IAF603: Preparing Data For Analytics

Impact of various government responses and policies on confirmed cases/fatality ratio of Covid-19

December 6th, 2020

Vathana Him
Andrew Lindberg
Dajonna Williams

Abstract

This research will examine the implementation of public health policy of three first world countries in response to the COVID-19 pandemic and how each approach affected the confirmed cases/fatality ratio. The three countries that are targeted are South Korea, the United States, and Sweden. These three countries are known in the public health sector for their varying response to the pandemic with South Korea having a strict approach, United States having a moderate approach, and Sweden having a relaxed approach. (Soistmann & Trigonoplos, 2020) Through a linear regression model, this project aimed to see how effective the implementation of stringency index, government response index, health containment index, and economic support index are in stabilizing the virus among the three countries.

Introduction

The novel coronavirus COVID-19 is one of the most prominent infectious diseases during the 21st century; its transmission rate is incomparable to other viruses (Gates, 2020). After the World Health Organization declared the disease a pandemic on March 11, 2020, several countries had diverging ideas of how to curb and prevent the spread of the disease (Hale et al.,2020; Soistmann & Trigonoplos, 2020). COVID-19 created a unique environment where governments had to react quickly and implement various plans to stop the spread of the virus. Because government response varies across different countries, it is important to examine how impactful these variances are in preventing the spread of the virus. This information is likely to be useful for decision makers in the event of a future disease similar to the coronavirus. Many people believe that the likelihood of future pandemics is a matter of when, not if they will happen.

In order to examine the variance in disease spread, this project will look at the quantitative measurement of public health policy regarding COVID-19 from three first world countries in South Korea, United States and Sweden to explore that effect on the recorded confirmed cases and fatality ratio. This analysis will be done through a regression test and model. Thus, this research will examine the possibility of a relationship between public health policy on Covid-19 confirmed cases and fatality ratio.

Description of dataset

Government Response dataset: This dataset was obtained from the University of Oxford's github repository, which quantitatively measures each country's responses to the Covid-19 pandemic (Hale et al., 2020).The University of Oxford dataset had 17 variables that were used to measure government responses ranging from school closure indexes to government response rate indexes. These variables were calculated based on a scale of 0-100 with 100 being the most strict. However, only StringencyIndex, GovernmentResponseIndex, ContaimentHealthIndex, and EconomicSupportIndex are variables targeted. For clarification, StringencyIndex is the overall measurement of the strictness of "lockdown style" policies. (Hale et al.,2020)

Confirmed/Confirmed death dataset: This dataset was obtained from John Hopkins github repository, which contained total confirmed cases and death cases on a day to day basis. The fatality ratio will be calculated based on the WHO formula of total death/total confirmed. Each dataset was updated on 9/21/2020, which was the retrieved date for this project.

Methods

The dataset that was collected from both repositories of John Hopkins and Oxford University was retrieved as a csv file and analyzed through Python. Because each respective dataset contains different formatting and data types, some data cleaning steps were performed before the analysis. These steps included: converting data type of each column into appropriate types, converting dates into appropriate format so they can be joined together through a database, dropping unnecessary columns, and creation of a new column that contains fatality ratio for exploration. Since the government response indexes dataset contains measurements for the United States by state name and dates and other targeted countries in Sweden and South Korea had measurements by dates, the United States dataset was cleaned heavily by taking an aggregate value of grouped dates and averaging that sum. The main purpose for this process was to create similar data formatting with other countries of interests.

After the cleaning process, each respective dataset was inserted into a separate database respectively. Two database tables were created for this assignment, the 'POLICY_DATA' table which contains daily government response data and the 'COVID_DATA' table which contains daily total number of confirmed cases and fatality ratio. The data types for each table were of type float, text, and datetime, which represents measurement values, location and date stamp respectively. Subsequently, these two tables were then joined together into a bigger table by using the date of each respective table as the key pair.

Analysis

The first phase of this analysis was done through descriptive analysis, which was produced with simple descriptive statistics and visualization methods through the seaborn package. This descriptive analysis was done separately based on each individual country's predictor variables: total case, total death, fatality ratio, stringency index, government response index, containment health index, economic support index, log of total case. However, for the purpose of this part of the analysis, total case ,total death and log of total case were not taken into consideration due to differences in population scale, thus fatality ratio was considered as a substitute.

