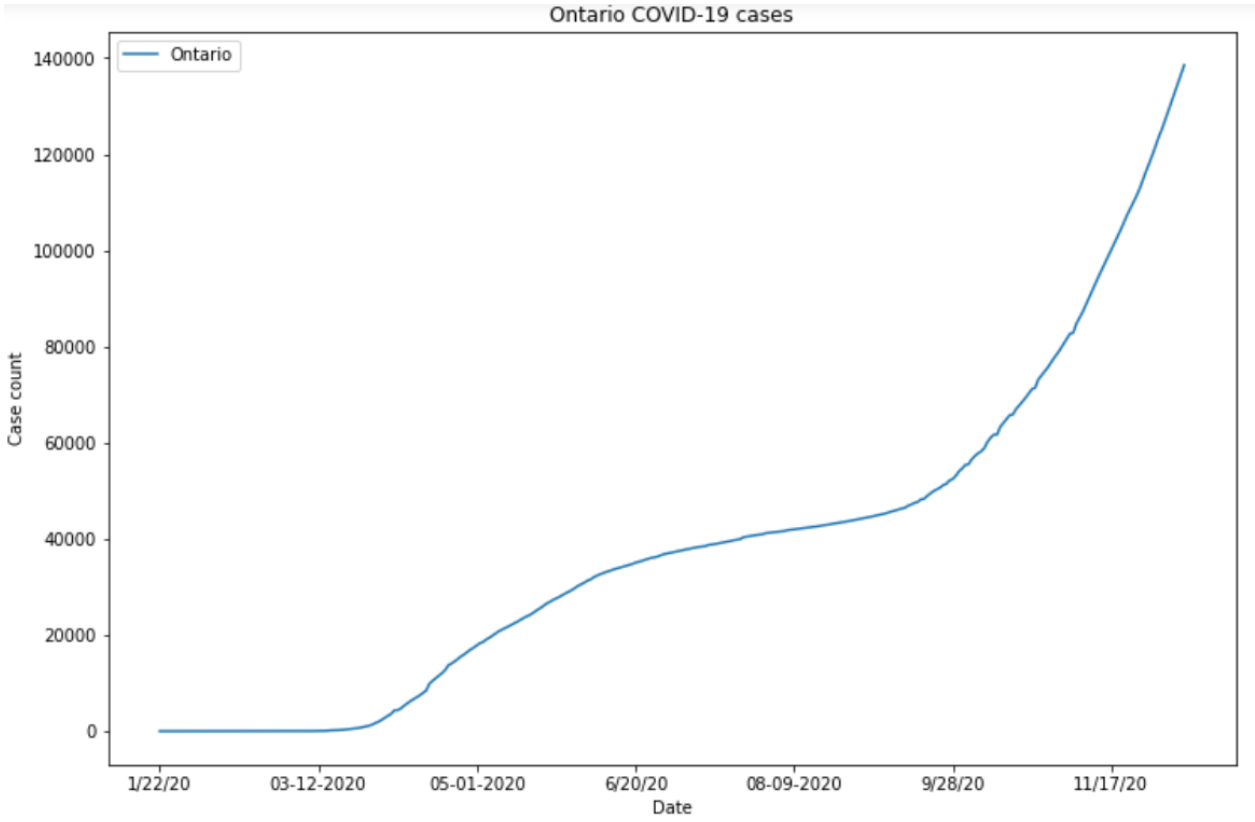
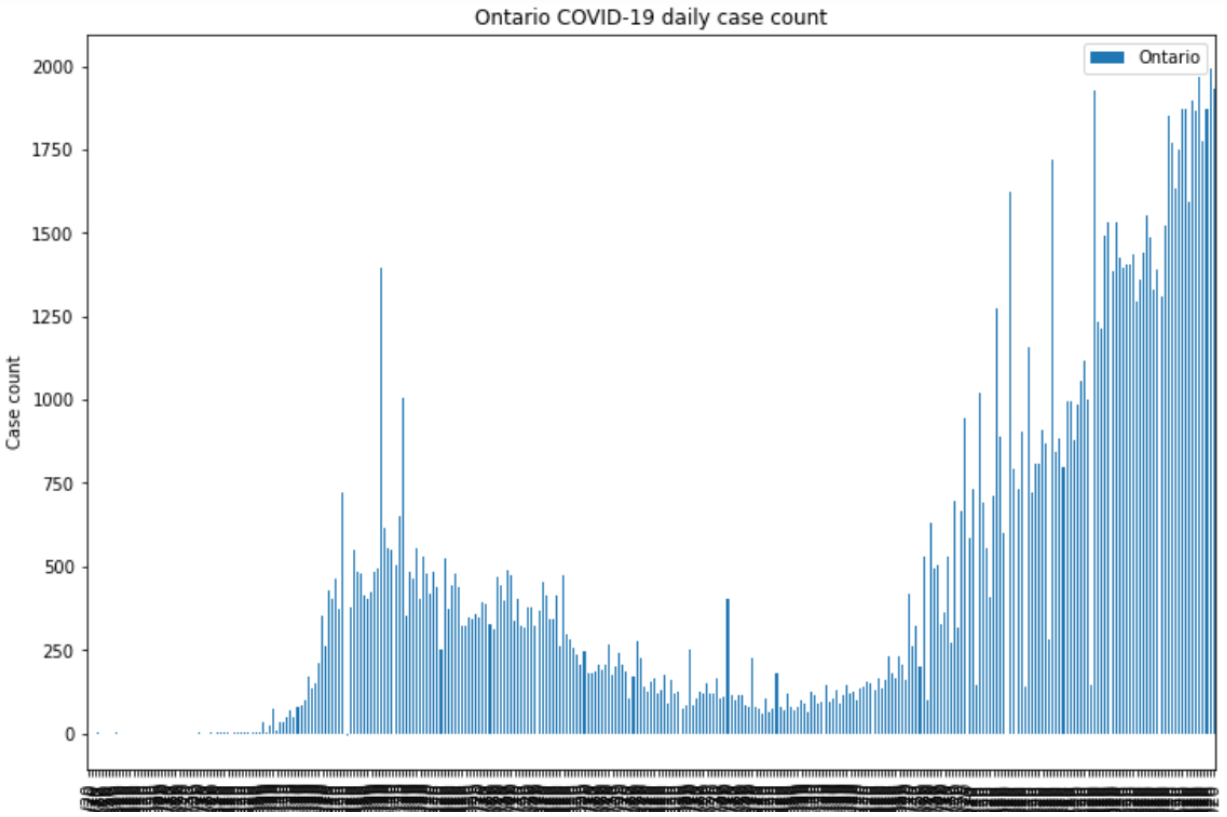
Time series forecasting using ARIMA

