# Exploring COVID-19 Policies and its Trends Throughout the Pandemic

Kelly Lei
Shin Naganuma
Christy Hui

University of California - Los Angeles

## 1. Abstract

The objective of this project is to determine if there are COVID-related policy trends that specific regions follow. This project also seeks to discover if there is an association between COVID policies and COVID cases by seeing if policies are effective in predicting COVID-19 cases. Using k-means with dynamic time warping, we detected 8 distinct clusters based on the strictness of all government policies. The regions that best represent these clusters are: Taiwan, Croatia, Afghanistan, Iran, New Zealand, Macao, Estonia, and China. We applied Poisson and negative binomial regression techniques on each of these 8 representative regions to see if policies can be used to predict COVID-19 cases effectively. These regression models used 16 policy indicators and 1 time variable (numbers that indicate the order of a date) as predictors. In the discrete regression approach, only 3 models out of the total 8 were deemed a good fit by the chi-square test. Thus, we have reason to believe that these two regression techniques (using policy indicators) are not good techniques for predicting COVID-19. This is also evidence for there being no strong association between COVID cases and COVID policies. We also performed random forest on 14 regions. Random forest was able to reliably predict most of the 14 regions' trends; furthermore, using variance importance plots, we found the top ten most important policies in predicting COVID.

Key words: COVID-19 policies, k-means, regression, random forest

## 2. Introduction

With an eventful last two years, it is important to look back at the COVID-19 pandemic and assess what worked and what did not work in trying to control the pandemic. One aspect to consider exploring is regional policies: whereas some regions are endlessly praised for their reaction to the pandemic and initiated lockdown relatively quickly, others are ridiculed for rejecting the notion and trying to continue as if everything was normal. A region's reaction to COVID-19 is directly related to its region-wide politics, and it would be natural to think that region-wide politics greatly affects how wild a pandemic ravages a region. However, is this truly the case?

The purpose of this report is to examine and find trends in policies throughout the world during the COVID-19 pandemic. To be more specific, we are interested in whether or not regions share certain patterns in the strictness of their policies. Can regions be grouped by their policies? Furthermore, we are interested in whether or not there is an association between policies and

COVID-cases. We will do this by seeing how effective COVID policies are in predicting COVID cases. Which policies did the best? Which did the worst?

We expect to find that while some regions will be more strict than others, multiple regions will have similar policy trends. Thus, we expect multiple regions to be able to be categorized by their policy trends. Furthermore, we hypothesize that there is a strong association between COVID cases and policies and that a region's policies contribute greatly in predicting COVID-19 cases. However, we also expect specific policies to perform better than others, and it would be of no surprise to us if we found that particular policies had little to no effect at all in mitigating the spread of COVID-19.

We infer this hypothesis due to the results of multiple research teams conducting similar investigations. For example, in a study titled "The effect of large-scale anti-contagion policies on the COVID-19 pandemic," Soloman Hsiang and his colleagues concluded that "anti-contagion policies have significantly and substantially slowed this [COVID-19 infections] growth" (Hsiang et al. 2020). Furthermore, they estimated that across China, South Korea, Italy, Iran, France, and the United States, region-wide interventions against COVID "prevented or delayed on the order of 61 million confirmed cases, corresponding to averting approximately 495 million total infections" (Hsiang et al. 2020). In another study titled "State variation in effects of state social distancing policies on COVID-19 cases," Brystana G. Kaufman and her colleagues found that social distancing policies were "associated with a 15.4% daily reduction... in COVID-19 cases" (Kaufman et al. 2021). They also found that social distancing policy prevented "nearly 33 million cases nationwide" (Kaufman et al. 2021) after 3 weeks. Ali Hadianfar and his colleagues, in their study titled "Effects of government policies and the Nowruz holidays on confirmed COVID-19 cases in Iran: An intervention time series analysis," concluded that the "Nowruz holidays and the implementation of social distancing measures in Iran were related to a significant increase and decrease in COVID-19 cases, respectively" (Hadianfar et al. 2021). However, Ali Hadianfar and his colleagues also found despite the "closing of universities and schools, no statistically significant change was found in the number of new confirmed cases" (Hadianfar et al. 2021). More evidence that some policies may not work well comes from Christopher R. Berry and his colleagues' "Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic," where they concluded that "shelter-in-place orders had no detectable health benefits" (Berry et al. 2021) because they did not "find detectable effects of these [shelter-in-place] policies on disease spread or deaths” (Berry et al. 2021). Thus, we hypothesize that although policies do have a great effect on the COVID-19 aftermath, certain policies will do better than others and some may not have an effect in predicting COVID-19 cases at all.

## 3. Data

### 3.1 - Data Collection

The data for this project was collected from two sources: the Oxford University's COVID Policy Tracker and Our World in Data's COVID-19 Dataset. The COVID Policy Tracker is a data hub containing multiple data sets from January 2020 to the present, each separated by a policy. In each policy data set, there exists a daily time series for around 180 regions. Every data set has a consistent format: for each region available, a policy level score is assigned for every day. This score represents the strictness of that specific policy for that day in that region and is calculated by the Oxford COVID-19 Government Response Tracker, which "tracks individual policy measures across 20 indicators" (Hale et al. 2021). The indices are simple "averages of the individual component indicators" (Hale et al. 2021) and use "ordinal indicators where policies are ranked on a simple numeric scale" (Hale et al. 2021).

Note that there are different types of policies in the data set: individual policies, flag variables, and summary policies. We define the 16 individual policies as government indices that contain information solely on the level of strictness. Examples of individual policies include policies representing stay at home requirements, restrictions on gathering, and facial coverings. Flag variables accompany individual variables and represent additional information corresponding to the type of policy it accompanies. These flag variables exist for some individual policy indicators but do not exist for others. An example of a flag variable is “c1_flag,” which represents whether or not the policy is geographically targeted and accompanies the “c1_school_closing” individual policy. Summary policy indicators are the four indices that aggregate certain types of individual policies into a single number. Examples of “summary” policies are the overall government response index (which takes into account all individual policy indicators) and the containment and health index (which takes into account all containment/closure and health system policy indicators).

Our World in Data’s COVID-19 Dataset is a data hub containing information on more than 200 regions for some metrics and as few as 47 regions for another. The time variate data sets are separated by the statistics it represents, such as the number of vaccinations, confirmed cases and deaths, and ICU data. Each data set includes information such as the date, the region, and the corresponding statistic of that day and region.

We combined the data into one data set that contains information for a total of 179 total regions. Each row represents a region and a date (i.e. one row represents Aruba on January 1, 2020, another row represents Japan on January 2, 2020, etc.). Each column represents a policy (i.e. school closing policy) or a COVID-related statistic (i.e. vaccinations). Note that each "policy" column of the data is normalized, which was not the case in the original data by Oxford University. This was done by dividing each point by it's column's maximum and then multiplying by 100, and it was normalized under the assumption that modeling/clustering would yield better results. However, it should be acknowledged that normalizing the data may not have an effect at all in aiding the techniques due to the categorical and static nature of the policy data.

**3.2 - Data Analysis - Missing Values**

In total, there are 121751 rows and 102 columns, with almost 30% of this data (29.45466%) consisting of missing values. Interestingly, 25.96512% of these missing values come from the Our World in Data's COVID-19 Dataset. The remaining amount of missing values (3.489539%) come from the Policy Tracker Data. It should be noted that the nature of these missing values is not missing at random. Values that were missing were only missing at the start of the time period. Once data started appearing after one day, there would be no more missing values. For example, data for Algeria's "c1_flag" column would only start appearing after March 12th, 2020. From that date onwards, missing values in that column for Algeria would not appear. This makes us believe that data is missing either because no data was being tracked at the given time (like a policy value) or the data point did not exist at that time (like vaccinations before December 2020). These missing values prove a problem in modeling and data analysis since they are not completely at random and cannot be imputed. To simplify computations and analyses, missing values were either ignored or deleted. However, it should be acknowledged that this may not be the greatest solution.

**3.3 - Data Analysis - A Closer Look at School Policies**

One of the best ways to understand the data is to look at one aspect or detail and examine it closely. To get a closer look at how policies work, we can look at one policy of interest: school policies. In prior research, Hafiandar and his colleagues claimed that school policies in the

Middle East had no statistically significant effect on COVID cases (Hafiandar et al. 2021). Is this the same case in global data? What does our data set say about school policies?

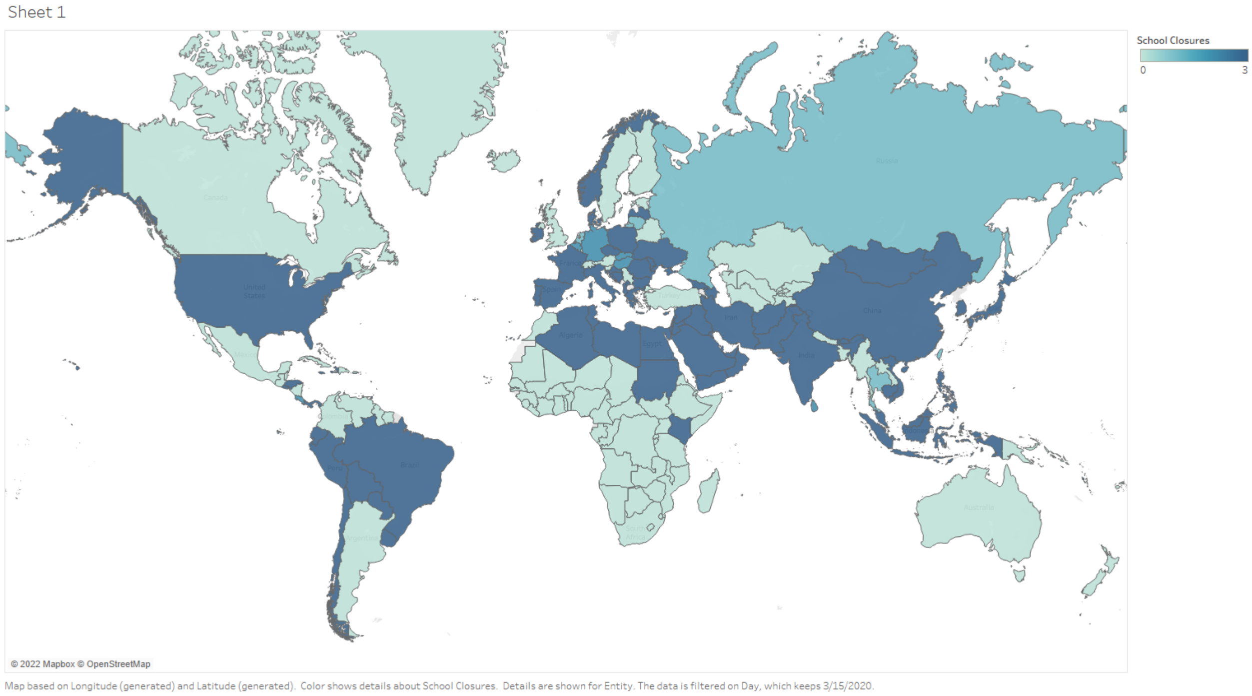



*Figure 1*

*Figure 1* shows a map of the world, colored by school policy on March 5th, 2020, a few days before the WHO declares COVID-19 a pandemic and former United States President Donald Trump declares COVID-19 to be a national emergency (AJMC 2021). One interesting thing to note is the relationship between policies and perception. Australia, for example, had lax policies in March 2020, despite the region's positive perception for its COVID-19 response and low cases (Haseltine 2021) at the time. On the other hand, the United States enforced strict school policies relatively quickly, which contrasts the overall negative view on how the region reacted to the pandemic (Wallace-Wells 2021). One other interesting thing to note is that areas of the world with densely populated areas enacted strict school policies (such as the United States, China, and South Korea) swiftly. This story, however, does not last.