In table 1, the descriptive analysis for Sweden is presented. It was observed that the mean fatality ratio was approximately 0.07, this resulted from the mean of stringency index, government response index, containment health index, and economic support index of approximately 32.8, 39, 36.8, and 50.7 respectively. Subsequently, in table 2, the analysis for South Korea is presented. The observed mean fatality ratio was approximately 0.01. The mean of stringency index, government response index, containment health index, and economic support index were approximately 50.0, 54.6, 58.0, and 36.2 respectively. Finally, in table 3, the United States had an observed mean fatality ratio 0.035. Their average stringency index, government response index, containment health index, and economic support index were approximately 53.2, 54.6, 54.7, and 53.7 respectively.

Based on this analysis, it was observed that South Korea had the lowest fatality ratio and Sweden with the highest fatality ratio. This difference was attributed to their strictness in the level of response to the pandemic based on the observed data of stringency, government response, containment health, and economic support index. Sweden had a dramatically lower response index than South Korea in every variable with the exception of the economic response index. However, the United States had higher response indexes and fatality ratios than South Korea. These contradictory findings suggest that the level of response indexes may have not had an impact on fatality ratios, and that further analysis is necessary to understand how a stricter response is related to a lower fatality rate in one comparison and a higher one in another.

Given that spread of the virus continued and policies changed over time, a time series graph of the response indexes was analyzed. In figure 1, it was observed that South Korea had implemented their response indexes at a stricter level earlier than the United States as the country was hit by the pandemic first. Subsequently, the max value of each response index of South Korea was higher than the United States with reference to table 2 and table 3 with the exception of the economic support index. This could explain why South Korea was able to maintain an overall lower mean fatality ratio and mean response indexes than the United States as their response indexes were more strict at the start of the pandemic and then progressively decreased.

The second part of this analysis was done through a regression analysis to help build a predictive model in determining the relationship between the response indexes and the log of confirmed case and/or fatality ratio. Prior to the start of the regression modeling, an exploration of the relationship between response indexes to log confirmed cases and fatality ratio was examined respectively. Scatterplots were used to examine this relationship to see if the independent variables in response indexes had a linear impact on the dependent variables in log confirmed cases and fatality ratio.

For this analysis, figure 3 and figure 4 were observed. Figure 3 represents the scatterplot of response indexes with respect to log confirmed cases. Based on this figure, it was observed that stringency, government response and containment health indexes had a roughly linear relationship with log confirmed cases. However, this linear relationship was not discovered for the relationship between Economic Support Index and log confirmed cases. Furthermore, in figure 4, it was observed that neither of the response indexes had a linear relationship with fatality ratio. Thus, it was decided that the regression analysis should be done with respect to the relationship between response indexes and log confirmed cases.

The regression analysis was done using pySpark. The initial step for building this model was to select the target predictor variables that were needed. The predictor variables included Stringency Index, Government Response Index, Containment Health Index, and Economic Support Index. By vectorizing all the assembled features or variables of this regression analysis, a regression predictive model was created for each respective country.

Based on figure 5, the models that were created had a pretty decent R^2 value with South Korea, Sweden, and the United States models having R^2 values of 0.76, 0.91, and 0.87. Subsequently, a 7:3 split ratio of the train and test data was done to further validify the accuracy of the model. For South Korea, the RMSE of the train and test models were 0.52 and 0.57 respectively. For Sweden, the RMSE of the train and test models were 0.58 and 0.58 respectively. Finally, for the United States, the RMSE of the train and test models were 0.62 and 0.55. The R^2 value of the test data was also approximately the same as the train data.

Results

Based on the result of this analysis, it was determined that the response indexes to Covid-19 had a positive linear impact on the log of total confirmed cases. With evidence from figure 3 and figure 4, it was determined that this linear relationship only persists with respect to the relationship between response indexes and log of total confirmed cases rather than response indexes and fatality ratio. Additionally, the linear relationship between the responses indexes to log of total confirmed cases prove to be a positive linear relationship as evidence from the supporting R^2 and RMSE of both the train and test datasets.