Exploratory data visualisation

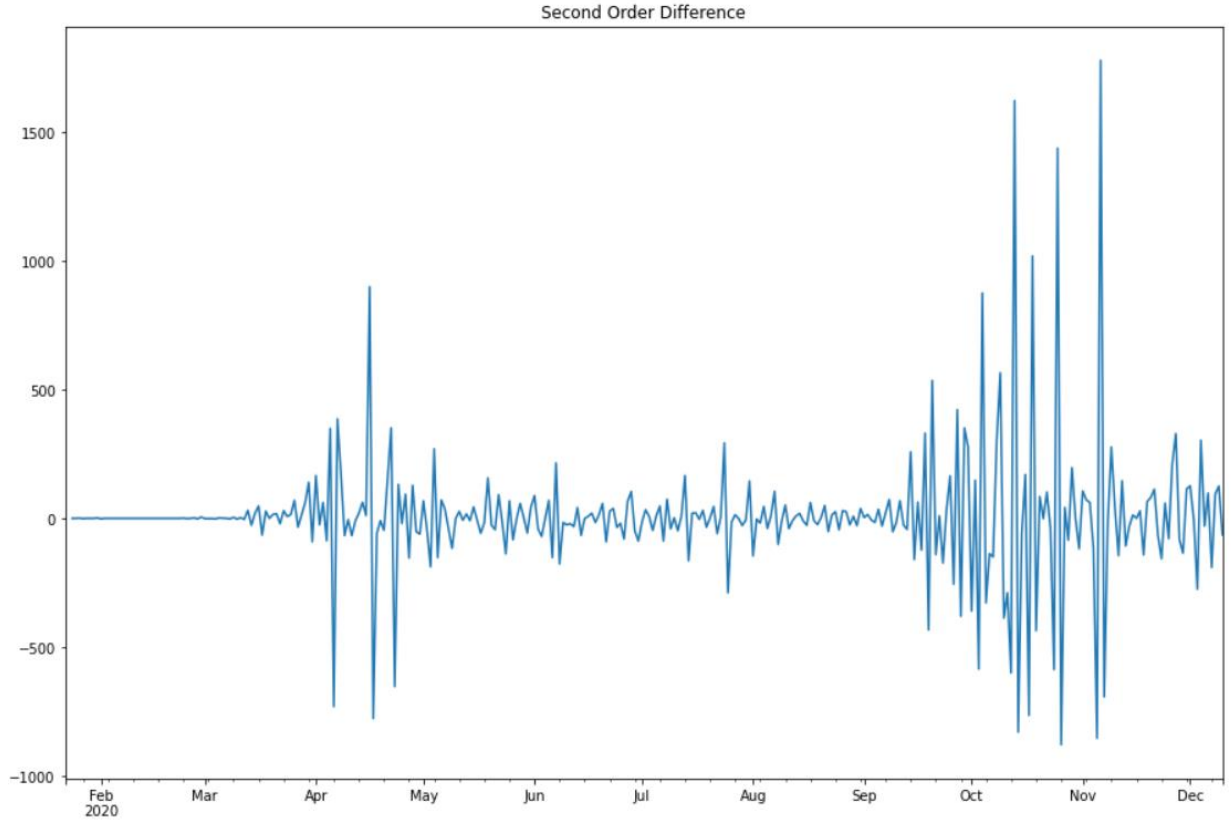


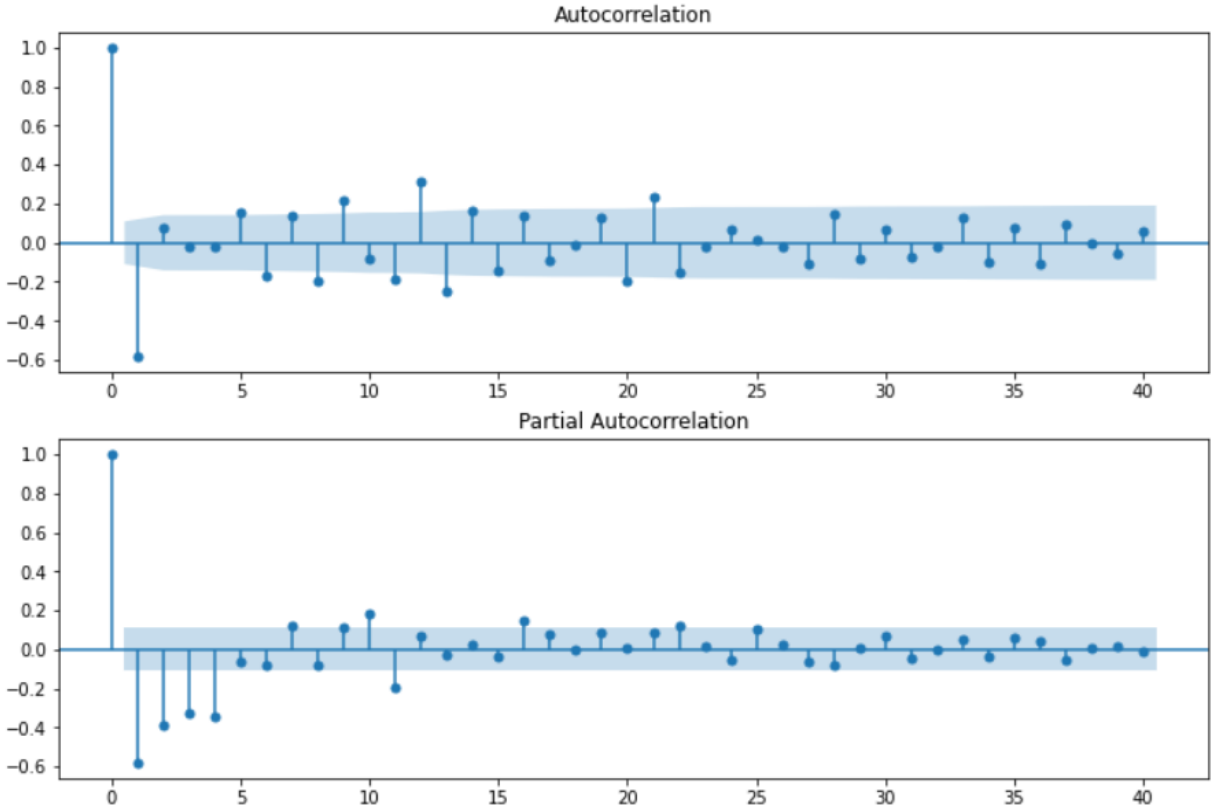


- COVID-19 virus spread was not prominent in Ontario until the end of March 2020. The virus started to spread in April and gradually increased in the summer from April to the end of September.
- COVID cases in Ontario increased exponentially from October to the beginning of December and is still increasing exponentially.
- It can be seen that the virus spread decreased from July to mid-September

