

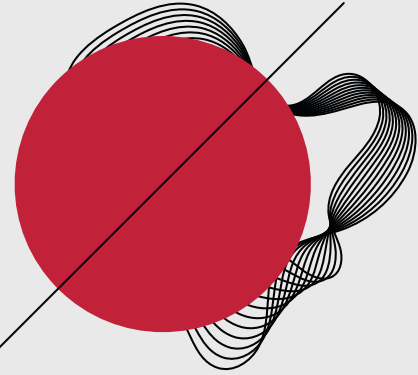
An Analysis of COVID-19 confirmed cases in New

York: Is the movement

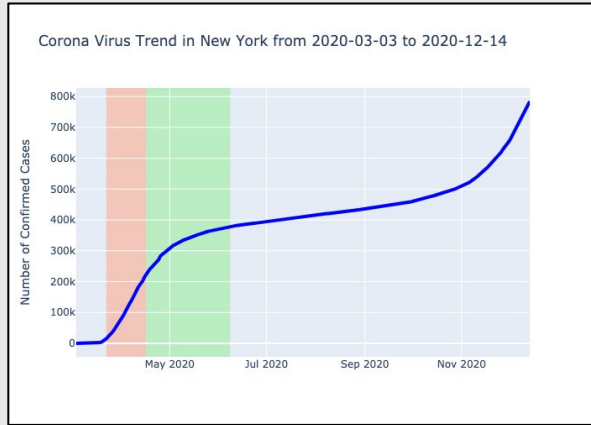
restriction important to prevent the COVID-19 pandemic?

MIE1624 Final Project

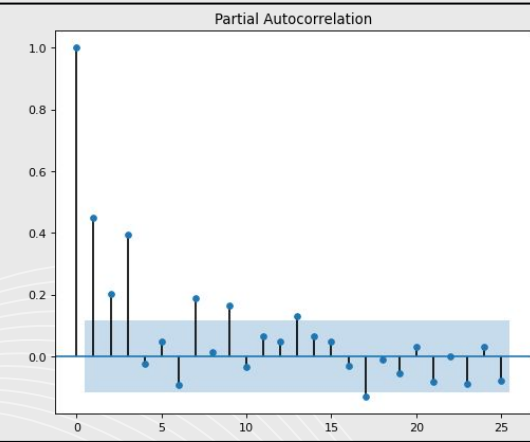
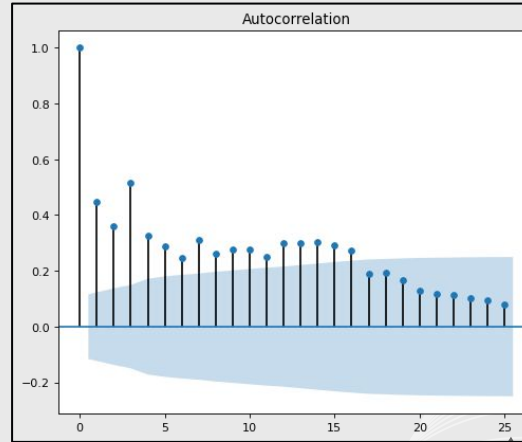
Jia Lei Qian (1002675544)



Time Serie Analysis on using JHU CSSE COVID-19 Daily Report Dataset-EDA



(Time serie trend plot on cumulative confirmed case in New York)



(ACF and PACF plots after transforming the data)

- ❖ **Overall increasing trend** of COVID-19 in New York from March to middle December.
- ❖ The orange area indicates that the number of confirmed cases increases rapidly. The second green highlight is between the extended stay-home order and phase 1 reopening.
- ❖ As the time of movement restriction increases, the growth rate of confirmed cases becomes slower, which implies **the effectiveness of movement restriction**.

