

**COVID-19 Report**

2/c Bill Sergio

Phone Number: +1 (631) 338-6830

Email: [William.J.Sergio@uscga.edu](mailto:William.J.Sergio@uscga.edu)

2/c Matt Varney  
Phone Number: +1 (559) 270-8432

Email: [Matthew.t.Varney@uscga.edu](mailto:Matthew.t.Varney@uscga.edu)

**Table of Contents**

**1 Purpose Page #3**

**2 Introduction  Page #3**

**3    Data Page #3-4**

**4    Origin of the Covid-19 Virus Page #4-5**

**5         What is the coronavirus? Page #5**

**6     The Spread of Covid-19 in the US**  **Page #5-6**

**7 Questions on Covid-19 in the US** **Page #7-12**

**8 The Spread of Covid-19 in China** **Page #12-13**

Purpose

The purpose of our project was to create mathematical models capable of predicting the number of cases of Covid-19 in a geographical region. The method that we applied for producing models was the Auto Regressive Integrated Moving Average. The overarching goal of this project was to analyze the factors linked to the spread of Covid-19 and, in doing so, predict the cases of a geographical region in a given day.

Introduction

Our group was approached by the President of the United States, Donald J Trump, who requested advice regarding the mysterious Covid-19 virus. Mr. Trump supplied our group with a data set from Kaggle detailing all the reported data on Covid-19. Mr. Trump stated the importance of being able to visualize the produced models and emphasized that the models would need to easily take in new data. Before defining our problem statement and ultimately producing mathematical models to predict COVID-19 and to suggest which methods of we must gain a better understanding of the provided data set.

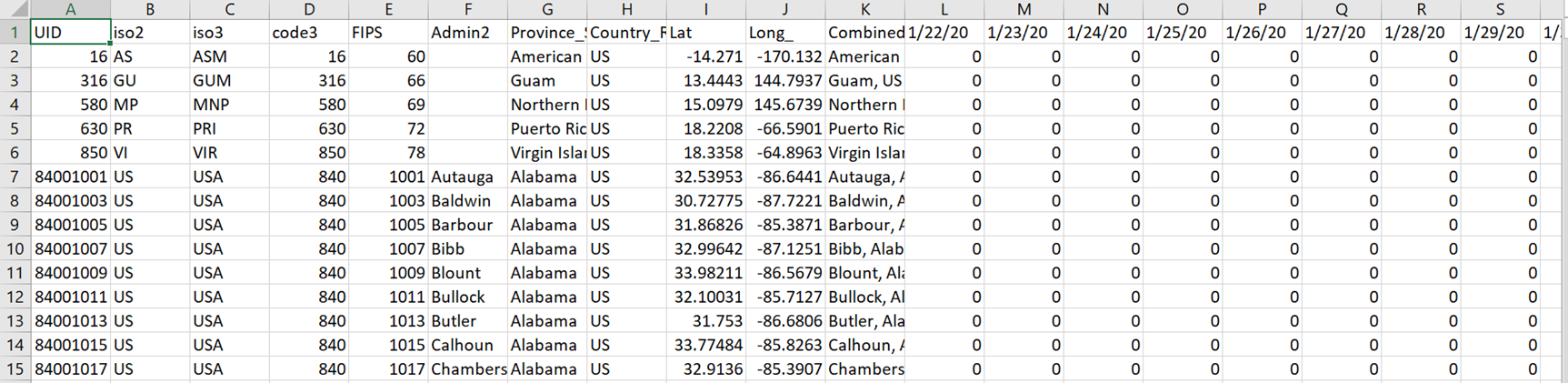
Data

Figure 1

The above is one of 3244 rows from our data set. The variables ‘UID’, ‘iso2’, ‘iso3’, ‘code3’, and ‘FIPS’ all contain different types of identifying codes and are not helpful for the purposes of our model. The variables ‘Province\_State’, ‘Country\_Region’, and ‘Combined’ are not used in our model but are important for us when using smaller geographical regions located within the United States. One such example of this use is for working with large hub airports. Without these variables we would be unable to run smaller models. The variables ‘1/22/20’ through ‘4/24/20’ are the bread and butter of our models and we would be unable to run the ARIMA model without them. They are the total reported cases in a geographical region on that given day. The way this excel spreadsheet is laid out does not easily fit into our needs, so we had to create a csv file for each geographical region we wished to model. The format of the csv files is as follows; date, cases. This made it extremely simple to run through the model without requiring extra code to find and clean the necessary data for each model we wished to run. This allowed us to feed the csv file names into one function that created a model for each file instead of writing up a new code for each new model we wished to create. This gave us a ton of room to answer new questions that required different geographical regions to be modeled and analyzed.