ARIMA model selection

- ARIMA model involves three parameters, Autoregressive (p), differencing(d), moving average (q)
- Differencing order is identified as 2, since the model is stationary after second order differencing
- Autocorrelation function and the partial autocorrelation function show that 1 and 2 are promising p and q values respectively
- Akaike Information Criterion confirm that 1 and 2 are appropriate p and q values, respectively

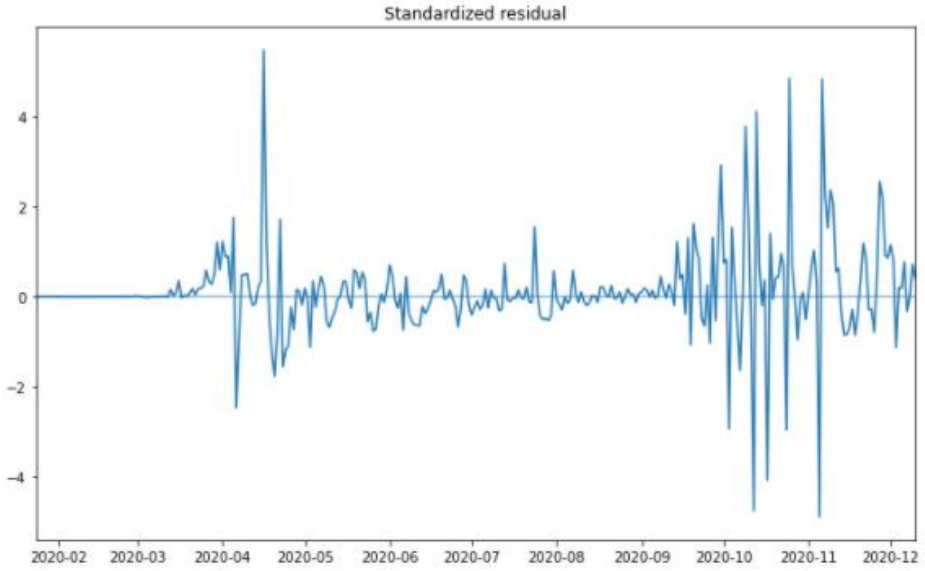