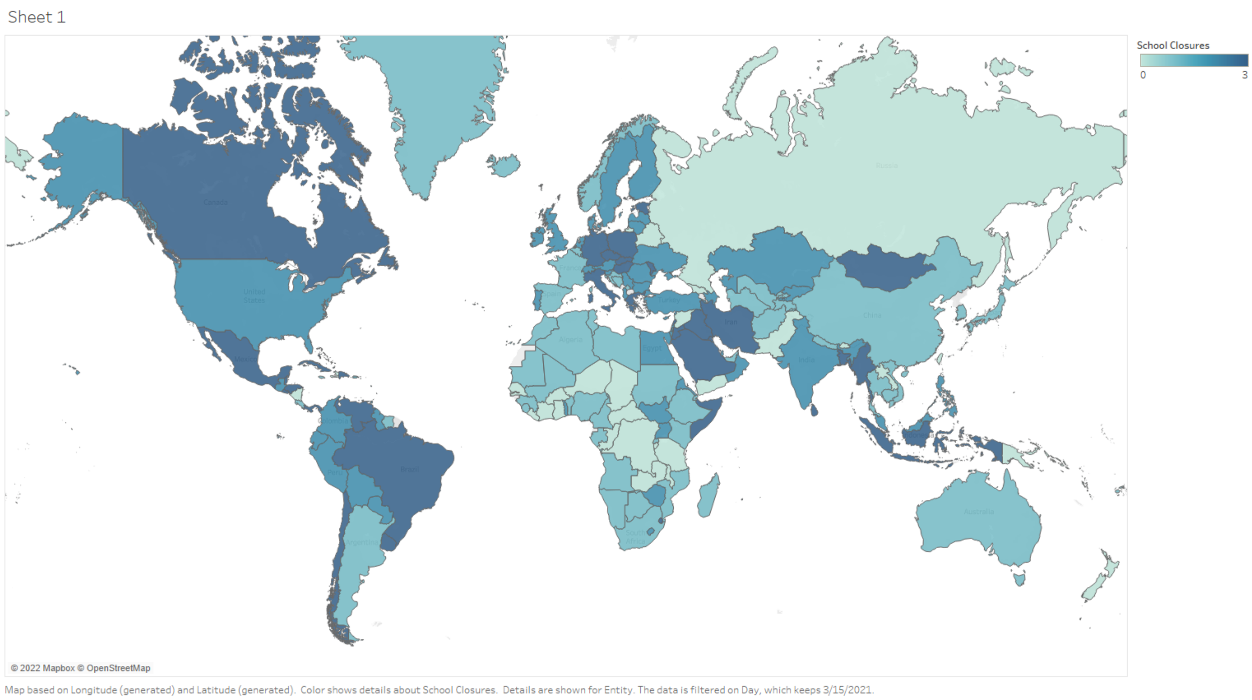



*Figure 2*

*Figure 2* shows the school policy level throughout the world just like the previous plot; however, this one is on March 15, 2021, exactly one year after the previous plot. Notice how in just a year, the aforementioned regions relax their policies while the other regions become more strict. This may be for a variety of reasons. After a year of lockdown, regions that have experienced it for longer may begin relaxing their policies due to societal pressure. Or due to strict policies, cases may have begun to decline; thus, regions that acted quickly did not have to lockdown as strictly a year later while regions that did not act had to act a year later. Nevertheless, there is a clear difference in attitude towards school policy throughout the world between March 2020 and March 2021.

While it is interesting to note differences between all regions in regards to school policy, it may be more useful to look at only a few. *Figure 3* shows a series of line plots.

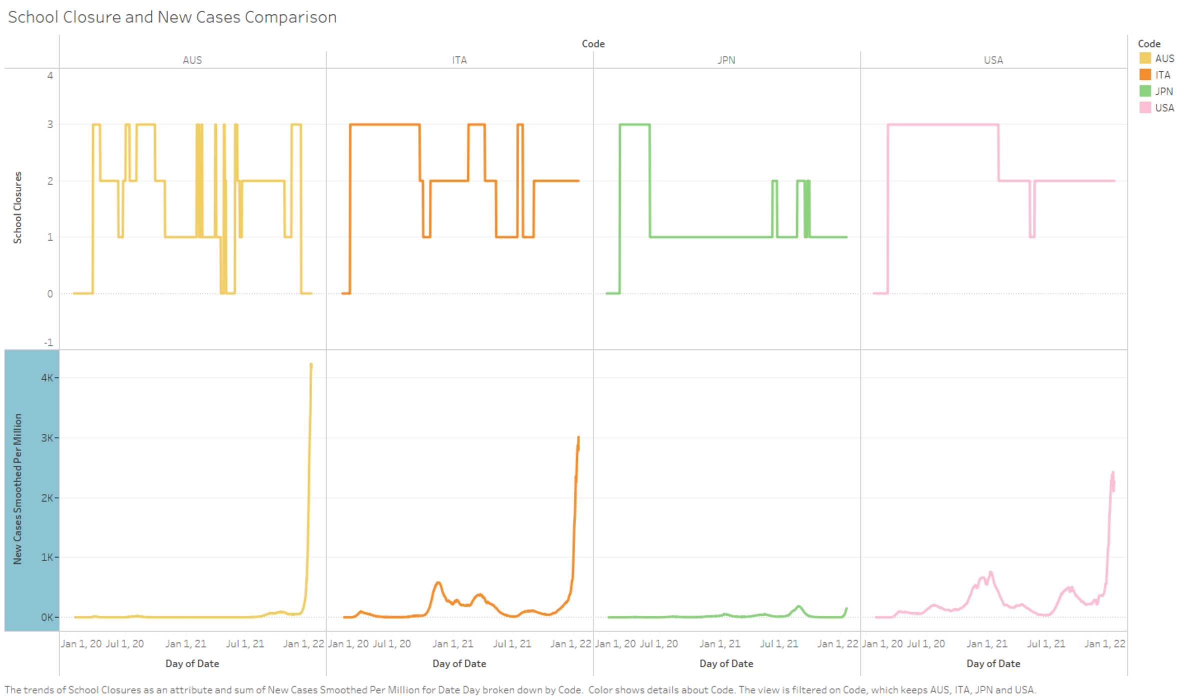



*Figure 3*

The first set of plots shows daily school closure policy levels from the end of 2020 to the start of 2022, separated by random regions (in this case, the four regions are Australia, Italy, Japan, and the United States.). The second set of plots is the corresponding new case counts. Comparing the series of plots, it is clear that school policy differs wildly between each region. Throughout two years, Australia seems to struggle with finding the right balance between closing and reopening schools. Italy fluctuates a lot in this area as well (especially in 2021), although they fluctuate less than Australia, cautioning on the side of being more strict. However, despite Italy being stricter than Australia, Australia's cases are few in-between compared to Italy between 2020 and most of 2021 (as shown in Italy's cases peaking at the end of 2020 and the start of 2021).

Interestingly, Italy's COVID situation is similar to the US. COVID cases seem to rise and fall in similar patterns, with the United States being a little more pronounced in each peak. However, the United States' school closure policies are entirely different from Italy's. When Italy decides to relax its school closure policies, it takes the US a few months to follow suit, if at all.

These two patterns (Australia vs. Italy and Italy vs. the US) may imply that school closure policies do not have a large effect on COVID cases (which supports Hafiandar and his colleagues' research). If there was strong correlation, these line plots would reflect the conclusion in multiple regions and we would see COVID cases drop as school policy becomes more strict. The implication that school policies do not have a large effect on COVID cases is further supported by the fact that Japan does not have a history of strict school closure policy (with them rarely escalating to a level of 3), but consistently has a low level of COVID cases.

However, we must be skeptical. There could be several different reasons that could explain this fluctuation (such as different policies that have a greater effect on controlling COVID). Also, we should note the possibility that school closure policies may be a response to COVID cases, rather than COVID cases being controlled by school closure policies. This would indicate that strict school closure policies do not cause lower cases, but the other way around. Nonetheless, there is a clear conclusion to be made: while Hafiandar may be correct (and our data shows information that can help prove it), there are many confounding variables, and concluding anything without more proper research would be disingenuous to the study.

**3.4 - Data Analysis - Correlation between Policies**

While looking at one policy can be insightful, the goal of this report is to study a multitude of policies for many regions in the lens of COVID-related statistics. To better understand this project, it can be helpful to examine the correlation between each policy.

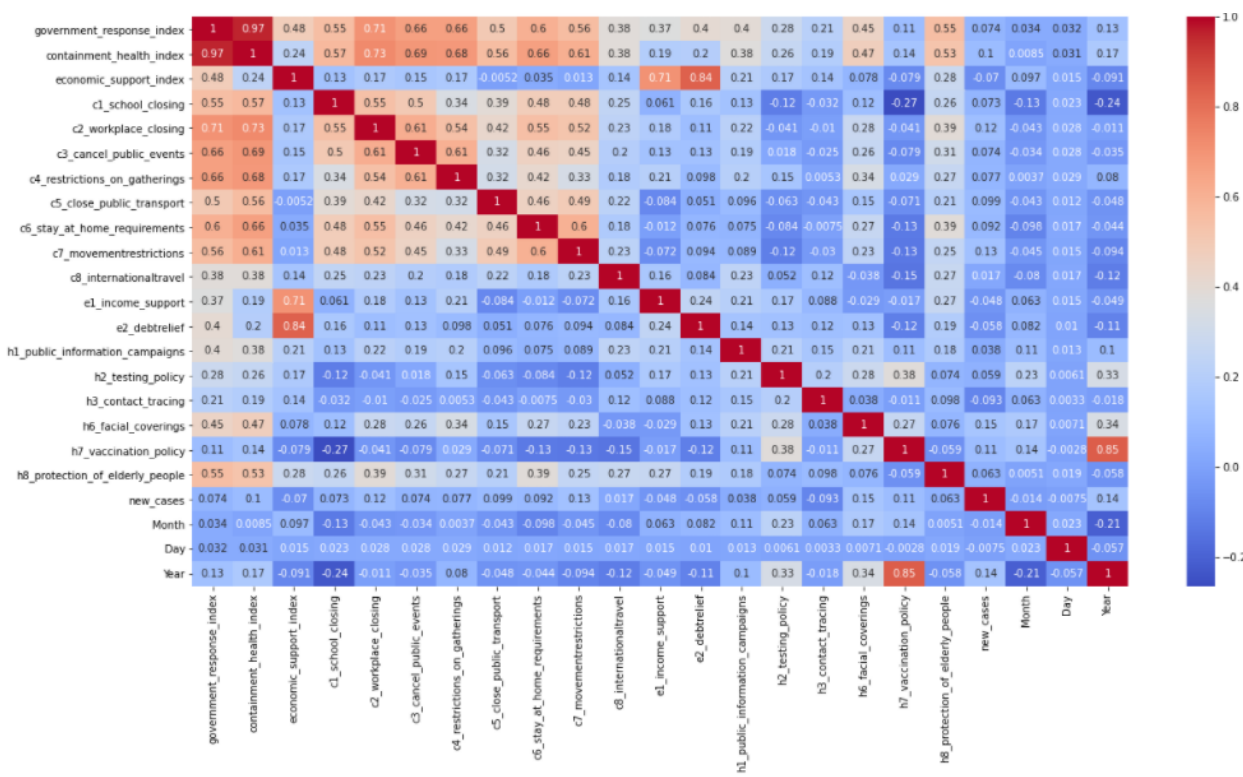



*Figure 4*

*Figure 4* is a correlation matrix between each policy. One interesting thing to note is the extremely high positive correlation (0.97) between the two summary policy statistics of government response index (a score that takes into account all four types of policy indicators) and containment health index (a score that takes into account two types of policy indicators: containment/closure policies and health system policies). This indicates that the government response index highly takes into account these two policies (containment/closure policies and health system policies) when calculated. Some more interesting things to note are the high positive correlations between the economic support index and debt relief and between vaccination policy and year. The former correlation makes sense, since the easiest way for a region to provide economic support is to provide debt relief. The latter correlation also makes sense, since vaccinations were unavailable in 2020, there was no vaccine policy to enforce. Once vaccinations became available in early 2021 (AJMC 2022), vaccination policies were finally able to be carried out.