Origin of the Covid-19 Virus

As you know the Covid-19 pandemic has the world under siege, but how did this happen? Well there is four credible theories as to how this pandemic started which we will discuss here. These theories are as follows; the Wuhan Bat theory, China accidently released it from a lab, the CIA released it as a biochemical attack on China, and that the Chinese government released it as a biochemical attack on the US and its allies.

1. Wuhan Bat theory

When the World Health Organization announced the Covid-19 outbreak they claimed that the virus originated from a seafood and animal market in Wuhan. The current belief of the World Health Organization is that the coronavirus started in a bat and was then passed on to another type of animal a pangolin. The pangolin is a heavily trafficked animal in Asia. The pangolin’s meat is considered a delicacy in many Asian cultures and its scales are used as traditional medicine. This would explain how it spread to humans in China, but the fact that it stated in bats and was then passed to pangolin’s seams extremely farfetched. There is currently no explanation as to how the coronavirus was passed to pangolins.

1. China accidently released it from a lab

Many people in the US government believe the coronavirus was accidently released by the Wuhan Institute of Virology due to poor safety and protection standards. The Wuhan Institute of Virology is located just a short distance from the so called “wet” market the World Health Organization claims the virus came from. This claim is also supported by US Government officials and the US Intelligence community. The US Intelligence community has been pouring a ton of resources into investigating the origins of the coronavirus. Many top US Government official’s willingness to back this theory appears to increase the strength of the argument. The US Secretary of State, Mike Pompeo said there is enormous evidence that the Covid-19 virus was not man made and came from a Wuhan laboratory.

1. The US released it as a biochemical attack on China

Top Chinese government officials have been pointing the blame for the coronavirus at the US. Lijian Zhao of the Chinese Foreign Ministry said that the US Army spread the coronavirus when they attended the World Military Games in Wuhan in October 2019. This claim has received little credibility do to the fac that there is zero evidence to prove it. The only two places you really see this claim is from china after people started to question the World Health Organizations bat claim and in the deep abyss of the world of conspiracy theories.

1. The Chinese government released it as a biochemical attack on the US and its allies

Although this claim was deemed “improbable” by the US Intelligence community it has gain traction in many countries such as India and has some alarming facts to back it up. Back in 1999 top generals in the Peoples Liberation Army wrote a Chinese military doctrine titled warfare beyond bounds. This doctrine details how China can defeat a technologically superior nation such as the United States. The military doctrine calls for attacks on the US leveraging Terrorism, economic warfare, Lawfare (similar to proxy war), germ warfare, and network warfare. The outlining of a global Pandemic used to grind the US economy to a halt and replace American goods with Chinese ones appears to a spitting image of Covid-19. An Indian news channel reported with supporting evidence on how China’s economic powerhouses were practically untouched coupled with Chinese medical supplies flooding the market. They also reported that all Chinese factories switched to producing medical supplies prior to the start of the Covid-19 outbreak in December. The timing of Covid-19 hitting the US and strangling its economy fits China’s overarching goal of shrinking the US economy while replacing it with the Chinese economy on the world stage.

What is the coronavirus?

The coronavirus is a family of virus which causes many illnesses in animals and/ or humans. According to the World Health Organization, “several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS).” Covid-19 an infectious disease caused by a coronavirus.

The Spread of Covid-19 in the US

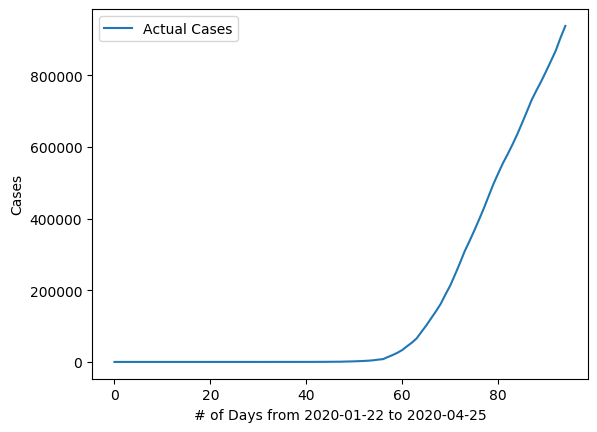
The first reported case of Covid-19 in the United States was on January 22nd. The increase in cases in the US was relatively small until around April 11th when the cases in the US skyrocketed. This is illustrated by Figure 3 bellow.