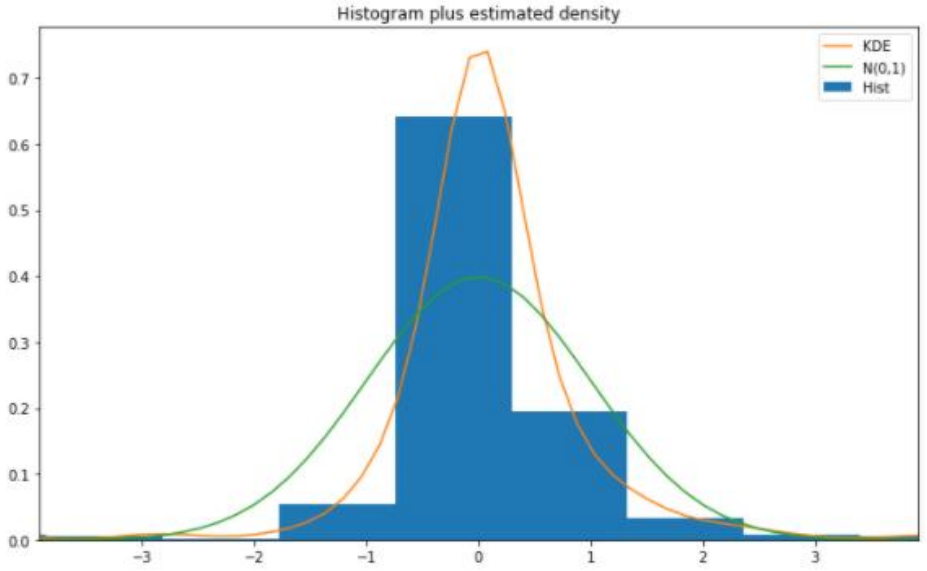


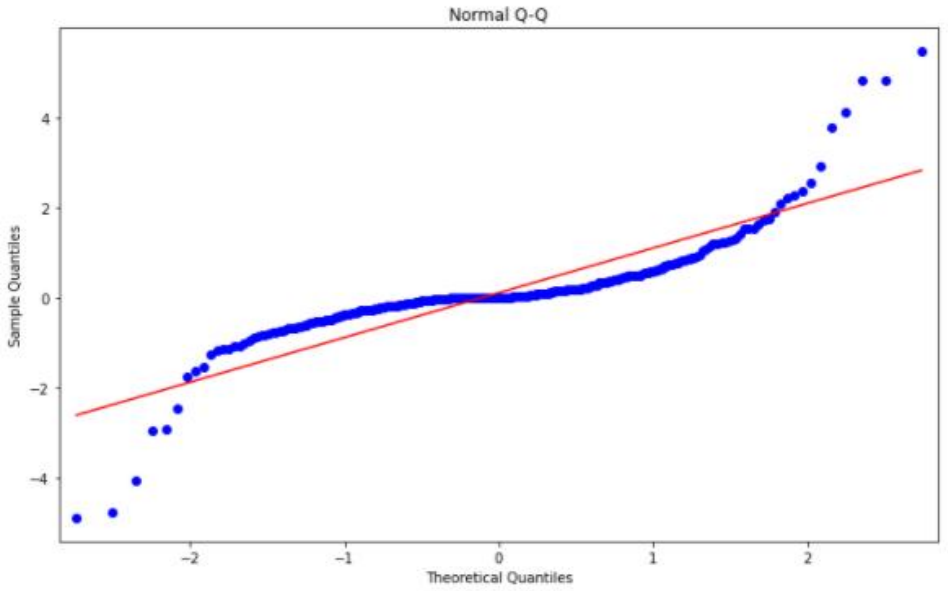
	p	q	aic	bic
5	1	2	4247.233437	4262.331644
8	2	2	4248.002226	4266.874984
2	0	2	4254.572479	4265.896134
7	2	1	4261.144515	4276.242721
4	1	1	4265.323771	4276.647425
1	0	1	4281.196165	4288.745268
6	2	0	4308.659029	4319.982684
3	1	0	4357.621008	4365.170111
0	0	0	4486.210471	4489.985022

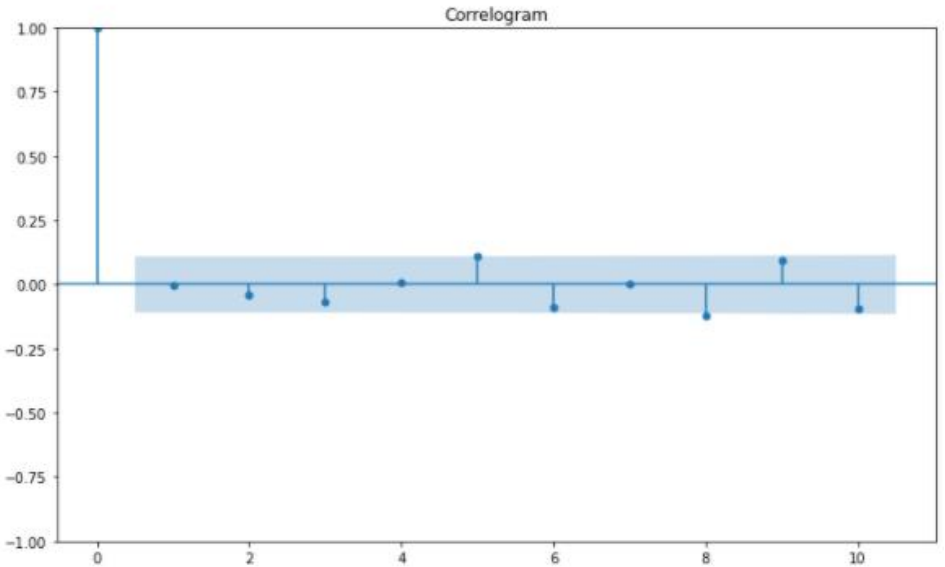
Model validation using Akaike Information Criterion