**3.5 - Data Analysis - K-means Clustering**

We also study patterns of policies using unsupervised learning methods. By exploring how algorithms group data points together, we can discover similar traits and perhaps even find the best policies that can predict COVID-19, given that an association exists between COVID cases and COVID policies. One method to achieve this is by clustering. In particular, we implement the k-means algorithm, which constructs clusters of data by “splitting samples into k groups and minimizing the sum-of-squares in each cluster” (Amidon 2020) with a distance metric (usually Euclidean distance). However, one of the biggest problems this method faces is its difficulty in shaping time series data, since it “is invariant to time shifts, ignoring the time dimension of the data” (Amidon 2020). This problem is solved by introducing dynamic time warping, “a technique to measure similarity between two temporal sequences that do not align exactly in time, speed, or length” (Amidon 2020). To summarize, dynamic time warping is an optimized distance solution that takes "the square root of the sum of squared distances between each element in X and its nearest point in Y" (Amidon 2020) and shapes an average centroid, “regardless of where temporal shifts occur amongst the members” (Amidon 2020). This allows for a more accurate k-means approach with time-varying observations. For this project, k-means with dynamic time warping is extremely powerful, and we gain a better understanding of the data as a result.

Since this data set has many variables, it can be difficult to determine which policies to cluster based off of. Fortunately, Oxford University's COVID Policy Tracker has the 4 summary indices: the government response index (which takes into account every policy index), the containment and health index (which takes into account every containment/closure and health system policy index), the stringency index (which takes into account every containment/closure policy and the public information campaign policy index), and the economic support index (which takes into account every economic policy index). These four summary indices were used as response indicators for a total of 4 k-means models. Furthermore, note that k-means is incapable of modeling when there are missing values. Only 77 countries out of the total 179 did not have missing values for the "government_response_index" variable; thus only those 77 countries were used in the clustering technique.

To determine the optimal number of clusters, we can find the number of clusters with the minimum sum of squares error (SSE). Below is the line plot showing the SSE multiple k-means models, each with a different number of clusters.

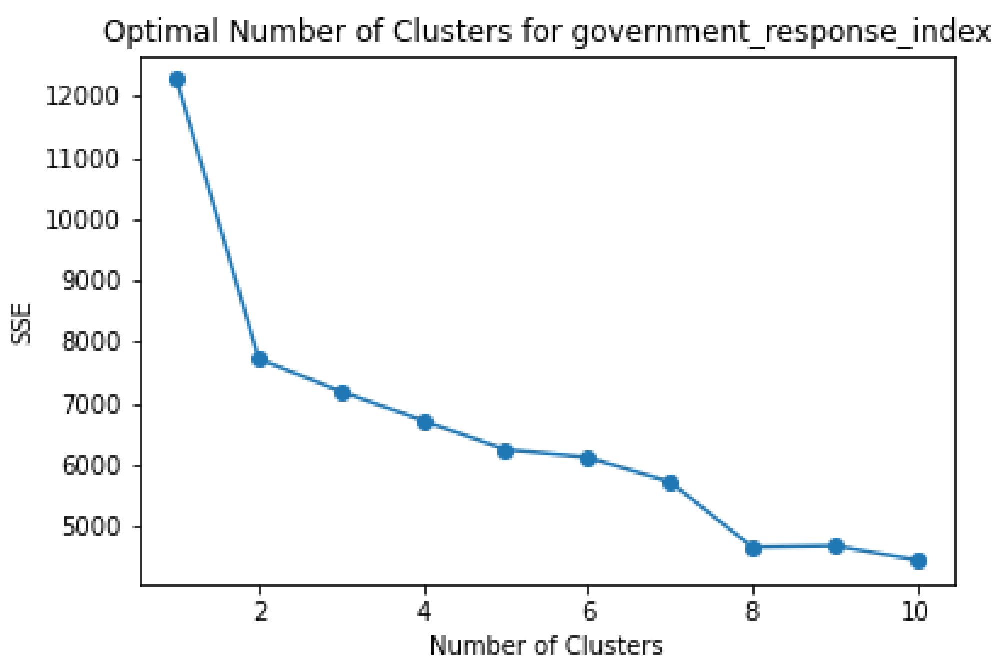



*Figure 5*

Notice that 10 clusters seem to yield the smallest SSE. However, this is an instance of overfitting. Thus, we used the elbow method to determine that 8 clusters was the most optimal.

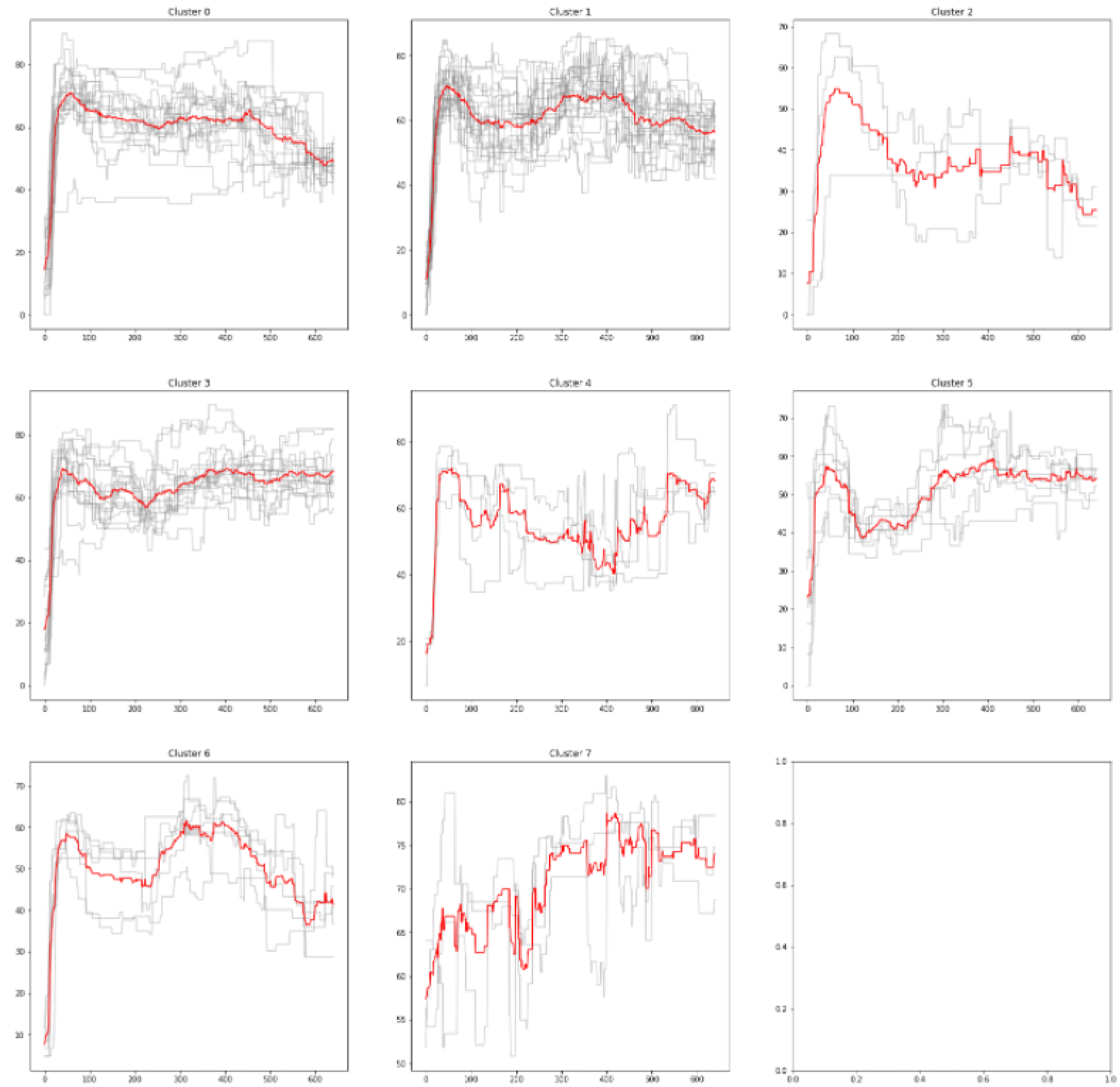



*Figure 6*

*Figure 6* shows each cluster and its trends in relation to the government response index. The y-axis represents the government response index score and the x-axis represents the number of days. Notice how each cluster seems very distinct and tell different stories. “Cluster 0” seems to represent regions that were consistent in their policies, but began to slowly relax their policies after one year of COVID, possibly due to low cases. “Cluster 1” seems to represent regions that have a similar start as “Cluster 0.” However, the regions in “Cluster 1” seem to be more inconsistent, relaxing their policies and then having to stricten them again in the later parts of the pandemic. This contrasts the regions in “Cluster 2,” which seem to have a consistent downward trend of strict policies. This cluster may represent regions with governments that may have trouble in responding to change during the pandemic (and thus simply relax their policies) or an extremely poor perspective on the pandemic. “Cluster 3” seems to be extremely similar to “Cluster 1,” but instead of relaxing their policies towards the end of 2021, “Cluster 3” seems to make their policies stricter. “Cluster 4” seems to represent regions that are extremely erratic in their policies, possibly due to weak leadership or governments that have polarizing views in regards to the pandemic. “Cluster 5” seems to represent regions that had strict policies at the start of the pandemic, but were forced to make their policies more strict as the pandemic continued. This pattern is seen in “Cluster 6” as well, albeit to a lesser degree. This may mean that “Cluster 5” and “Cluster 6” represent similar regions. The final cluster, “Cluster 7” seems to represent regions with extremely erratic approaches to COVD-19 policy, but slowly get more stringent as the pandemic continues.