The relatively high R^2 values for the models of each respective country proved that the observed data fit the model relatively well. Thus, providing evidence that variances between the fitted value and the observed data was relatively small. Additionally, the RMSE value for both the train and test datasets also provided another indicator of the validity of this model as RMSE values for both train and test set for each respective model were reasonably similar with only slight variations in RMSE values. This further provided support for the notation that a positive linear relationship does exist between response indexes and log of total confirmed cases.

Conclusion

______With divergence in public policy responses to the pandemic among different countries, it was clear that governmental response rate had an impact on the spread of COVID-19 pandemic as seen through this project. It was concluded that a positive linear relationship does exist between crucial factors such as the strictness of government “lockdown” policies and containment of the virus to the log of confirmed cases of the virus. South Korea can be used as an example, as the country began to be hit with the virus, the strictness level of their policy was high, which then helped stabilize confirmed cases (figure 3). The level of strictness in public policy that impacted lockdown, containment, and government response rate did help slow the spread of the virus by reducing the rate of confirmed cases as it was hypothesized before the start of the project.

However, it should be noted that this conclusion should be taken with limitations. This analysis was done for only three countries and more countries should be taken into consideration. Additionally, other confounding variables such as cultural and political factors were not taken into consideration.

Table 1: Sweden Descriptive Statistics

	Total_Case	Total_Death	Fatality_Ratio	Stringency_Index	Government_Response_Index	Containment_Health_Index	Economic_Support_Index	log_total_case
count	218.000000	218.000000	218.000000	218.000000	218.000000	218.000000	218.000000	218.000000
mean	37245.972477	3238.013761	0.073813	32.868119	38.976560	36.833899	50.745413	3.746791
std	31984.976245	2416.440040	0.046332	16.418287	13.876847	12.178615	24.034117	1.632382
min	1.000000	0.000000	0.000000	0.000000	10.260000	12.120000	0.000000	0.000000
25%	2784.500000	115.750000	0.041443	32.410000	39.740000	35.610000	62.500000	3.444331
50%	31476.000000	4081.000000	0.076789	38.890000	45.510000	42.420000	62.500000	4.497955
75%	72756.500000	5651.750000	0.114499	46.300000	48.080000	45.450000	62.500000	4.861872
max	84985.000000	5835.000000	0.138975	46.300000	50.640000	48.480000	62.500000	4.929342

Table 2: South Korea Descriptive Statistics

	Total_Case	Total_Death	Fatality_Ratio	Stringency_Index	Government_Response_Index	Containment_Health_Index	Economic_Support_Index	log_total_case
count	230.000000	230.000000	230.000000	230.000000	230.000000	230.000000	230.000000	230.000000
mean	10000.204348	199.195652	0.016341	49.968087	54.616261	57.964043	36.195652	3.598567
std	5346.839501	116.549710	0.008260	17.380605	16.318512	17.061932	21.595345	1.092507
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	8343.250000	82.250000	0.009857	43.520000	53.210000	53.790000	25.000000	3.921330
50%	10976.500000	259.500000	0.020873	54.170000	56.730000	62.500000	50.000000	4.040464
75%	13326.750000	287.750000	0.022511	56.940000	62.500000	64.770000	50.000000	4.124724
max	20842.000000	331.000000	0.023964	82.410000	80.130000	85.610000	50.000000	4.318939

Table 3: United States Descriptive Statistics

	Total_Case	Total_Death	Fatality_Ratio	Stringency_Index	Government_Response_Index	Containment_Health_Index	Economic_Support_Index	log_total_case
count	2.290000e+02	229.000000	229.000000	229.000000	229.000000	229.000000	229.000000	229.000000
mean	1.890331e+06	76382.868996	0.034796	53.245123	54.571834	54.737488	53.675476	5.023705
std	1.946225e+06	65814.474998	0.020845	25.404093	23.694261	22.495144	31.271783	2.130536
min	1.000000e+00	0.000000	0.000000	0.445849	6.706415	7.886792	0.235849	0.000000
25%	6.427000e+03	108.000000	0.019315	55.668679	48.699245	53.523962	22.169811	3.808008
50%	1.390746e+06	84133.000000	0.035055	61.660566	65.844151	64.829434	72.169811	6.143248
75%	3.118008e+06	133291.000000	0.054932	70.605472	70.434528	69.904528	74.056604	6.493877
max	6.201726e+06	187765.000000	0.072000	79.193019	74.177358	74.199434	75.471698	6.792513