Model evaluation





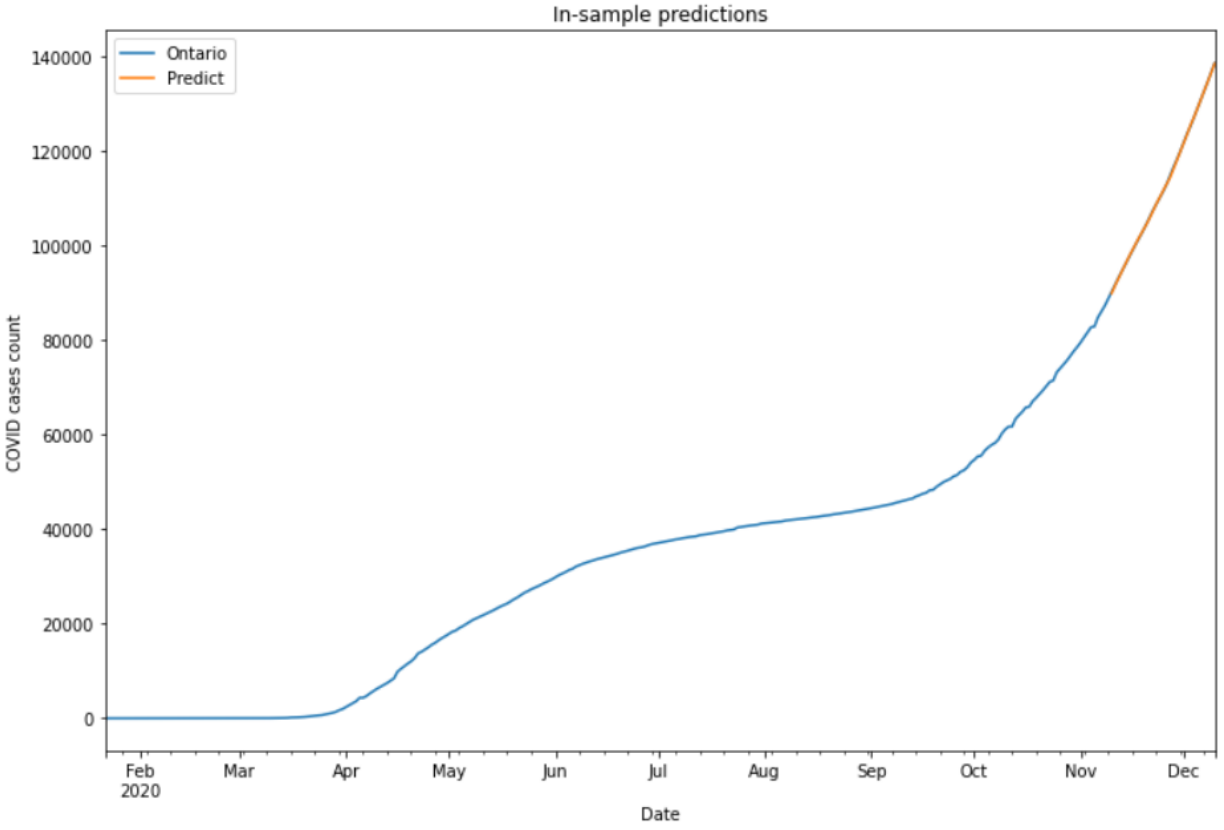


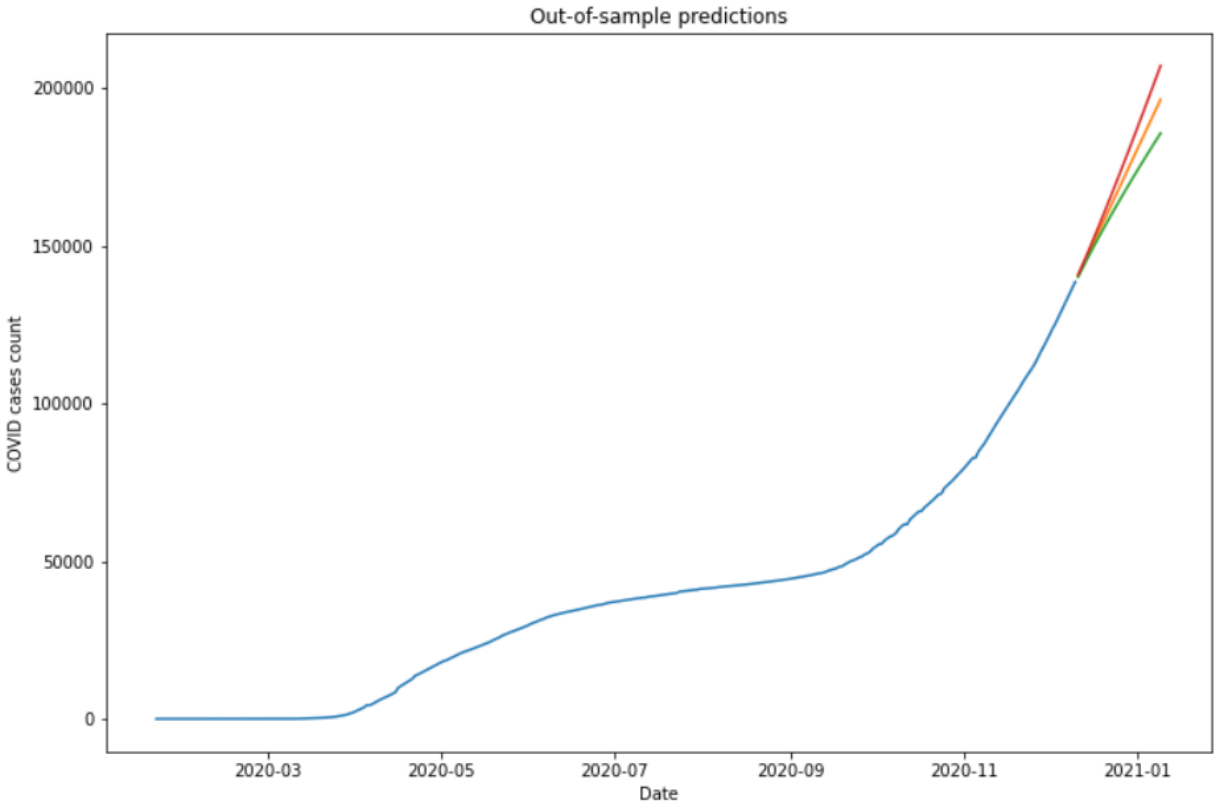


- The residual plot show that the data is made stationary
- Histogram and the Q-Q plot show that the residuals are normally distributed and that our model is good for prediction.
- Correlogram show that the residuals are not correlated to the data.
