10/9/2022

Project 1

Early Pandemic COVID-19 Case Analysis

Mathematical Modeling of Data

Anthony Arnold
NID: 4674586

Objective:

The objective of this project is to analyze the early pandemic data of COVID-19 cases for five counties in the state of Florida using Microsoft Excel. The analysis will consist of the modeling of positive cases over a period (days) and the estimation of the growth rates for each county. Also, each county will be compared with one another to find similarities and/or differences. The exploration of possible causes of said comparisons will be highlighted in the analysis process as well.

Dataset:

"covid-19-data---daily-report-2020-04-14-0943"	**(File 1)** File Type: pdf
"Florida_COVID19_Case_Line_Data"	**(File 2)** File Type: csv
"time_series_covid19_confirmed_US"	**(File 3)** File Type: csv

Analysis Process:

The first step to the analysis process was to choose which counties to work with. I used File 1 and File 2 to guide me in my decision process. I chose Miami-Dade, Orange, Osceola, Sumter, and Flagler as the counties to analyze for this project. In File 3, I filtered the data to only include the five counties in question. This made it easier to locate the data I would work with for this project. Next, I simplified the data by formulating the dates into the number of days starting from their first cases reported and modified the number of cases by grouping them in thousands. With the data now prepared for analysis, I graphed the five counties' COVID-19 cases over a select number of days. The trendline feature in Excel allowed me to compare different data models to see which one fit the data best. I almost decided to go with the power models since these had the highest R^2 values, however, I pondered on the use of the exponential models even though they contain lower R^2 values. The reason behind this is that the nature of an infectious virus over a population tends to act exponentially since susceptible people have a higher chance of getting the virus when more people are infected. Torn between the two models, I decided to model the data

twice. This makes it easy to predict the outcome of the pandemic with the safety precautions set into place by the United States as well as the prediction of the number of cases if no precautions were taken at all. First, I used the power model since it is the best model that fits data when precautions are taken to combat COVID-19. This means it would be the most ideal model to fit the data given.

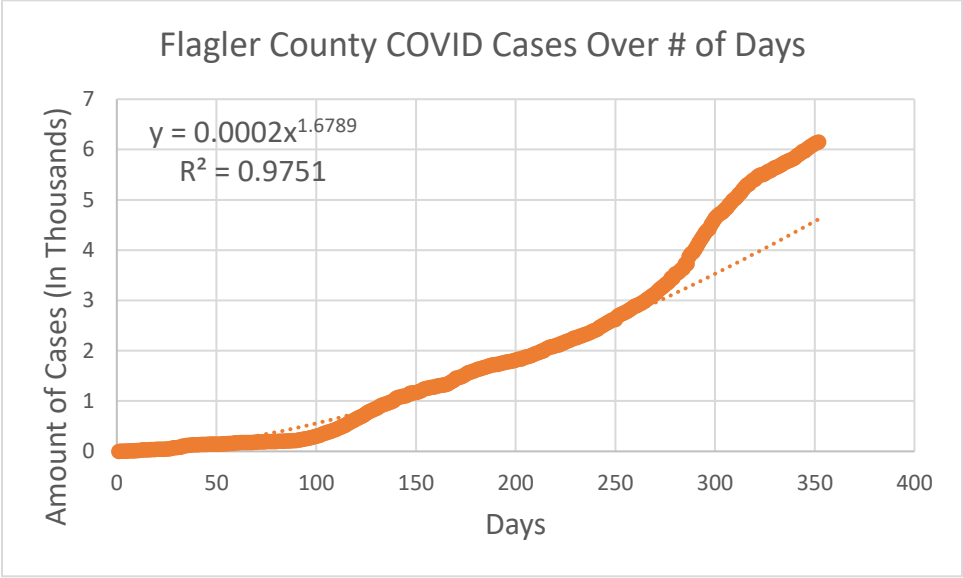

Figure 1.1 (Flagler County)