Figure 1:

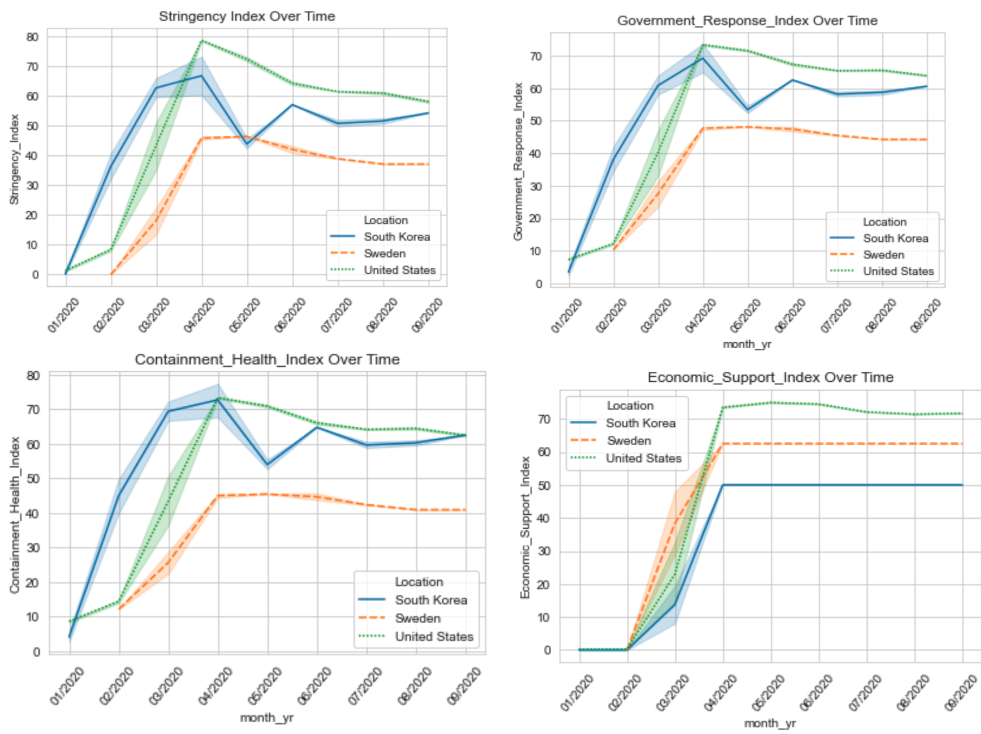


Figure 2:

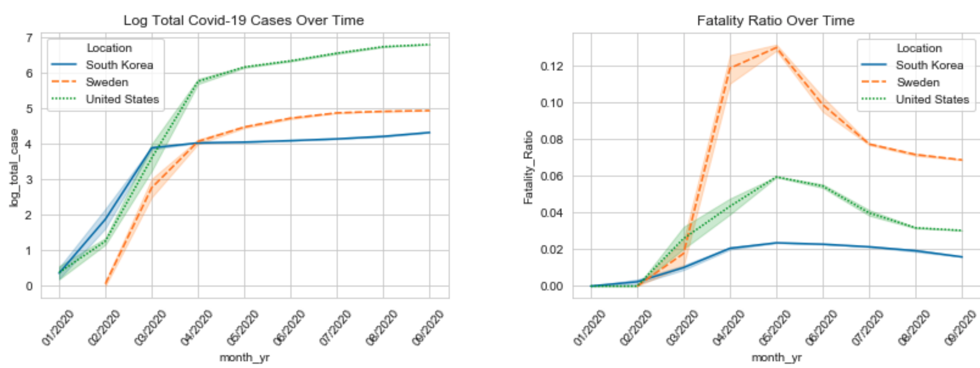


Figure 3:

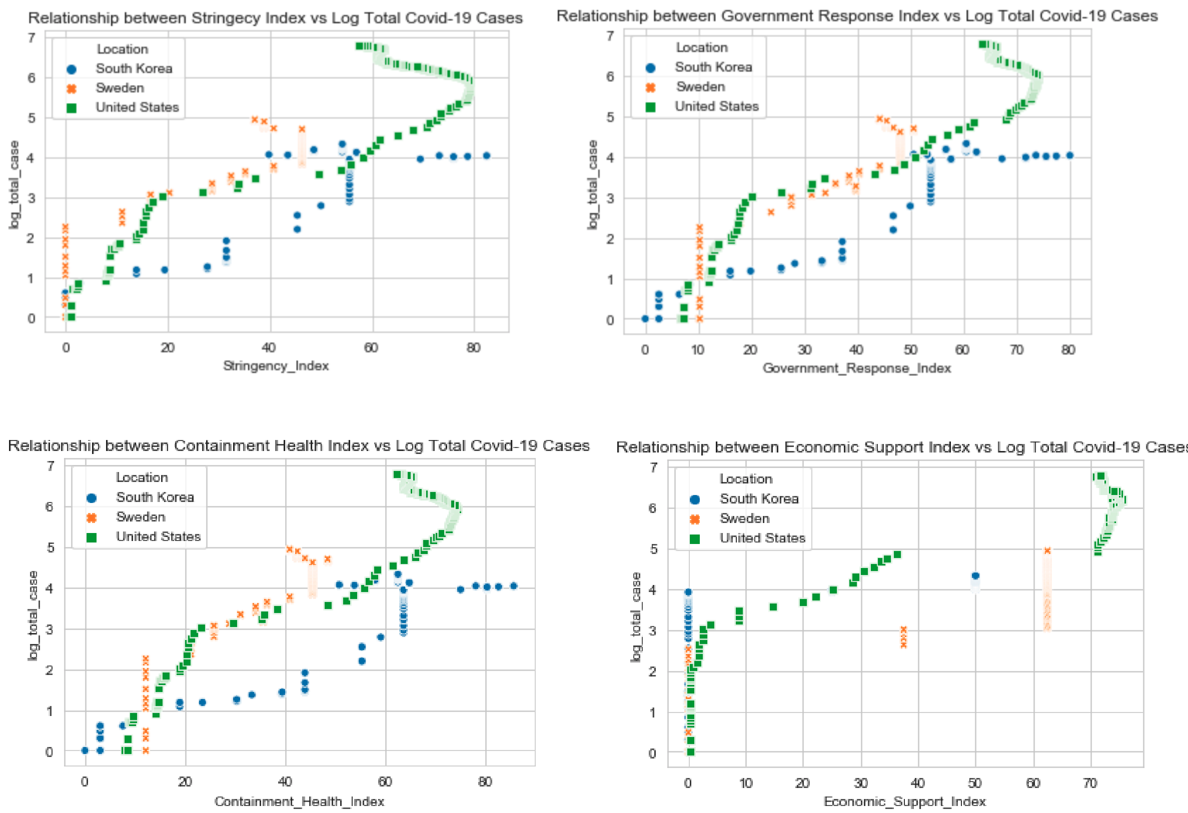
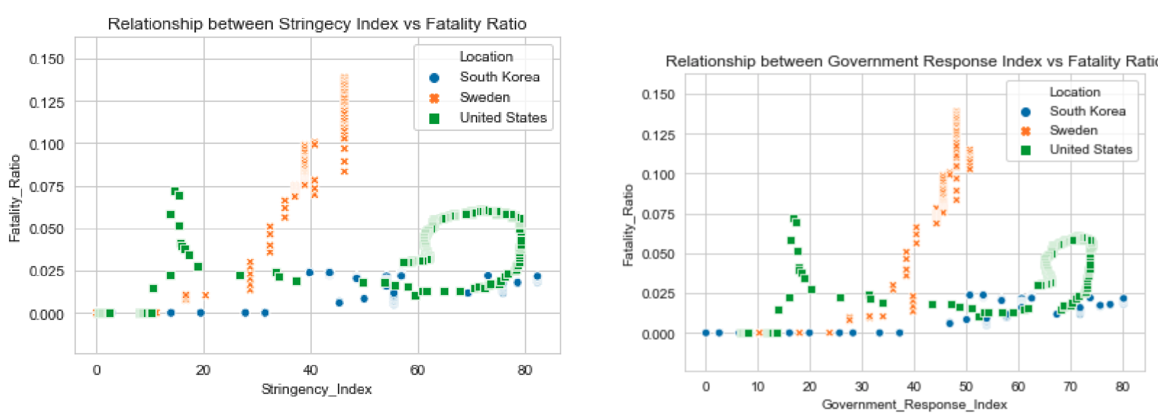
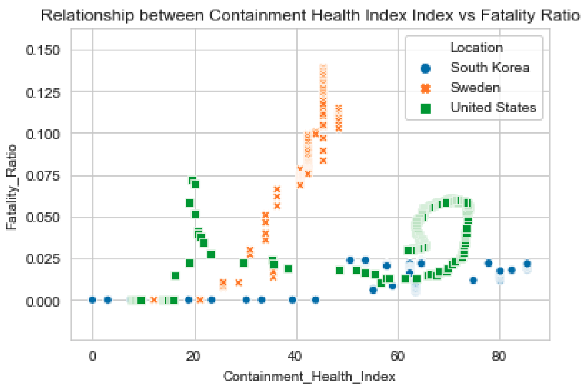