Examining trends of the line plots is extremely helpful in answering whether or not there are distinct patterns of COVID-19 policy enacted by certain regions. However, it also is helpful to look at the regions represented by each cluster to get a sense of actual regions being represented by each cluster. By finding the region with the minimum dynamic time warping distance relative to the average dynamic time-series sequence, we can find the representative regions for each cluster.

Taiwan, Croatia, Afghanistan, Iran, New Zealand, Macao, Estonia, and China are the representative regions of each cluster, respectively. Interestingly, these clusters and their representative regions are consistent with the region’s public perception of its COVID response. For instance, Taiwan, the representative region of “Cluster 0,” is often praised for its early and consistent response to the pandemic (Cheng 2021). “Cluster 0” represents this story: with few oscillations and a near straight line for 400 days, it is apparent that k-means did an exceptional job in characterizing regions that have a consistent and quality COVID-19 response. On the other hand, according to the model, Afghanistan best represents “Cluster 2,” a cluster that seems to constitute regions that have a poor/relaxed response to the pandemic. Again, this is consistent with public perception: Afghanistan faced many challenges in controlling the pandemic in 2020

(Lucero-Prisno 2020). Due to these observations, it is clear that clustering was extremely effective in categorizing the different types of policies.

*Table 1* describes all models which were performed, and were analyzed the same way as the government response index.

| **Policy Indicator** | **Optimal Clusters** | **Representative Regions** |
|---|---|---|
| government response index (all indicators) | 8 | Taiwan, Croatia, Afghanistan, Iran, New Zealand, Macao, Estonia, China |
| containment and health index (all C and H indicators) | 3 | Belarus, Taiwan, Cambodia |
| stringency index (all C indicators, plus H1) | 3 | Macao, Cambodia, Taiwan |
| economic support index (all E indicators) | 5 | Russia, Belarus, Senegal, Lithuania, Nigeria |

*Table 1*

All of these models and "optimal" regions were observed the same way as the government response index. Notice how the same regions (Taiwan, Macao, Belarus) were observed multiple times, which may indicate that these regions have very distinct patterns in how they approached the COVID-19 pandemic. Furthermore, note that a majority of these regions are either Asian or European countries.

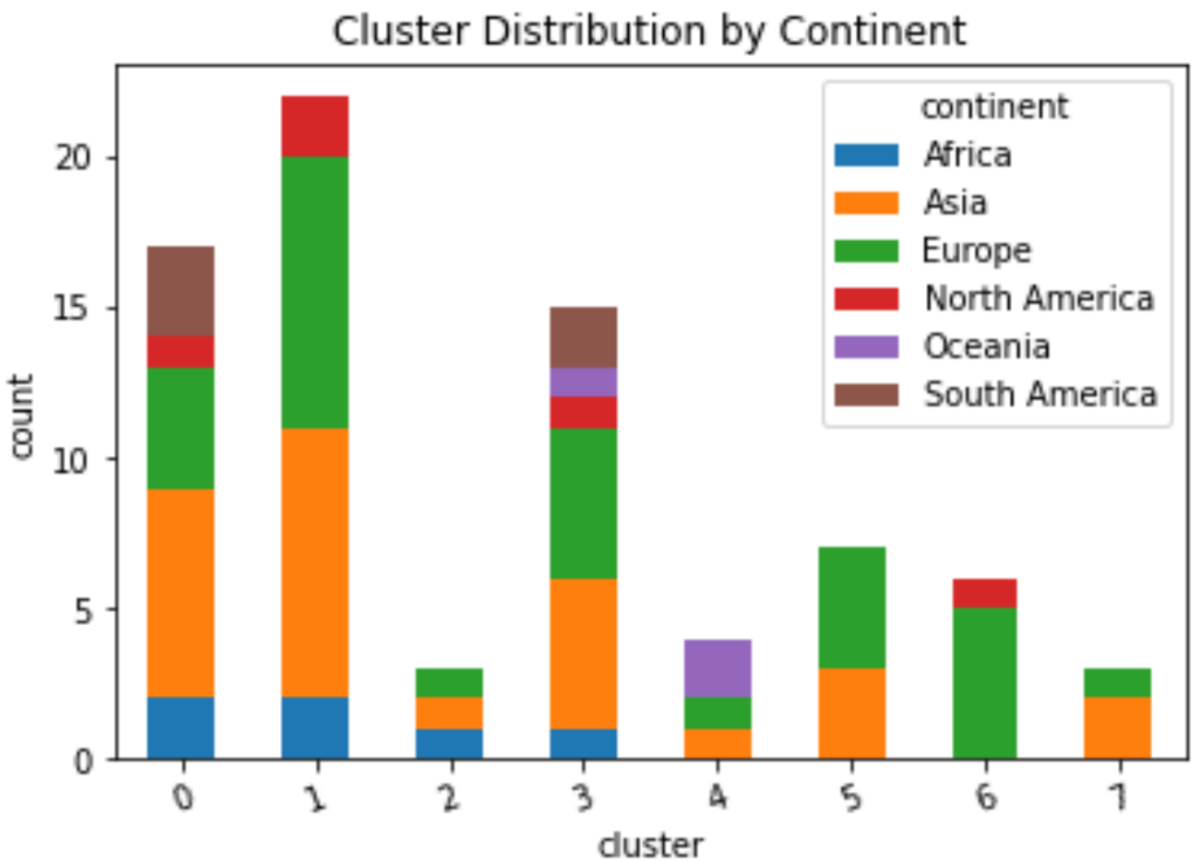



*Figure 7*

Figure 7 shows the cluster distribution by continents. As shown in the distribution, due to the sheer amount of Asian and European countries represented in the data set and in each cluster, it would be natural for a majority of the representative regions to be Asian or European. To be exact, there are a total of 30 European regions and 28 Asian regions, which encompasses around 75% of all the regions within the data set. It is also interesting to point out the variance of continents in which each cluster contains. For example, “Cluster 0” contains regions from 5 continents: Africa, Asia, Europe, North America, and South America. “Cluster 1” contains regions from 4 continents: Africa, Asia, Europe, and North America. “Cluster 3” contains regions from all 6 continents. Furthermore, there does not exist a cluster that contains a region from only one continent. This may imply that COVID policies are varied, even within continents. Despite countries being close to each other, it does not seem like their policies would be similar. Geography does not seem to play a role in COVID policy.

Due to this analysis, we have reason to believe that certain regions can be distinctly classified by their policies (based on overall government response index). The regions, separated by policies using k-means, were distinct, not only through trends seen by the plots, but also in real life (as seen when comparing real-life perception). Further studies can be done on each region, and further studies must be done on more specific policies. Furthermore, it was noted that each cluster was not similar in geography; further studies can be done to see if these regions have a cultural or economic commonality. Nonetheless, it is clear that k-means clustering was able to distinctly capture policy trends based on overall government policy.

## 4. Analysis Methods - Regression Techniques

### 4.1 Method Introduction

With a deeper understanding of the data set, we can now seek to discover whether or not an association exists between COVID-cases and COVID policies. By running modeling techniques and finding whether or not policy indicators are the best predictors for predicting COVID cases, we can determine if a strong association exists. Furthermore, we can also determine which policies are most effective in predicting cases (given that policies can predict COVID cases). This information can be helpful in getting a better understanding of the policies and, perhaps, allow more specific research to be conducted, which could be particularly helpful for future pandemics. However, note that conclusions based on these techniques does not mean that the policies are best suited in controlling COVID-19 or other pandemics. This would be a premature conclusion, since data is not always an accurate representation of the real world.

Before modeling, it must be noted that the blanket statistic of “COVID cases” is not an accurate measure of the pandemic. In some regions, testing kits may not be as readily available, causing cases to be not representative of how rampant COVID was in the region. Furthermore, there is the case that COVID cases are seen less on the weekends, due to testing sites not being available at that time or people simply not going as much. Thus, for the response variable in modeling techniques, we use “new cases smoothed,” which removes noise in data to better represent the time-series data.

Note that we will only use the 16 “individual” policies. Since flag policies are only additional information (thus are not needed) and only exist to overcomplicate the project, we decided to not leave them in as predictors. Furthermore, since summary policies are not as specific and will only add complicated correlation to models (due to being based off of the "individual" policies), we also decided against using these policies as predictors. This leaves for a total of 16 "individual" policies, which we use solely for predicting COVID cases. We add 3 more variables (year, month, and day) in order to accommodate for time in the regression models.

For modeling, we decided against using only summary policies. Note that summary policies are summary scores of individual scores that are already a summary of strictness throughout a whole country. If we were to model using only individual policies, we could get a more specific understanding of which real-world, applicable policies (rather than a summary score of a summary score) have a strong association between policies and cases. Modeling with only summary policies will lead to more generalized results, and although that can be helpful, individual policies yield richer and more specific results.

Lastly, due to the scope of the data, we chose to run regression techniques on only the representative regions of each cluster in the government response index separately. In other words, regression was only applied to Taiwan, Croatia, Afghanistan, Iran, New Zealand, Macao, Estonia, and China separately. A different approach where we run models on all countries (179 of them) would complicate the problem immensely, since each country would need a dummy variable. Furthermore, since we are curious on which policies are strongest in predicting COVID cases, keeping the country variable constant would make the process more interpretable and easier to execute.

The first types of modeling techniques to consider are regression techniques. By examining the contribution of each policy, comparing the standardized coefficients, and training and testing data, we can infer which policies are more effective than others in predicting COVID-19 cases (smoothed). The approach to finding significant predictors was as follows: (1) the first regression technique used was Poisson Regression; (2) in the case that any assumptions fail, we considered Negative Binomial Regression; (3) every time we fit a model, a chi-square test was performed to determine the model's goodness of fit.

Note that the time period for the data being modeled is between 2020-03-01 and 2021-12-01. This date range was arbitrarily chosen.

**4.1.1 - Poisson Regression Explained**

Since Covid data is count-based and discrete, we first consider Poisson regression, which models the logarithm of the expected value. The equation of a Poisson regression can be written as

$$log(\widehat{Y}) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \quad \text{(Eq. 1)}$$

There are four assumptions:

1. The response variable Y is count per unit of time
2. The observations are independent of each other
3. $E(Y) = Var(Y) = \mu$
4. Linearity: $log(Y)$ is proportional to X

**4.1.2 - Negative Binomial Regression (NB2) Explained**

a. **General Model**

In order to account for the possibility of failing the third assumption, we can also consider a more flexible regression model: the negative binomial regression (NB2). Rather than assuming the mean and variance are equal, this model assumes that the mean and variance has the following relationship:

$$Var(Y) = E(Y) + \alpha \times E(Y)^p \quad \text{(Eq. 2)}$$

As shown in Eq.2, $Var(Y)$ is always greater than $E(Y)$ for a positive value of α, a constant that indicates the inflation of "among-unit variance". As α approaches 0, the model will be reduced to a Poisson distribution. When the result of a chi-square test implies that the Poisson model is not a good for for the data, we will use $p = 2$ to build an NB2 model that will, hopefully, be effective in predicting COVID cases.

b. **Finding a value for α**

To find a value for α in Eq. 2 after fitting a Poisson regression, we need to fit an Auxiliary OLS regression, which can be expressed by the following equation:

$$\frac{(y_i - E(Y_i))^2 - E(Y_i)}{E(Y_i)} = \alpha \times E(Y_i) + 0 \quad \text{(Eq. 3)}$$

In Eq.3, the $y_i$ is the ith observation of the original response variable, and $E(Y_i)$ is the expected value of the response variable given by a Poisson regression.