Figure 3

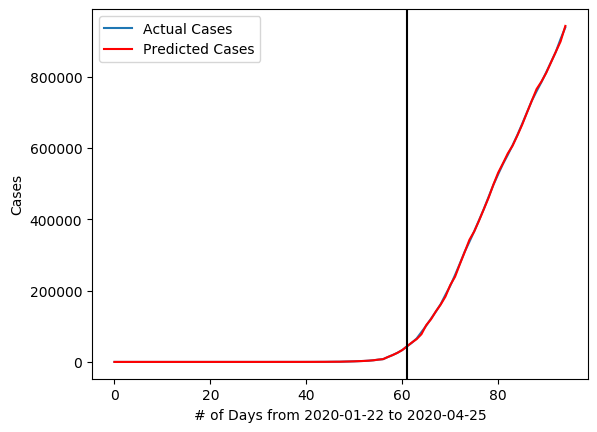
By using an ARIMA(1,2,0) we attempted to create a model to capture the cases in a given day of Covid-19 in the US. As you can see in Figure 4 bellow you can see our model was very good at prediction cases in the US, but how good was it? We don’t currently have any metrics to compare it to the future models we will create within the US and around the world.

Figure 4

To solve this problem, we then implemented three Error measurements. These three measurements are Mean Squared Error, Mean Absolute Error and Mean Absolute Percent Error. The equation for Mean Squared Error is . After plugging into the equation, we got a value of 9146615. The equation for Mean Absolute Error is . After plugging into the equation, we got a value of 2292.3. The equation for Mean absolute Percent Error is . After plugging into the equation, we got a value of 0.9 percent. Of the three Error measurements we calculated we deemed that Mean Squared Error is the most meaningful metric. This is due to the fact that it normalizes the error and allows it to easily be compared to other models.

Posing questions on Covid-19 in the US

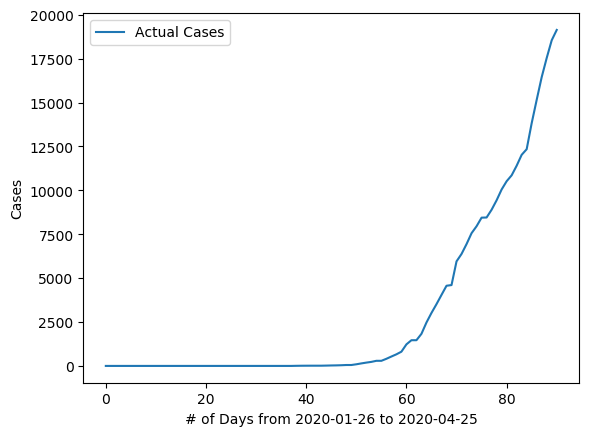
1. Is there a Correlation between Large Hub Airports and the spread of Covid-19?

Before we get started on answering this question lets define a few terms and make assumptions. We define a large hub airport as a US airport that accounts for at least one percent of all US air travel. We are assuming that all airports have a similar impact on the counties they reside in. The airports we selected to study to answer the question are as follows; PHX airport in Maricopa County Arizona, LAX airport in Las Angeles County California, DEN airport in City and County of Denver Colorado, FLL airport in Broward County Florida, MIA airport in Miami-Dade County Florida, ATL airport in Clayton County and Fulton County Georgia, HNL airport in the City and County of Honolulu Hawaii, ORD airport in Cook County and DuPage Illinois, MDW airport in Cook County Illinois, BWI airport in Anne Arundel County Maryland, BOS airport in Suffolk County Massachusetts, EWR airport in Essex County and Union County New Jersey, JFK airport in New York City New York, LGA airport in New York City New York, CLT airport in Mecklenburg County North Carolina, PDX airport in Multnomah County Oregon, and SEA airport in King County Washington. You can see the locations of these airports on figure 5 below.

A picture containing text, map

Description automatically generated

Figure 5

We will compare LAX airport and JFK/LGA airports. In order to do so we will first look at the cases of Covid-19 in Los Angeles County Figure 6 and New York City Figure 7.