- So, our model has captured the necessary information from the data to be deployed for prediction

Time series forecasting

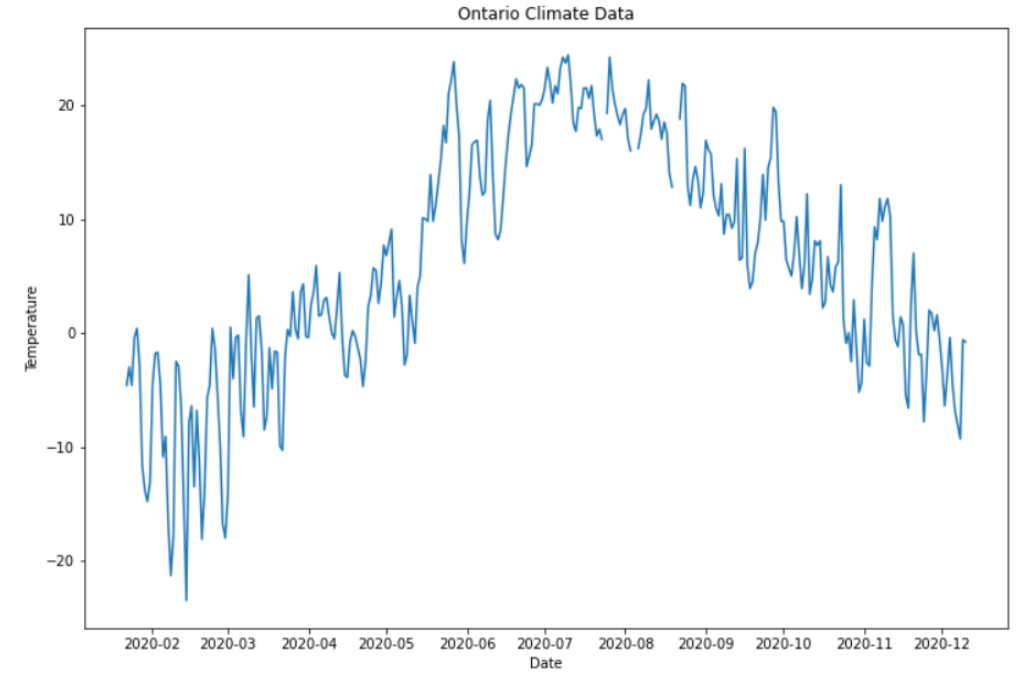
- In-sample prediction of the time series show good prediction accuracy
- Out-of-sample prediction show that the mean cumulative predicted cases at the end of 1 month will be 196338