Relationship between Stringecy Index vs Log Total Covid-19 Cases
Location
South Korea
Sweden
United States
log_total_case
Stringency_Index
Relationship between Government Response Index vs Log Total Covid-19 Cases
Location
South Korea
Sweden
United States
log_total_case
Government_Response_Index
Relationship between Containment Health Index vs Log Total Covid-19 Cases
Location
South Korea
Sweden
United States
log_total_case
Containment_Health_Index
Relationship between Economic Support Index vs Log Total Covid-19 Cases
Location
South Korea
Sweden
United States
log_total_case
Economic_Support_Index

Figure 4:


Relationship between Stringecy Index vs Fatality Ratio
Location
South Korea
Sweden
United States
Fatality_Ratio
Stringency_Index
Relationship between Government Response Index vs Fatality Ratio
Location
South Korea
Sweden
United States
Fatality_Ratio
Government_Response_Index



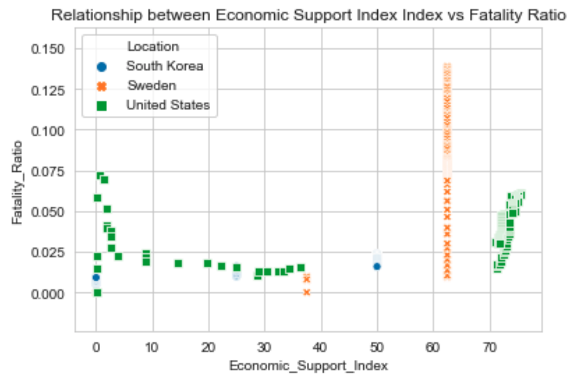


Figure 5:

```
South Korea Coefficients: [0.0,0.02558256997108275,0.004602145578880234,0.012443870337403488]
South Korea Intercept: 1.4931330114778036
South Korea RMSE: 0.522679
South Korea r2: 0.760505
+------------------+-------------------+--------------------+
|        prediction|     log_total_case|            features|
+------------------+-------------------+--------------------+
|1.5725688917077827|                0.0| [0.0,2.56,3.03,0.0]|
|1.5725688917077827| 0.3010299956639812| [0.0,2.56,3.03,0.0]|
|1.5725688917077827|0.47712125471966244| [0.0,2.56,3.03,0.0]|
| 2.109517054995326| 1.1760912590556813|[19.44,19.87,23.4...|
|2.2885151165764364|  1.255272505103306|[27.78,25.64,30.3...|
| 2.527078582966084|  1.380211241711606|[31.48,33.33,39.3...|
| 2.527078582966084| 1.4313637641589874|[31.48,33.33,39.3...|
+------------------+-------------------+--------------------+
only showing top 7 rows

South Korea R Squared (R2) on test data = 0.7451769616421524
South Korea Root Mean Squared Error (RMSE) on test data = 0.575011

Sweden Coefficients: [0.0,0.03422443998083798,0.029225008223191216,0.019589942473522172]
Sweden Intercept: 0.31605774799948344
Sweden RMSE: 0.575814
Sweden r2: 0.872788
+------------------+--------------+--------------------+
|        prediction|log_total_case|            features|
+------------------+--------------+--------------------+
|1.0214076018679588|           0.0|[0.0,10.26,12.12,...|
|1.0214076018679588|           0.0|[0.0,10.26,12.12,...|
|1.0214076018679588|           0.0|[0.0,10.26,12.12,...|
|1.0214076018679588|           0.0|[0.0,10.26,12.12,...|
|1.0214076018679588|           0.0|[0.0,10.26,12.12,...|
|1.0214076018679588|           0.0|[0.0,10.26,12.12,...|
|1.0214076018679588|           0.0|[0.0,10.26,12.12,...|
+------------------+--------------+--------------------+
only showing top 7 rows

Sweden R Squared (R2) on test data = 0.8796650866668344
Sweden Root Mean Squared Error (RMSE) on test data = 0.575167
```

```