**4.1.3 - Chi-Square Test**

We will perform a Chi-Square test each time we fit a model to check the goodness of fit and determine the next steps needed to improve the model. The hypotheses for the test are:

$H_0$: There is no significant difference between the observed and the expected value.

$H_a$: There is a significant difference between the observed and the expected value.

## 4.2 - Results for Regression Models

### 4.2.1 - Regression for All Countries Data with Cluster and Time

***Poisson Regression***

After obtaining the clustering result based on "government_response_index", we decided to use the clustering result and time as explanatory variables to see if cluster outperforms time in fitting COVID new cases. We used data between 2020-03-01 and 2021-12-01, and we set the data during the first 512 days as training set and the rest 129 days as testing set.

By fitting a Poisson regression, the summary shows that almost all 9 predictors have extremely small and significant p-values ($\leq 0.000$). The deviance for the model is $5.07 \times 10^{8}$ and the Pearson chi-square value is $1.59 \times 10^{9}$, and we then used a chi-square test to check the goodness of fit:

| Chi-Square Value | Degree of Freedom | P-Value ($\alpha = 0.05$) | Conclusion |
|---|---|---|---|
| $1.59 \times 10^{9}$ | 37985 | 0 | Reject the Null |

*Table 2*

At $\alpha = 0.05$, a p-value of 0 makes us reject the null and conclude that there is a significant difference between the observed values and the fitted values. In other words, the Poisson regression is not an accurate model for data on all countries. Upon closer inspection, the third assumption ($E(Y) = Var(Y) = \mu$) is violated. Due to this result, we try negative binomial regression.

***Negative Binomial Regression***

By fitting the auxiliary OLS regression model on the data, we can obtain an optimal value for $\alpha$, as seen in (Eq. 3). We also need to perform a t-test to check if the value of $\alpha$ is significant:

| $\alpha$ | T-Value | Degree of Freedom | P-Value ($\alpha = 0.05$) | Conclusion |
|---|---|---|---|---|
| 10.11673 | 22.09771 | 37985 | $\leq 0.0001$ | $\alpha$ is statistically significant |

*Table 3*

Plugging in the significant α value to Eq. 2, we obtain the relationship between the mean and variance of new cases smoothed for China:

$$Var(Y) = E(Y) + 10.11673 \times E(Y)^2 \quad \text{(Eq. 4)}$$

Fitting a negative binomial with $\alpha = 10.11673$, we obtain the following equation:

$$log(\widehat{Y_{all}}) = 0.0045\, day *+ 7.5725\, cluster_0 *+ 7.2231\, cluster_1 *+ 4.7876\, cluster2 *+ 6.6814\, cluster_3 * + 6.8718\, cluster_4 *+ 5.1021\, cluster_5 *+ 6.0304\, cluster_6 *+ 6.6424\, cluster_7 * \quad \text{(Eq. 5)}$$

The negative binomial regression shows that all cluster indicators and time are significant, and our new Pearson chi-square value for the improved model is $30500$.

- Model Validation

| Chi-Square Value | Degree of Freedom | P-Value ($\alpha = 0.05$) | Conclusion |
|---|---|---|---|
| 30500 | 37985 | 1 | Fail to reject the null |

*Table 4*

*Table 6* shows that at $\alpha = 0.05$, a p-value of 1 makes us fail to reject the null hypothesis. Thus, we conclude that the NB2 regression is a valid approach to model all countries' daily new cases smoothed using cluster indicators and time. The RMSE for the testing set is 19070.
However, this model cannot help us distinguish the statistical significance between cluster indicators and time since they all have p-values extremely close to 0.

**4.2.2 - Regression Example: China**

***Poisson Regression and its Problems***

The representative region for Cluster 7, given by the government response index, is China. By fitting a Poisson regression on the 512 days of the training set, the summary shows that almost all policy predictors have extremely small and significant p-values (≤ 0.000). However, the deviance for the model is 8857.6 and the Pearson chi-square value is $1.04 \times 10^4$, and a chi-square test is performed to check the goodness of fit:

| Chi-Square Value | Degree of Freedom | P-Value ($\alpha = 0.05$) | Conclusion |
|---|---|---|---|
| $1.04 \times 10^4$ | 511 | 0 | Reject the Null |

*Table 5*

At $\alpha = 0.05$, a p-value of 0 makes us reject the null and conclude that there is a significant difference between the observed values and the fitted values. In other words, the Poisson regression is not an accurate model for data on China. Upon closer inspection, the third assumption ($E(Y) = Var(Y) = \mu$) is violated. Due to this result, we try NB2 regression.

***Negative Binomial Regression on the Training Set***

By fitting the auxiliary OLS regression model on the data, we can obtain an optimal value for α, as seen in (Eq. 3). We also need to perform a t-test to check if the value of α is significant:

| α | T-Value | Degree of Freedom | P-Value ($\alpha = 0.05$) | Conclusion |
|---|---|---|---|---|
| 0.866765 | 14.175875 | 511 | $\leq 0.0001$ | α is statistically significant |

*Table 6*

Plugging in the significant α value to Eq. 2, we obtain the relationship between the mean and variance of new cases smoothed for China:

$$Var(Y) = E(Y) + 0.866765 \times E(Y)^2 \quad \text{(Eq. 5)}$$

Fitting a negative binomial with α = 0.575456, we obtain the following equation:

$$\begin{aligned} log(\widehat{Y_{China}}) = \; & -0.0001\, day + 0.0005c_1 - 0.0088c_2{*} + 0.0083c_3 + 0.0016c_4 \\ & + 0.0006c_5 + 0.0047c_6 + 0.0013\, c_7 - 0.0303c_8{*} - 0.0026e_1 - 0.0116e_2{*} \\ & + 0.0228h_1{*} - 0.0022h_2 + 0.0228h_3{*} + 0.0168h_6{*} + 0.0052h_7 - 0.0016h_8 \end{aligned} \quad \text{(Eq. 6)}$$

The negative binomial regression helps us exclude 11 statistically insignificant predictors, and our new Pearson chi-square value for the improved model is 339.

- Model Validation

| Chi-Square Value | Degree of Freedom | P-Value ($\alpha = 0.05$) | Conclusion |
|---|---|---|---|
| 339 | 511 | 0.999 | Fail to reject the null |

*Table 7*

*Table 7* shows that at $\alpha = 0.05$, a p-value of 0.999 makes us fail to reject the null hypothesis. Thus, we conclude that the NB2 regression is a valid approach to model China's daily new cases smoothed using policy indicators and time, and the significant predictors are workplace closing, inernational travel, debt relief, public information campaigns, contact tracing, and facial coverings.

***Prediction on the Testing Set***

- Plotting the Results

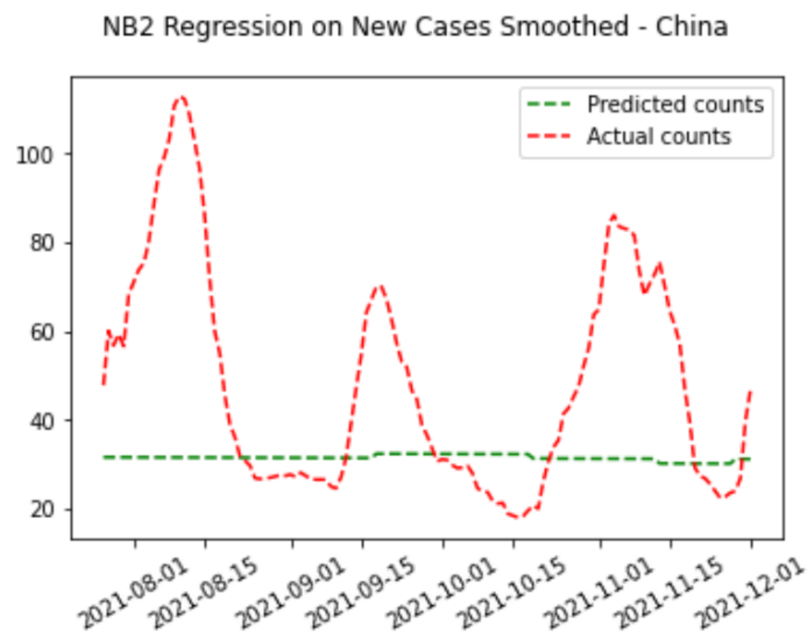



*Figure 9*

As can be observed in *Figure 9*, although the RMSE for the fitted values is 30.42, the fitted values of the NB2 model on China's daily new cases could not capture any fluctuations, and instead captures a smoother pattern.This tells us that policy indicators with numbers indicating dates are not sufficient to explain the monthly trend of COVID new cases in China.

#### 4.2.3 - Summary Results for Discrete Model Approach

| Cluster | Representative Country/Region | Model | α | Number of Significant Predictors | Testing RMSE | Chi-sqr Goodness of Fit | |
|---|---|---|---|---|---|---|---|
| 0 | Taiwan | NB2 | 0.114942 | 10 | 25.414 | Y(0.618) | **Settled for NB2** |
| 1 | Croatia | NB2 | 0.065545 | 14 | 2158.096 | N(0.0) | |
| 2 | Afghanistan | NB2 | 0.072083 | 16 | 10715.269 | N(0.0) | |
| 3 | Iran | | 0.052868 | | | | **Neither of the two worked: alpha was not significant** |
| 4 | New Zealand | NB2 | 0.285514 | 11 | 38.773 | Y(0.999) | **Settled for NB2** |
| 5 | Macao | Poisson | N/A | **0** | 1.536 | Y(1) | **Settled for Poisson but no significant predictors** |
| 6 | Estonia | NB2 | 0.151947 | 13 | 581.524 | N(0.0) | |
| 7 | China | NB2 | 0.866765 | 6 | 30.42 | Y(0.999) | **Settled for NB2** |

*Table 8*

*Table 8* shows the results of using regression on the 8 representative regions chosen by k-means clustering based on the government response index. To reiterate, a Poisson regression model was fit on each region with all 16 "independent" policies and a variable "day" indicating the order of date. If any assumptions failed, a negative binomial regression would be used. Each time a regression model was used, a chi-square test would be applied to test the goodness of fit. Notice how each region has a distinct number of significant predictors (i.e. Taiwan has 10, China has 6, and so on). This implies that different regions (and possibly different clusters) have unique significant policies that can be used to predict COVID cases smoothed. Furthermore, note that Macao is the only region where Poisson regression can be used effectively (according to chi-square). Poisson regression was deemed inappropriate for the other regions because the third assumption, $E(Y) = Var(Y) = \mu$, was not satisfied with each. However, note that the negative binomial model only improved the deviance and chi-square value for 4 regions: Taiwan, New

Zealand, Macao, and China. For the other 4 regions, the Poisson nor negative binomial regression technique was able to fit for COVID cases (according to the chi-square test). Notice that none of the predictors are significant in the NB2 model for Macao, so we have reason to believe that the policy indicators and the order of the date cannot be used to predict COVID cases.