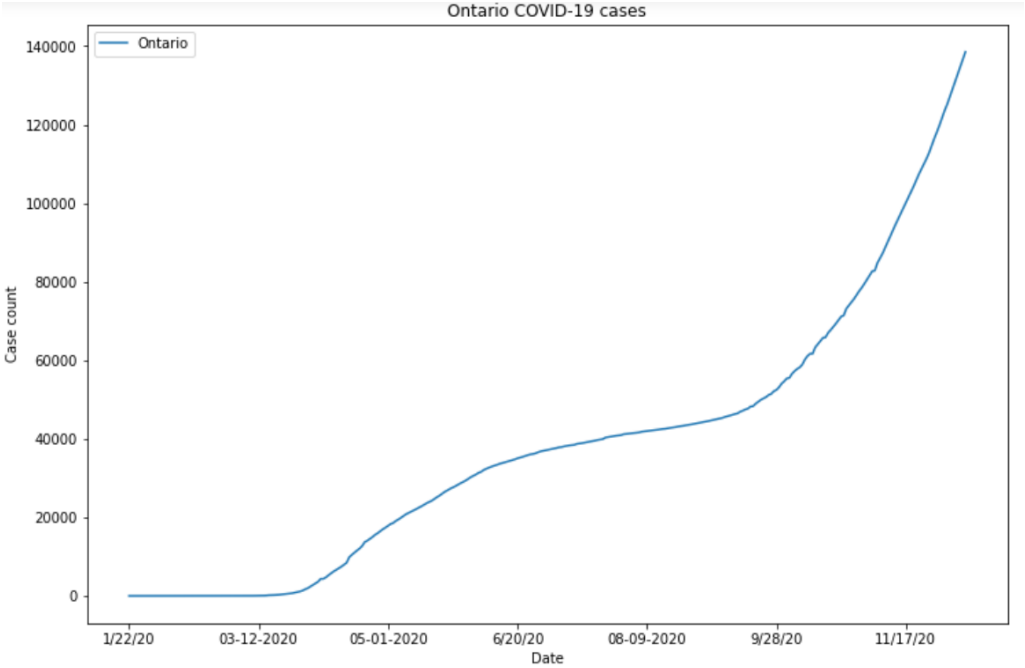




Date	Predicted cases
2021-01-05	188621.847662
2021-01-06	190551.006582
2021-01-07	192480.165502
2021-01-08	194409.324422
2021-01-09	196338.483342

Factors influencing COVID cases growth and conclusion





- The graphs show that COVID-19 cases growth in Ontario is considerably correlated with the temperature of the region
- Number of cases started to increase as the temperature increased. However, the growth stabilised in the peak summer due to government measures
- As the temperature dropped, cases increased exponentially. This could be because of the relaxations in COVID-related restrictions
- It is important to note that temperature is not the only factor that influences COVID cases growth. Other factors such as social distancing adherence, lockdowns and relaxations, etc. also contribute to the cases growth.
- Timeline of COVID related measures taken by government:

  March 17 - Declares state of emergency (some business closed)
  March 23 - Closure of non-essential business for 14 days
  March 30 - Closure of outdoor amenities for 14 days
  April 14 - Extension of state of emergency for 4 more weeks
  May 1 - Seasonal business and construction reopens
  May 6 - Emergency orders extended to May 19
  May 6 - Nurseries and garden centers reopen
  May 9 - Parks reopen
  May 14 - Phase 2 province reopen
  May 27 - Emergency order extended to June 9 (further extended to July 22)
  July 13 - Province enters phase 3 of reopening
  August 14 - Indoor recreation allowed
  September to October - COVID restrictions relaxed