United States Coefficients: [0.0,0.022037878227385933,0.01795725062881903,0.028676119054643304]
United States Intercept: 1.282603918439405
United States RMSE: 0.626869
United States r2: 0.918693
+------------------+------------------+--------------------+
|        prediction|    log_total_case|            features|
+------------------+------------------+--------------------+
|1.5960537239290862|0.3010299956639812|[1.02264150943396...|
|1.6180231930692683|0.6989700043360189|[1.75698113207547...|
|1.8016432272133471|0.9030899869919435|[7.89000000000000...|
| 1.809490720480118|1.0413926851582251|[8.15188679245283...|
| 1.809490720480118|1.0791812460476249|[8.15188679245283...|
|  2.06753330087492|2.2013971243204513|[15.2584905660377...|
|2.1398858622887804|  2.74350976472843|[16.3420754716981...|
+------------------+------------------+--------------------+
only showing top 7 rows

United States R Squared (R2) on test data = 0.9049541697894158
United States Root Mean Squared Error (RMSE) on test data = 0.551224
```

References:

Cheng, C., Barceló, J., Hartnett, A.S. *et al.* COVID-19 Government Response Event Dataset (CoronaNet v.1.0). *Nat Hum Behav* 4, 756–768 (2020).

Estimating mortality from COVID-19. (2020). Retrieved September 23, 2020, from https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19

England Journal of Medicine, 382(18), 1677–1679. https://doi.org/10.1056/NEJMp2003762

Gates, B. (2020). Responding to Covid-19—A Once-in-a-Century Pandemic? New

Soistmann R.,Trigonoplos P.,*How 9 countries responded to Covid-19—And what we can learn to prepare for the second wave*. (n.d.). Retrieved September 23, 2020, from http://www.advisory.com/research/global-forum-for-health-care-innovators/the-forum/2020/05/covid-19-covid-19-responses

Hale, Thomas, Noam Angrist, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster. “Variation in Government Responses to COVID-19” Version 6.0. Blavatnik School of Government Working Paper. May 25, 2020. Available: www.bsg.ox.ac.uk/covidtracker

Hale, Thomas, Noam Angrist, Emily Cameron-Blake, Laura Hallas, Beatriz Kira, Saptarshi Majumdar, Anna Petherick, Toby Phillips, Helen Tatlow, Samuel Webster (2020). *Oxford COVID-19 Government Response Tracker*, Blavatnik School of Government.

John Hopkins GitHub repository (2020). *CSSE_COVID-19_Daily_Report* [data file and codebook].doi:https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_daily_reports/09-25-2020.csv

Kickbusch, I., Leung, G. M., Bhutta, Z. A., Matsoso, M. P., Ihekweazu, C., & Abbasi, K. (2020). Covid-19: how a virus is turning the world upside down. https://doi-org.libproxy.uncg.edu/10.1038/s41562-020-0909-7

W. (n.d.). Estimating mortality from COVID-19. Retrieved October 23, 2020, from https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19

Yong, E. (2020, July 22). America Should Prepare for a Double Pandemic. Retrieved November 04, 2020, from https://www.theatlantic.com/health/archive/2020/07/double-pandemic-covid-flu/614152/