To interpret the coefficients for the regions that did work (Taiwan, New Zealand, and China), we say "for a one unit change in the predictor variable, the difference in the logs of expected counts of the response variable is expected to change by the respective regression coefficient, given the other predictor variables in the model are held constant" (UCLA SCG). For instance, the estimated coefficient for $c_3$ (cancel public events) in China's regression model is 0.0083. This means that when all else is held constant, if China were to increase its strictness for canceling public events by one unit, the difference of its daily new cases smoothed (when applied to a logarithmic function) will be expected to increase by 0.0083 units. It is important to note, however, that these equations only explain the increase or decrease of strictness resulting from a change in a certain policy. It is impossible to generalize any causational relationships between new daily cases and government policies, since regression models only indicate whether or not the information of a certain policy is useful in predicting the counts.

**4.2.4 Shortcomings of Discrete Model (Poisson and NB2)**

In Section 4.1.1, we discussed the assumptions of Poisson regression, and one of the four that is included in both Poisson and NB2 is that "the observations are independent of each other." However, the original form of our data is time series, which means that the information of data on one day could be dependent on the data in previous days. Therefore, in the discrete modeling approach, we have ignored the autocorrelation between each observation and the time series aspect in our data. This could be the reason why we did not obtain good predictions that captured trends and fluctuation in each month.

Due to the fact that Poisson regression and negative binomial regression were unable to effectively build a fit model (with 16 individual policy indicators and 1 time variable), we have reason to believe that these two regression techniques are not very helpful in predicting new daily COVID cases (smoothed). This failure also supports the conclusion that there does not seem to be a strong association between COVID policies and COVID cases. This does not mean, however, that COVID cases cannot be predicted using policy indicators, nor does it prove that

COVID cases cannot be predicted using regression techniques. This also does not definitively prove that COVID policies and COVID cases have no association; it merely supports the notion and should not be used as the sole reason to conclude that there is no association. More research should be done on the topic, especially since we only trained these regression models on 8 regions out of the 179 total. No definitive conclusion should be made.

## 5. Analysis Methods - Random Forest

### 5.1 - Random Forest Explained

While regression can be extremely useful in finding significant variables, it may not be the best method for predicting a complicated response variable, such as a time series of COVID cases. To help us get a better understanding of whether policies can predict pandemic cases, we adopt a random forest approach instead of applying regression.

The random forest algorithm is the construction of many decision trees; in brief, a decision tree is the product of a recursive process where different conditions based on each predictor of a training set are evaluated. When the evaluation that best splits the data is found, it becomes a node and the process repeats until the best node combination that predicts the training set is found. A random forest is a combination of these trees, taking the average of the output of each decision tree. The random forest is extremely effective due to its high accuracy in predictions and ability to solve overfitting. Furthermore, the curse of dimensionality (the higher the dimensions, the more data points one needs) does not apply as much to this algorithm compared to other techniques due to the nature of trees not computing every single feature. However, random forests have low interpretability, thus making it hard to understand the data and formulate strong conclusions.

Note that the data is a time-series; thus, we transform the time series data into a supervised learning problem. After that, we perform anchored walk-forward validation (WFV), which is a specific technique of cross-validation. In anchored WFV, a segment within a fixed time frame is selected. Then, a smaller time window within the fixed time frame is chosen to optimize the system. The rest of the time samples within the larger fixed time frame is chosen as validation. The process is then repeated, with the fixed time frame elongated by a certain period of time. Eventually, this process will yield an effective random forest model with the most appropriate parameters. For example, in a hypothetical anchored walk-forward validation situation with a grand total of 1000 days, say that the first iteration's fixed time frame is 100 days. Let's also say that the smaller time window used for optimization is the first 90 days, with the last 10 days being used for validation. In this hypothetical situation, the next iteration will contain a total of 110 days as a fixed time frame: 100 days (from the first fixed time frame) used for optimization and the next 10 days as validation. This process will repeat until 1000 days are used for optimization.

**5.2 - Random Forest Approach**

Despite the low interpretability of random forests, we were motivated to use the algorithm due to its tendency to be strong in predictions. There were 2 main goals:

1. Determine which policies most were effective in predicting COVD-19, if any
2. Determine if it is possible to predict the next day's COVID-19 cases by knowing past policy values

To solve the first goal, we simply look at the variable importance plot. The variable importance plot helps explain how important each predictor is in predicting the response variable; in this case, since we use the 16 individual policies and as predictors to predict COVID-19 cases smoothed (per million), the variable importance plot will be clearly explain how important each policy and time variable is in predicting the COVID cases smoothed per million. To solve the second goal, we look at the prediction rate of the random forest model against a testing data set.

Note that missing values were not imputed but deleted (as is the case for all techniques in this project). Furthermore, to account for time and perform walk-forward validation, the data was transformed by changing the date variables of each country to numerical values representing days from when missing values stopped appearing. For example, if a country has missingvalues up to March 10th, 2020, the date variable starts at 0, which represents the date March 10th, 2020. March 11th is represented by 1, March 12th is represented by 2, and so on. It must also be noted that, due to time constraints, only a total of 14 countries (the representative countries from each k-means clustering model) was chosen to run random forest on: Taiwan, Croatia, Afghanistan, Iran, New Zealand, Macao, Estonia, China, Belarus, Cambodia, Russia, Senegal, Lithuania, and Nigeria.

For walk-forward validation, the first iteration's fixed time frame is 30 days. The smaller time window used for optimization is (n – 30) days, with the next day being used as validation. Furthermore, the time window shift is 30 days. 30 days was chosen for two main reasons: a 30 day shift simply means a month shift; furthermore, 30 days accounts for quarantine lengths in most countries (which ranges from 14-28 days). Note, however, 30 was an arbitrary number, and it was not tuned. There could be a better number of days chosen that would give more accurate predictions. We acknowledge that in order to get the best prediction rates for each region, we need further investigation on the set of policies the random forest algorithm takes and the optimal date shift for specific region by looking at validation. Note that to account for the time shifts, note that we also "save" the previous days' policies. For example, when predicting school policy on day 31, the algorithm will also have information on day 30 as well.

**5.3 - Random Forest Results**

The Appendix section provides the prediction rates of all 14 countries that random forest was performed on (See page 34-36). Notice how a majority of the predictions work decently well, especially as time continues. For example, for Taiwan, the random forest technique struggles to predict the first 10 test days. However, the algorithm is eventually able to predict the trends of Taiwan's cases quite well, ending with a mean relative error (MRE) of 0.175. Iran and New Zealand are examples of regions that predict even better. Notice how the trends of each line is followed almost exactly (such as the rise of cases at around the 23rd day for New Zealand in predictions but there being a rise of cases at 20 days in the real world). Russian cases predicted extremely well, in fact. The prediction cases do not seem to lag behind and, instead, follow almost exactly with the actual cases line. However, of course, there are exceptions. The algorithm seemed to struggle with regions that had no cases, such as Macau. The prediction fluctuates above 0 until eventually beginning to predict 0 at around 21 cases. Cases in Senegal were also poorly predicted, with the algorithm only beginning to predict correctly at 25 days. We note that these results are not as accurate as we had hoped; however, this is partly due to the very few amount of test cases and training data. For many regions, the amount of days that would be fed into the algorithm would only be around 600, which is an extremely small number of observations for a time series. The algorithm was only able to test for around 30 days for each region, which disallows the algorithm to catch up to patterns more accurately. Nonetheless, cases did seem to predict quite well, which leads us to believe that Random Forest may be a good algorithm to predict COVID daily cases using COVID policies.

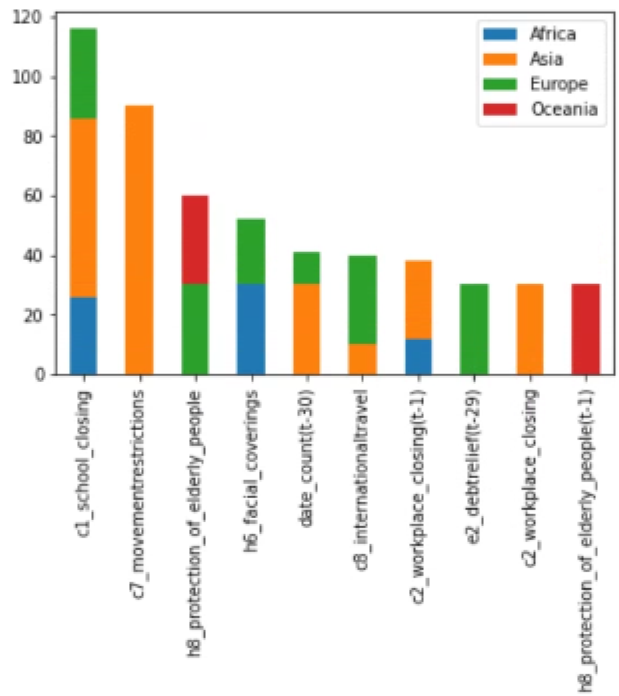



*Figure 10*

In all 14 random forest models, a variance importance plot is outputted, showing the importance of each predictor (policy or date variable) in predicting the response (COVID cases smoothed per million). Of each model, the top 3 variables were chosen to represent the "best 3 variables of the model." Figure 10 is shows the tally of the top 10 variables. In sum, this bar chart shows the top variables that were important in predicting COVID cases smoothed per million, taking into account all 14 regions and their models. Thus, the most important policies in predicting COVID cases are (1) school closing, (2) movement restriction policies, (3) the prediction day's protection of elderly people policies, (4) facial covering policies, (5) the date variable, (6) international travel policies, (7) workplace closing policies, (8) debt relief policies, (9) workplace closing policies, and (10) the previous day's policy of the protection of elderly people. Some of these policies were expected, such as school closures. When schools are closed, children, who are more prone to getting sick than adults, do not spread the sickness. This, in turn, leads to less cases. Movement restriction policies being important was also expected; when people do not move around as much and stay inside their homes, it is much easier to contain a virus due to there being less human contact. However, debt relief being one of the top policies was not expected, since it does not directly seem to affect COVID cases. Debt relief being one of the top policies may be due to the fact that when there are a lot of people sick, regions tend to give more relief because people cannot work.

In sum, random forest as an algorithm seems to do quite well in predicting COVID cases, due to it being able to, largely, capture trends. Furthermore, by using multiple variance importance plots, we tallied the amount of time each policy was counted in the top 3. The top 10 policies with the most counts were: (1) school closing, (2) movement restriction policies, (3) the prediction day's protection of elderly people policies, (4) facial covering policies, (5) the date variable, (6) international travel policies, (7) workplace closing policies, (8) debt relief policies, (9) workplace closing policies, and (10) the previous day's policy of the protection of elderly people. We have reason to believe that these policies are the most important in predicting COVID 19 cases.

## 6. Project Concerns

### 6.1 Skeptical Concerns

While the data is extremely rich and has the potential to accurately paint the pandemic experience, it is important to be skeptical and acknowledge the faults of the project. One of the biggest flaws is that this data set is heavily reliant on the Oxford University COVID Policy Tracker. Any inaccuracies or incorrect scores that were made during the process of calculating the scores will lead to heavily unreliable analyses and conclusions. This project was done under the assumption that all of this data was accurate on a region-level basis. This is due to the fact that there is no reason to believe that the scores were done in an inaccurate and biased way. However, if the results were biased or inaccurate, this project would not be reliable at all and conclusions would be entirely null.