- ❖ Autocorrelation function (ACF) graph and partial autocorrelation (PACF) graph help **find the initial number of ARIMA models**.
- ❖ The ACF describes the degree to which the present value of the series is related to its past value. Meanwhile, the PACF finds the residuals are correlated with the next lag value.
- ❖ From the PACF, we have a significant peak at lag 0. It is significant at every 3 lags. From the ACF plot, it is significant until lag 16.
- ❖ **PACF and ACF suggest AR(16) process with seasonal length of 3.**

Time Serie Analysis on using JHU CSSE COVID-19 Daily Report Dataset-SARIMA Model

SARIMA Model: SARIMA model is ARIMA model with a seasonal component. The components of the model is shown as below:

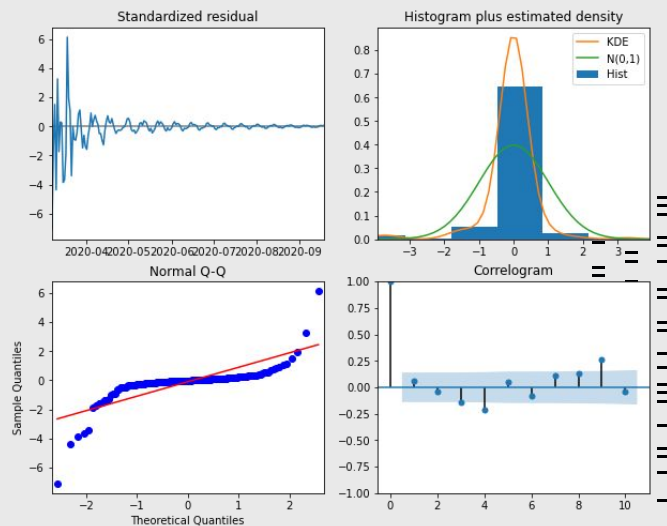
$$SARIMA \left(\underbrace{p, d, q}_{non-seasonal} \right) \left(\underbrace{P, D, Q}_m \right)_m$$

p and seasonal P: lags of the stationarized series
d and seasonal D: differencing that must be done to stationarize series
q and seasonal Q: lags of the forecast errors

Grid Search SARIMA: can choose the worst and best case model by trying different parameters based on AIC criteria

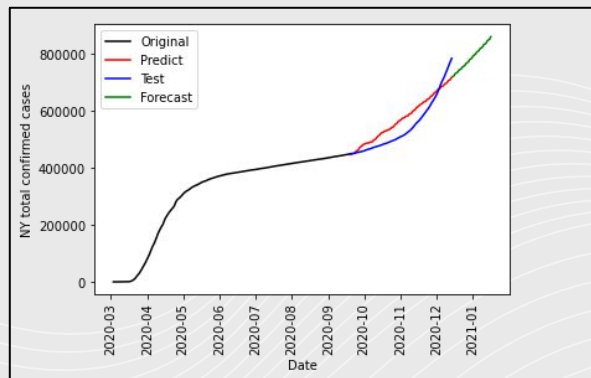
	Parameters	AIC
Worst Model	(0,1,0)(0,1,0)[3]	-246.001
Basic Model	(16,1,0)(0,1,0)[3]	-307.285
Best Model	(2,1,2)(2,1,2)[3]	-310.445

- ❖ **The standardized residual plot:** The residuals over time don't display any obvious seasonality
- ❖ **Histogram with estimated density plot:** The KDE line follows normal distribution, which indicates the residuals are approximately normally distributed.
- ❖ **The Q-Q-plot:** Shows that the ordered distribution of residuals (blue dots) follows the linear trend of the samples. However, several outliers exists.
- ❖ **The Correlogram plot:** Residuals have low correlation with their own lagging version