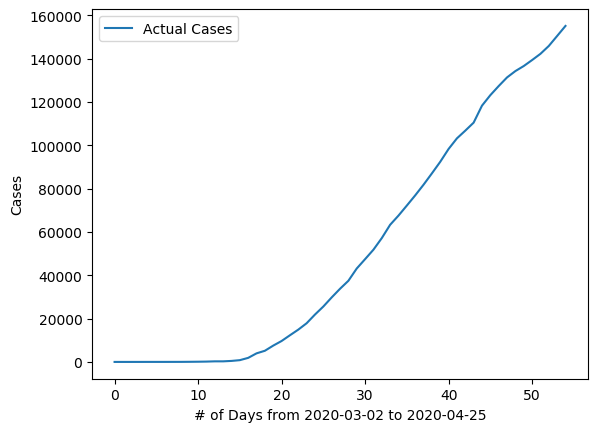
Figure 6

Figure 7

Los Angeles County reached 20,000 cases on April 25th while New York City reached 20,000 cases on March 22nd. On April 25th cases in New York City reached 160,00. That means that there is 8 times more cases in New York City then Los Angeles County. Given that LAX and JFK/LGA are of similar size with LAX servicing roughly 43 million passengers a year and the JFK/LGA combination servicing roughly 45 million passengers a year. This tells us that we can’t conclude that there is a correlation between Large Hub airports and the spread of Covid-19.

1. Is there a correlation between population density and the spread of Covid-19?

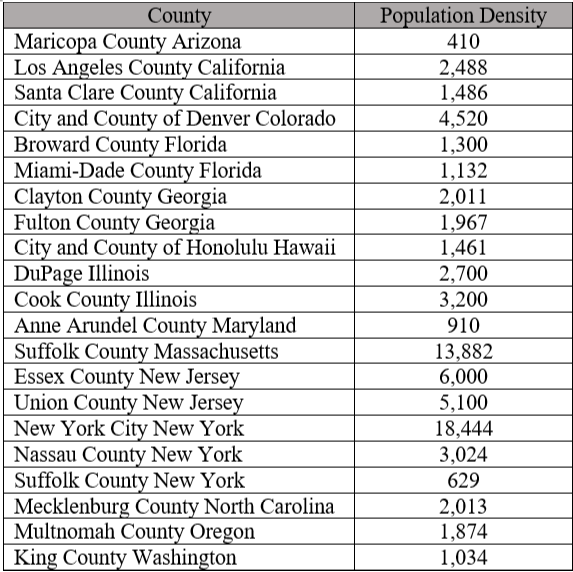
Before we get started on answering this question lets define make some assumptions. We are assuming that counties have been taking a similar response as the same time to the Covid-19 virus. The counties we selected to study to answer the question are as follows; Maricopa County Arizona, Los Angeles County California, Santa Clare County California, City and County of Denver Colorado, Broward County Florida, Miami-Dade County Florida, Clayton County Georgia, Fulton County Georgia, City and County of Honolulu Hawaii, DuPage Illinois, Cook County Illinois, Anne Arundel County Maryland, Suffolk County Massachusetts, Essex County New Jersey, Union County New Jersey, New York City New York, Nassau County New York, Suffolk County New York, Mecklenburg County North Carolina, Multnomah County Oregon, and King County Washington. You can see the locations of these Counties on figure 5. In Figure 8 below is a compressive list of all county’s population densities.