It is vital to note that COVID data is severely underreported in a multitude of regions. Some regions are transparent while others attempt to hide how widespread COVID is in their region; also, regions with a lower GDP are underrepresented in COVID statistics, since test-kits are not as readily available as in regions with a higher average income, such as the United States. Due to these reasons (and possibly many more that cause COVID-related statistics to not accurately display how strong the pandemic hit a region), a degree of skepticism should be exercised. These statistics may be heavily misrepresented, which jeopardizes not only the analyses and conclusion of this project but also its integrity.

Furthermore, this project operated under the assumption that missing values were not missing at random. Missing values seemed to only appear when data was not being tracked or when the data point did not exist at that given time. As a result, imputing values would be detrimental in this project, and deleting missing values was deemed the only solution. However, it must be noted that this is an assumption. If missing values were, indeed, random and should have been imputed, this project would be inaccurate.

It should also be noted that this project does not represent regions within regions (like a state in the US). Data on an inner-region is wildly varied, especially if the area allows its states/provinces to act independently on specific mandates. For example, whereas a state in the United States (like California) may be extremely strict in mask policies, another (like Wyoming) may be extremely relaxed. The mask mandate score for the United States on that day does not fully represent each state (and even each county in each state), thus, should not be treated as such. Instead, these scores should be viewed as a loose representation of COVID policy in the whole region, and there may be inconsistencies as a result.

Time constraints lead to many limitations in this project. We could not observe every single region and analyze it completely. We could not run clustering on every single policy (and ran clustering only on 4 summary policies). We could not apply random forest regression on all regions nor were we able to use a wider range of dates (which would have changed which variables were most important). Although we were careful in concluding anything without concrete evidence, note that any conclusions made could be a result of time constraints. If more time could be made for this project, more resources could be allocated and different results may occur.

### 6.2 Ethical Concerns

There do not seem to be any ethical concerns in our project due to the fact that it is on a regional scale. Since most ethical concerns are on privacy, which deals with data on the individual level, we believe that our data does not have any direct ethical concerns. However, we do acknowledge that weaknesses/biases in our project may lead to unethical conclusions made by unaffiliated and biased parties. For example, one weakness of our project is that it does not represent regions within regions, only taking into account a large region (such as the United States) rather than a state/province (such as California). This weakness can be used to construct inaccurate conclusions with underlying and unethical reasons. To elaborate, one of the conclusions we reach in this project is that regions share a distinct trend of policy strictness. One of these trends is one that heavily oscillates (indicating that the policies are constantly being relaxed and harshened). Afghanistan, which is a region with this trend, also struggles with a high number of COVID cases. One may use this information to wrongfully conclude that Afghanistan and its peoples are underdeveloped (ignoring the fact that our project does not take into account the actions of smaller regions or its peoples), with a goal based on racism and islamophobia.

**7. Conclusion**

By using k-means with dynamic time warping, we were able to find 8 distinct clusters that describe overall government policy trends throughout the COVID-19 pandemic; thus, we have reason to believe our hypothesis that regions can be clustered by their pandemic policy. These clusters each have a region that best represents it, calculated by finding the region with the smallest dynamic time warping distance relative to the average dynamic time-series sequence. In order, Taiwan, Croatia, Afghanistan, Iran, New Zealand, Macao, Estonia, and China are the regions that best represent each cluster. Upon closer inspection, each region has a distinct pattern (i.e. policies become relaxed or more strict in different times or the pandemic). Furthermore, these clusters are representative of the regions, as shown by public perception (i.e. Taiwan, in "Cluster 0" implies mostly strict policies, and that is represented by the fact that Taiwan is praised for its COVID-19 response). Due to the distinctness of each cluster, we have a reason to believe that regions of the world have similar approaches to the pandemic, and can be clustered into different patterns.

K-means with dynamic time warping was also able to cluster other policies. These policies, the containment and health index (which takes into account every containment/closure and health system policy index), the stringency index (which takes into account every containment/closure policy and the public information campaign policy index), and the economic support index (which takes into account every economic policy index), are more specific than the overall government response index. These models have 3, 3, and 5 distinct clusters, respectively. Each k-means model has distinct clusters and show varying levels of policy strictness throughout the pandemic, similar to the model based on the government response index. When taking into account containment and health policies, Belarus, Taiwan, and Cambodia are the regions that best represent each different clustered pattern. For stringency index, Macao, Cambodia, and Taiwan best represent each pattern. Finally, Russia, Belarus, Senegal, Lithuania, and Nigeria best represent each cluster for economic policies during the pandemic. More research must be done on each k-means model in order to conclude whether or not clustering was effective.

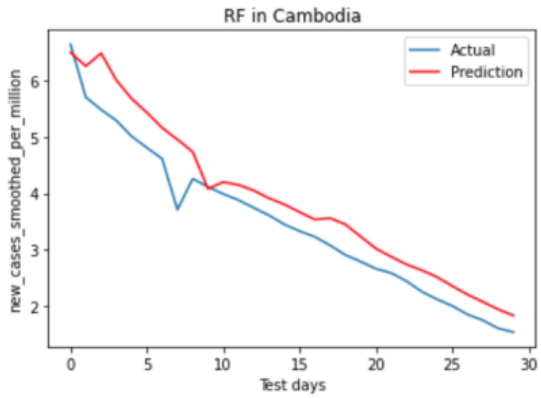
We applied two regression techniques (Poisson and negative binomial regression) to the 8 representative regions of each cluster (which was generated by the k-means model with overall government response index as the response variable). Out of the 8 representative regions, only 3 (Taiwan, New Zealand, and China) were a good fit model according to the chi-square test. Furthermore, the discrete regression models have ignored the time series aspect of the data by treating each observation as independent. Thus, we conclude that these two regression techniques are not strong models in predicting COVID-19 cases using policy indicators. This also acts as support against our hypothesis that there is a strong association between COVID policies and COVID cases. However, note that this does not mean that all regression techniques are unable to predict COVID-19 cases. Furthermore, this conclusion does not mean that all modeling techniques are unable to predict COVID-19 cases (and, by extension, it is impossible to predict COVID-19 cases) using policy indicators. Finally, we should not conclude that an association between COVID cases and COVID policies do not exist. None of these findings should be used to conclude that our hypothesis is false; it is merely support against it. More research must be done in trying to predict COVID-19 before we reach these conclusions.

A random forest with anchoring walk-forward validation was attempted on 14 regions chosen by the 4 summary statistics clusters: Taiwan, Croatia, Afghanistan, Iran, New Zealand, Macao, Estonia, China, Belarus, Cambodia, Russia, Senegal, Lithuania, and Nigeria. The technique was quite accurate in mapping the trends of COVID-19 cases, although in most cases, there was a lag between predicted and actual cases. In a few cases, trends were captured perfectly (Russia and Belarus) and others were unable to capture them at all (Macau and Senegal). We have reason to believe, despite Macau and Senegal not doing very well, that COVID-19 can be predicted using policy indicators in most countries. Furthermore, with the variance importance plot, we tallied the top 3 policies of each random forest model. The ten factors in predicting COVID-19 cases were (1) school closing policies, (2) movement restriction policies, (3) the prediction day's protection of elderly people policies, (4) facial covering policies, (5) the date variable, (6) international travel policies, (7) workplace closing policies, (8) debt relief policies, (9) workplace closing policies, and (10) the previous day's policy of the protection of elderly people. While some of these were expected (school closing policies and movement restrictions), others were not (debt relief policies). Nonetheless, we have reason to believe, due to the variance importance plots, that these 10 policies are effective in predicting COVID-19 policies.

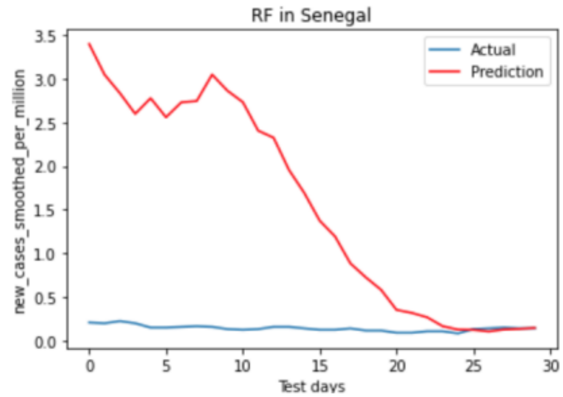
This project was done in order to understand COVID-19 policies in regions and its trends throughout the world. It was also done in order to see if COVID-19 cases could be predicted. Both yielded desirable results, and we learned plenty in this project. This information can be useful for those who are trying similar projects. When trying to use policy indicators to predict COVID-19, random forest may be best.

For future assessment and for those trying to replicate the project, we recommend more than 10 weeks. It took a great deal of time in trying to learn about time-series analysis and data. Furthermore, we recommend for results to be assessed further. Some parts of this project are inconclusive or do not say much in the context of the real world due to the fact that we needed more time on the project to research the field.

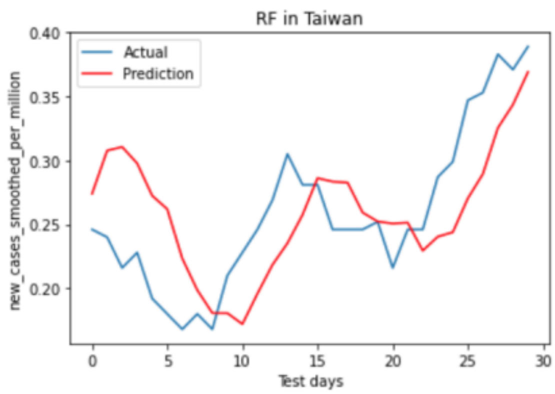
**Appendix: Prediction Plots Using Random Forest**