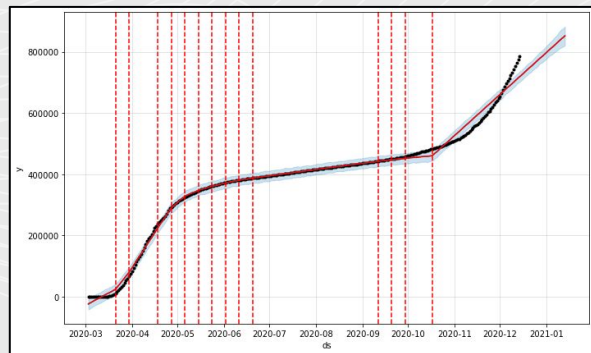


(Diagnosis of the best model) ▲

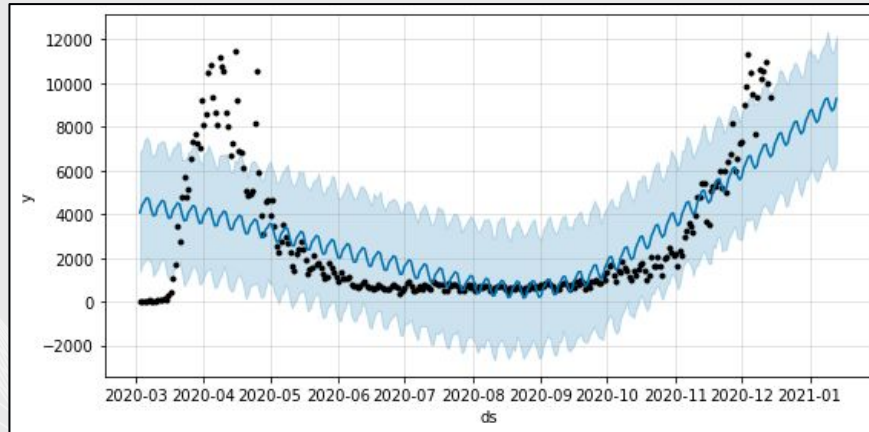
Time Serie Analysis on using JHU CSSE COVID-19 Daily Report Dataset-Performance



(Forecast of COVID-19 cumulative confirmed cases in New York in the upcoming 30 days using SARIMA)



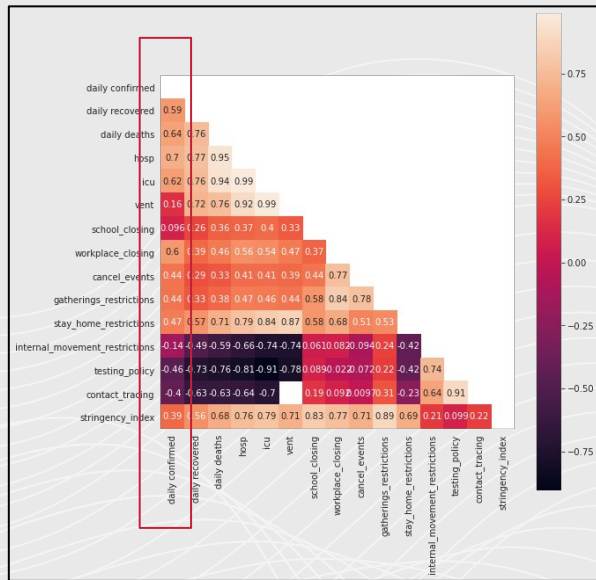
(Forecast of COVID-19 cumulative confirmed cases in New York in the upcoming 30 days using Prophet)



(Forecast of COVID-19 daily confirmed cases in New York in the upcoming 30 days using Prophet)

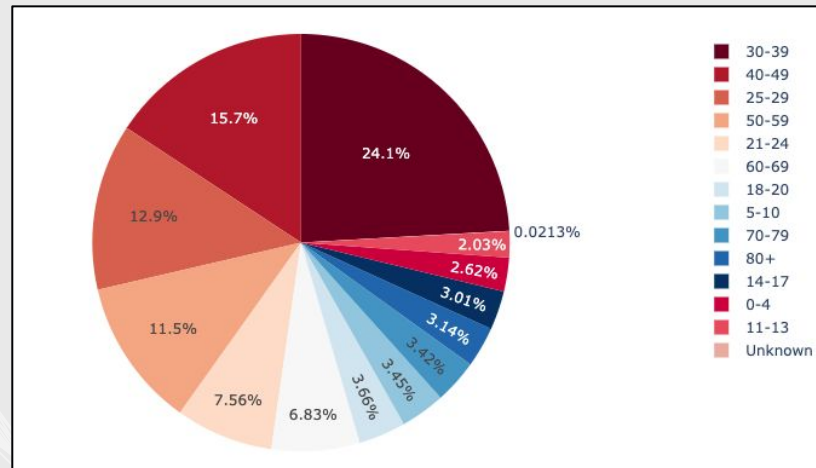
- ❖ The overall forecasting trend is increasing without following the original trend perfectly. It indicates that the daily confirmed tendency is **not as serious as the original dataset** even though there is an increasing trend of daily confirmed in New York next month.
- ❖ Prophet:
 - Notice that the overall **trend generated by Prophet is similar to the result generated by SARIMA**. Middle April and the middle of June are the critical periods that start with the movement restriction policy. The result shows that the pandemic has improved.
 - The daily confirmed case will keep increase, but the increase rate is relatively slower.