Figure 8

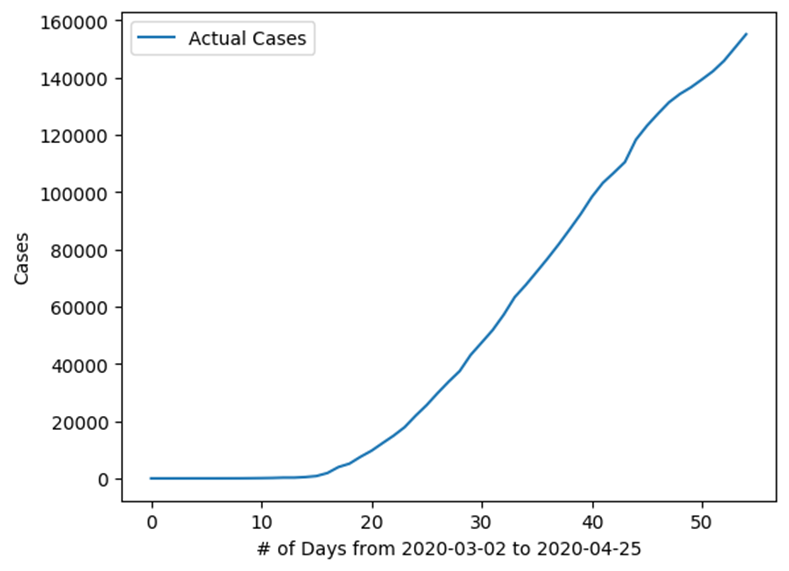
In order to try and find a correlation we will compare New York City and Suffolk County Massachusetts. According to Figure 8, New York City has a population density of 18,444 people per square mile and Suffolk County Massachusetts has a population density of 13,882 people per square mile. Due to this if there is a strong correlation between population density and the spread of Covid-19 we would expect to see similar case numbers and a similar increase in the cases over time.

Figure 7

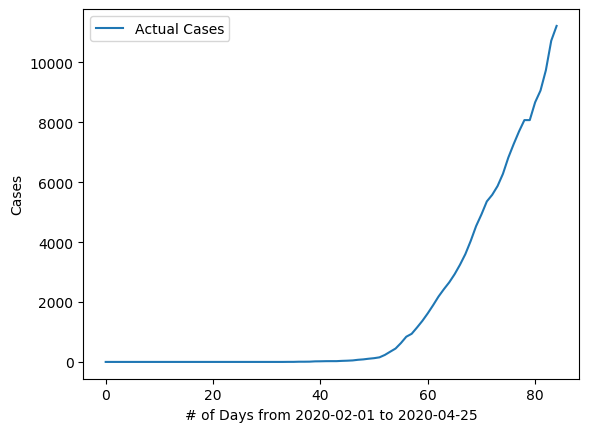


Figure 9

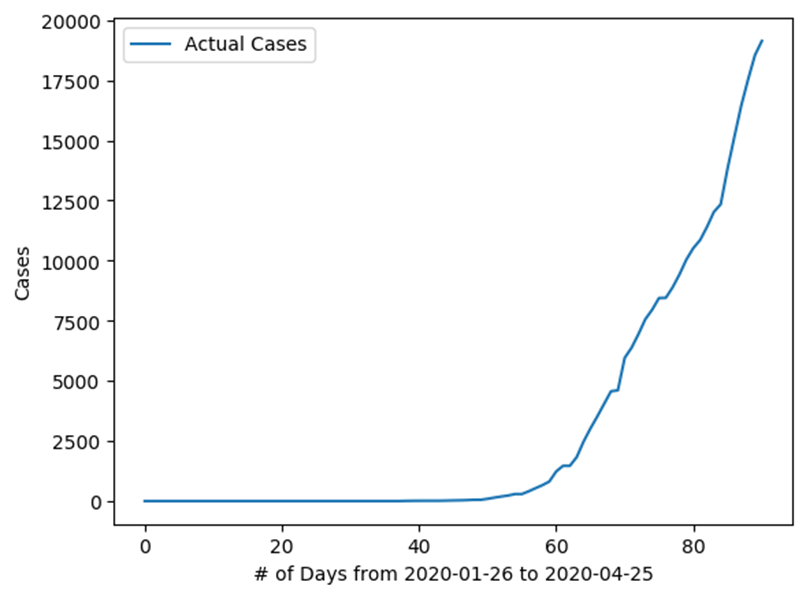
On April 25th cases in New York City reached 160,00 while cases in Suffolk County Massachusetts reached 12,000. Now this seems very odd so let’s compare Suffolk County Massachusetts to Los Angeles County which has a population density of 2,488. Now Los Angeles is less dense by around 10,000 people per a square mile so if there is in face a correlation between population density and the spread of Covid-19 we would expect to see Suffolk County Massachusetts have higher case numbers and a larger increase in the cases over time compared to Los Angeles County.

Figure 6

Now when comparing Figures 6 and 9 you should notice that both have a similar looking increase with Figure 6, Los Angeles County, having a slightly larger increase and 8,000 more total cases. This tells us that there is no strong correlation between population density and the spread of Covid-19, but population density most defiantly plays a part in the spread of Covid-19. We have concluded though that there is many factors causing Covid-19 to spread and that it is a combination of those factors which causes Covid-19 is spread differently in different geographical regions.