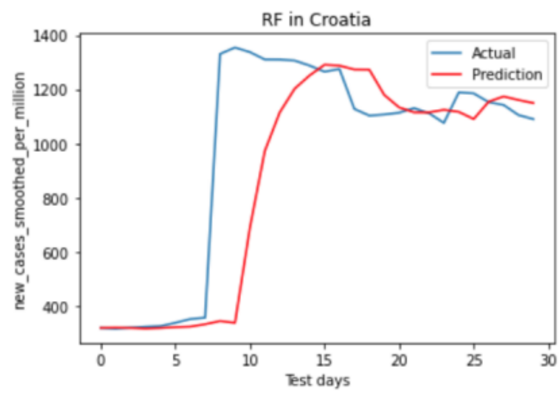

MRE = 0.134



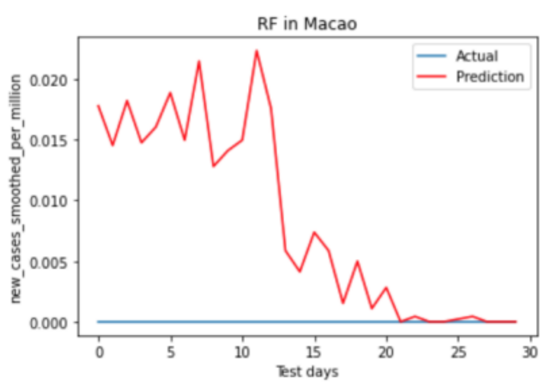

MRE = 9.102



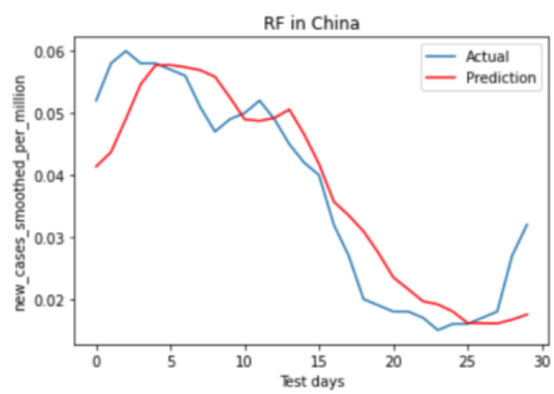

MRE = 0.175



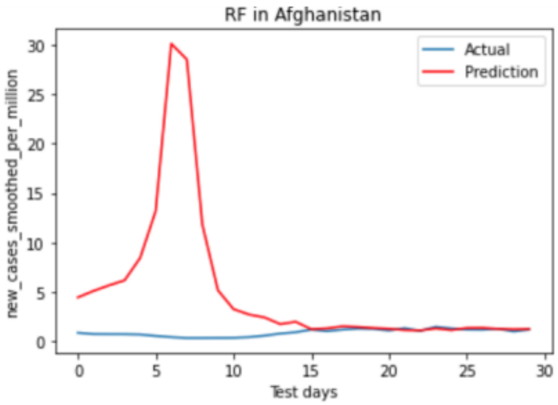

MRE = 0.116



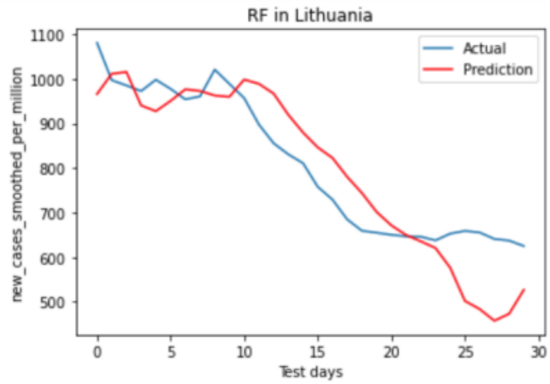

MRE NaN



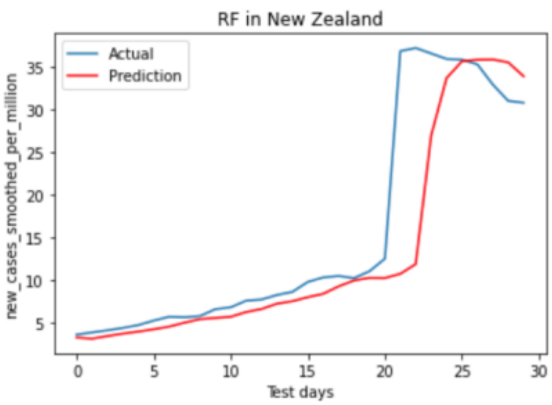

MRE = 0.163



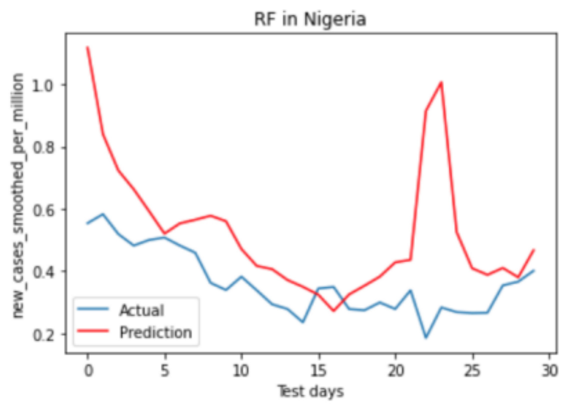

MRE = 9.766



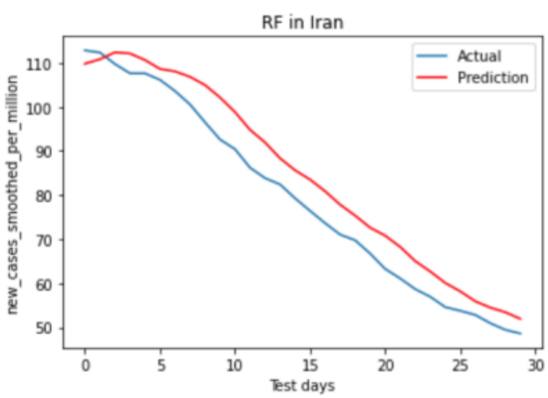

MRE = 0.095



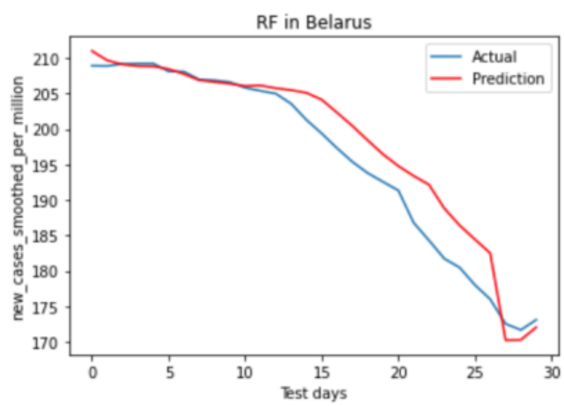

MRE = 0.169



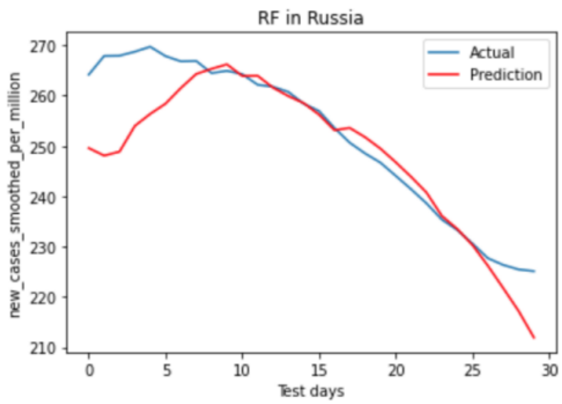

MRE = 0.547



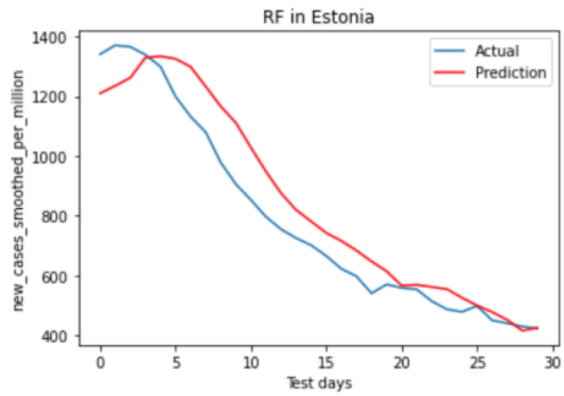

MRE = 0.075




MRE = 0.015




MRE = 0.020




MRE = 0.104

Use this link to go back to Section 5.3: Random Forest Results

**References**

Hsiang, S., Allen, D., Annan-Phan, S. *et al.* 2020. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* 584, 262–267. (https://doi.org/10.1038/s41586-020-2404-8)

Kaufman, B.G., Whitaker, R., Mahendraratnam, N. *et al.* 2021. State variation in effects of state social distancing policies on COVID-19 cases. *BMC Public Health* 21, 1239. (https://doi.org/10.1186/s12889-021-11236-3)

Hadianfar, A., Yousefi R., Delavary M. *et al*. 2021. Effects of government policies and the Nowruz holidays on confirmed COVID-19 cases in Iran: An intervention time series analysis. *PLoS ONE* 16(8): e0256516. (https://doi.org/10.1371/journal.pone.0256516)

Berry, C.R, Fowler, A., Glazer, T., *et al*. 2021. Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic. *PNAS* 118(15): e2019706118. (https://doi.org/10.1073/pnas.2019706118)

Hale, T., Angrist, N., Goldszmidt, R. *et al.* 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav* 5, 529–538 (https://doi.org/10.1038/s41562-021-01079-8)

Cameron, A., Beltekian, D., Gavrilov, D. *et. al*. 2021. Data on COVID-19 (coronavirus) by Our World in Data. *Nat Hum Behav* (https://github.com/owid/covid-19-data)

AJMC. 2021. "A Timeline of COVID-19 Developments in 2020." Retrieved Feb 10, 2022 (https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020)

Haseltine, W.A., 2021. "What Can We Learn From Australia's Covid-19 Response?" Forbes. Retrieved Feb 12, 2022 (https://www.forbes.com/sites/williamhaseltine/2021/03/24/what-can-we-learn-from-australias-covid-19-response/?sh=54aea0f93a01)

Wallace-Wells, B., 2021. "The Good, the Bad, and the Embarrassing in America's COVID-19 Response." The New Yorker. Retrieved Feb 12, 2022 (https://www.newyorker.com/news/the-political-scene/the-good-the-bad-and-the-embarrassing-in-americas-covid-19-response)

AJMC. 2022. "A Timeline of COVID-19 Developments in 2021." Retrieved Feb 12, 2022 (https://www.ajmc.com/view/a-timeline-of-covid-19-vaccine-developments-in-2021)

Amidon, A., 2020. “How to Apply K-means Clustering to Time Series Data.” Towards Data Science. Retrieved Feb 15, 2022 (https://towardsdatascience.com/how-to-apply-k-means-clustering-to-time-series-data-28d04a8f7da3)

Cheng, T., 2021. “How has Taiwan navigated the pandemic?” Retrieved Feb 20, 2021 (https://www.economicsobservatory.com/how-has-taiwan-navigated-the-pandemic)

Lucero-Prisno, D.E., Ahmadi, A., Essar, M.Y, *et. al*. Addressing COVID-19 in Afghanistan: What are the efforts and challenges?. *Journal of Global Health*. 10(2): 020341 (2020) DOI: 10.7189/jogh.10.020341

Harris, C.R., Millman, K.J., van der Walt, S.J. *et al.* 2020. Array programming with NumPy. Nature 585, 357–362 DOI: 10.1038/s41586-020-2649-2.

McKinney, W. *et al.* 2010. “Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference.” p. 51–6.

Tavenard, R., Faouzi, J., Vandewiele, G., *et al.* 2020. “Tslearn, A Machine Learning Toolkit for Time Series Data” vol. 21 p. 1-6. https://jmlr.org/papers/v21/20-091.html

Hunter, J.D., 2007. "Matplotlib: A 2D graphics environment" vol 9 p 90-5. IEEE Computer Soc. DOI: 10.1109/MCSE.2007.55

Seabold, S., Perktold, J. 2010. “statsmodels: Econometric and Statistical Modeling with Python.” Proceedings of the 9th Python in Science Conference.

Van Rossum, G. 2020. "The Python Library Reference, release 3.8.2" Python Software Foundation

Pedregosa, F., Varoquaux, G., Gramfort, A., *et al.* 2011. "Scikit-learn: Machine Learning in Python" JMLR 12, p. 2825-2830

Virtanen, P., Gommers, R., Oliphant, T., *et al*. 2020. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." vol 17. p. 261-72. *Nature Methods* DOI: 10.1038/s41592-019-0686-2