Study of the Impact of Movement Restriction on COVID-19 using Guidotti and Ardia (2020) and San Francisco Datasets -EDA



(Correlation Matrix of Guidotti and Ardia(2020) dataset)

- ❖ No factor highly correlated with the daily confirmed, and among all policy factors, **workplace closing has the highest correlation with the daily confirmed cases**
- ❖ Daily confirmed is **negatively correlated with internal movement restrictions**, and the association is small.



(COVID-19 Patients age wise in US from 2020/03/13 to 2020/12/14)

- ❖ The top three age groups with the highest percentage of daily case confirmed rate are 30-19, 40-49, and 25-29, respectively.
- ❖ -Exclude unknown, the least three age groups infected the COVID-19 are 11-13, 0-4, and 14-17.
- ❖ **Middle-age people are the majority of people infected with the disease**
- ❖ Since the middle-aged make up the bulk of society's workforce, I would like to study the relationship between daily confirmed cases and internal movement restrictions.

Study of the Impact of Movement Restriction on COVID-19 using Guidotti and Ardia (2020) and San Francisco Datasets -Model

Dep. Variable:	y	R-squared:	0.334
Model:	OLS	Adj. R-squared:	0.329
Method:	Least Squares	F-statistic:	69.51
Date:	Thu, 17 Dec 2020	Prob (F-statistic):	3.45e-25
Time:	18:12:03	Log-Likelihood:	-773.52
No. Observations:	280	AIC:	1553.
Df Residuals:	277	BIC:	1564.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	9.6029	0.303	31.644	0.000	9.005	10.200
0.0	-4.7195	0.843	-5.598	0.000	-6.379	-3.060
1.0	8.6260	0.407	21.196	0.000	7.825	9.427
2.0	5.6964	0.370	15.415	0.000	4.969	6.424

(OLS Regression Result)

group1	group2	meandiff	p-adj	lower	upper	reject
0.0	1.0	13.3455	0.001	10.5726	16.1184	True
0.0	2.0	10.4159	0.001	7.7019	13.13	True
1.0	2.0	-2.9296	0.001	-4.0743	-1.7849	True

(Post-Hoc Test result)

- ❖ Both results of linear regression and ANOVA are significant.
- ❖ Requiring closing will lead to fewer confirmed cases, implying a positive impact of movement control.
- ❖ Through the post-hoc test, one can conclude that **all restriction groups differ in their mean of daily confirmed cases, implying internal movement restrictions effectiveness.**
- ❖ Internal movement restriction does have a positive impact on reducing the spread of the pandemic. As an increasing trend forecasted previously, **the movement control is required to extend** to reduce the tendency of COVID-19 spread in New York

F_onewayResult(statistic=69.5059664790454, pvalue=3.4458498700867286e-25)

(One way ANOVA result)



Reference

- ❖ Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1
The direct link to data set is: <https://github.com/CSSEGISandData/COVID-19>
- ❖ Guidotti, E., Ardia, D., (2020), "COVID-19 Data Hub", Journal of Open Source Software 5(51):2376, doi: 10.21105/joss.02376.
The direct link to data set is: <https://covid19datahub.io/>
- ❖ Data acquired from <https://datasf.org/opendata/> on December 17 2020. Aggregated data on this page is made freely available under Open Data Commons Public Domain Dedication and License(<https://opendatacommons.org/licenses/pddl/1.0/>)
The direct link to data set is:
<https://data.sfgov.org/COVID-19/COVID-19-Cases-Summarized-by-Age-Group/sunc-2t3k>
- ❖ Levenson, E. (2020, March 26). Why New York is the epicenter of the American coronavirus outbreak. Retrieved from <https://www.cnn.com/2020/03/26/us/new-york-coronavirus-explainer/index.html>
- ❖ Kerr, A. (2020, October 06). Timeline: How COVID-19 Took Over NYC. Retrieved from <https://www.investopedia.com/historical-timeline-of-covid-19-in-new-york-city-5071986>
- ❖ Graves, A. (2020, July 21). Time Series Forecasting with a SARIMA Model. Retrieved from <https://towardsdatascience.com/time-series-forecasting-with-a-sarima-model-db051b7ae459>
- ❖ Prophet Retrieved December 17, 2020, from <https://facebook.github.io/prophet/>