1. Is social distancing and quarantining effective at stopping the spread of Covid-19?

This is a question we have been asking ourselves throughout the entirety of our project, but we didn’t really have much to go off of within the US. The geographical region the we can find that didn’t implement so form of social distancing and quarantining was Sweden. Now we can’t exactly compare Sweden to the US for a few reasons. One the population of the US is 309 million people while the population of Sweden is 10.333 million people. We couldn’t possibly compare those Sweden is almost 30 times smaller than the US. That would be like comparing the strength between an ant and a bodybuilder. Second, the climate of Sweden extremely homogeneous while the US’s is extremely diverse. Lastly, Sweden has a homogeneous population and the US has an extremely diverse one. Due to this diverse population the US has vastly different gene pool. The gene pool in the US is like an Ocean while Sweden’s is like a small pound. Therefore, we cannot say with certainty the effectiveness of social distancing and quarantining.

The Spread of Covid-19 in China

Now before we start to discuss the spread of Covid-19 in China we must first make you aware of some concerns we have concerning Chinas data. The US Intelligence Community has reported that the numbers China has been reporting to the World Health Organization are fictional potentially created using an ARIMA or another similar model. It is unsure if this is to save Beijing embarrassment or if there is a more sinister explanation. China is the most populated country in the world. China accounts for 18 percent of the world’s population. China also has been repeatedly been claiming that they have no new cases while at the same reporting to the World Health Organization new cases. According to figure 10 bellow China appears to have Covid-19 under control.

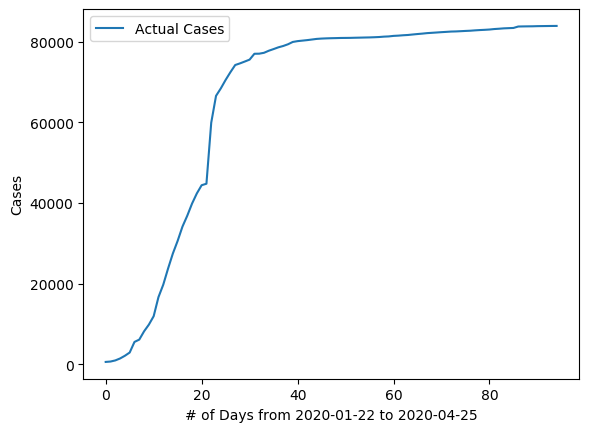


Figure 10

On April 25th China had a little over 80,000 cases. If you recall New York city on April 25th had 160,000 cases alone. This seems odd and should cause questions to be raised about China’s reporting of Cases. For now, let’s put that aside and create a model using ARIMA(1,2,0). In Figure 11 below is the model created for China.

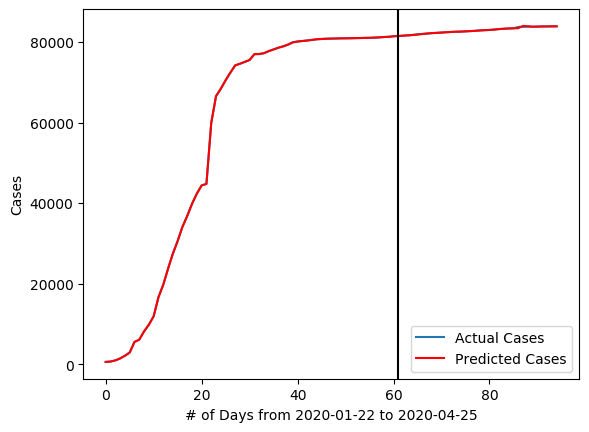


Figure 11

Next we implemented three Error measurements. These three measurements are Mean Squared Error, Mean Absolute Error and Mean Absolute Percent Error. After plugging into the equation for Mean Squared Error, we got a value of 5337.5. After plugging into the equation for Mean Absolute Error, we got a value of 35.3. After plugging into the equation for Mean absolute Percent Error, we got a value of 0.04 percent. Now let’s compare China’s model to the US’s model. If you recall the United States of Americas model had a Mean Absolute Percent Error of 0.9 percent. To put that in perspective that means if there were 1,000 cases the US model was on average off by 90 cases, while China’s model was off by 4 cases on average. So that means China’s model was on average off by 32 cases while the US’s model was off by 9,000 cases on average. That means on average the US model was off by 280 times more cases then China’s model.