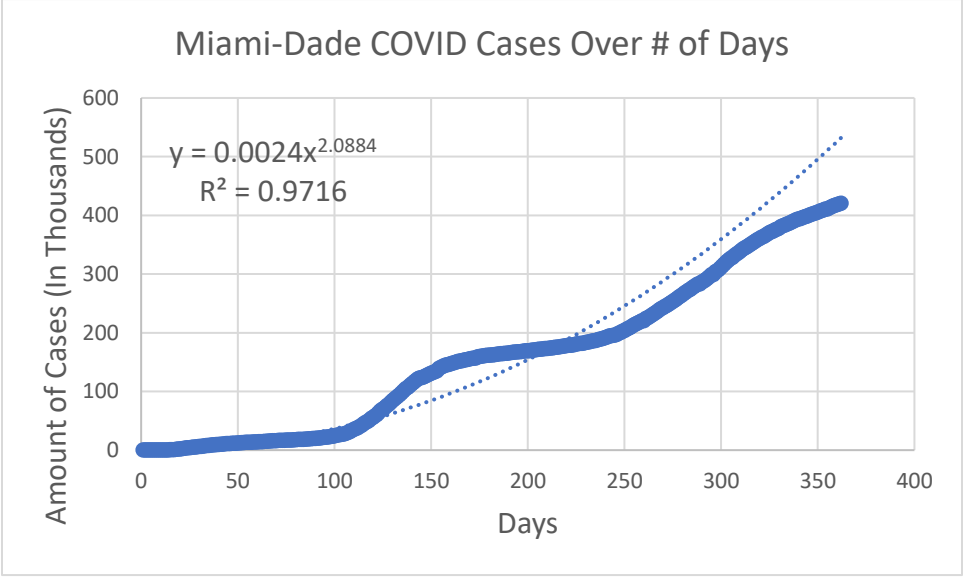

Figure 1.2 (Miami-Dade County)

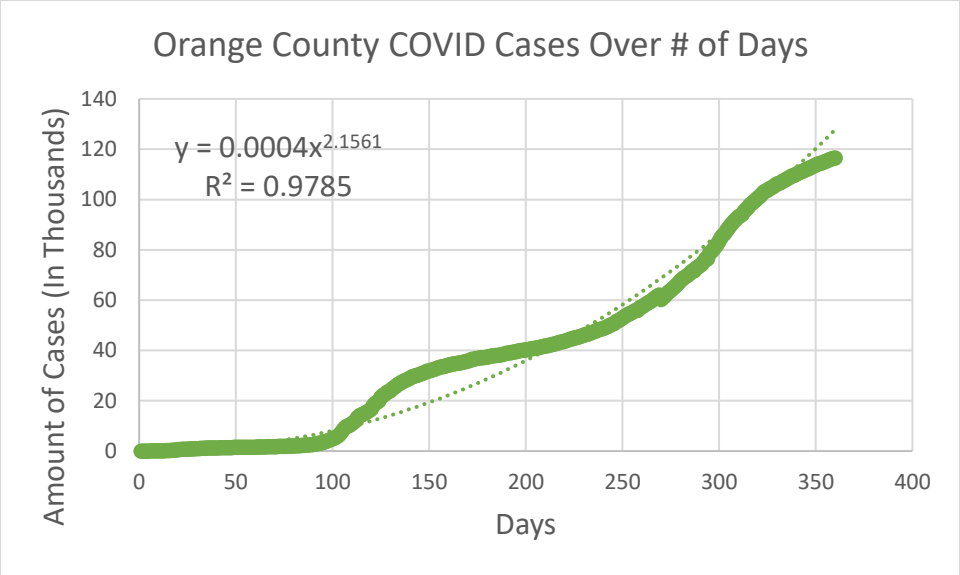

Figure 1.3 (Orange County)

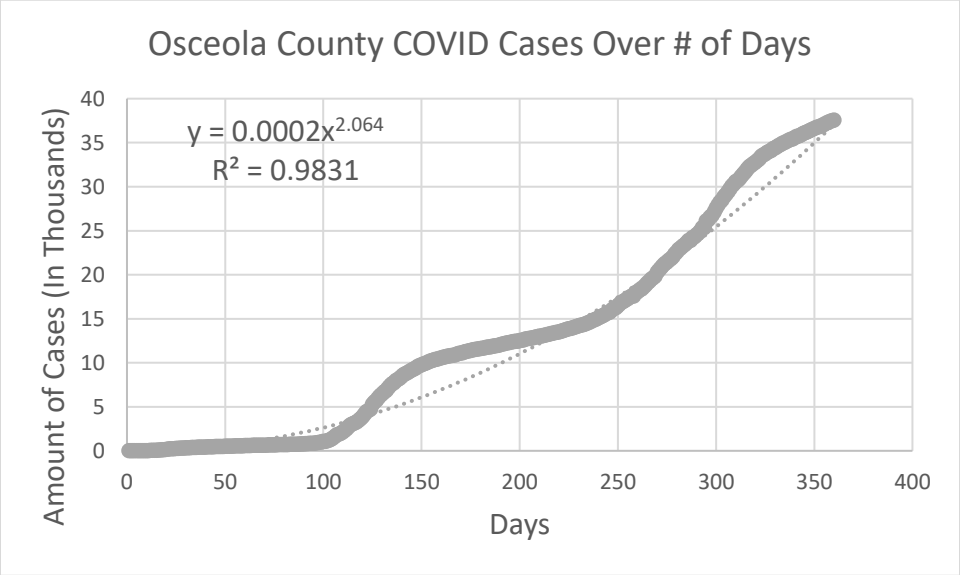

Figure 1.4 (Osceola County)

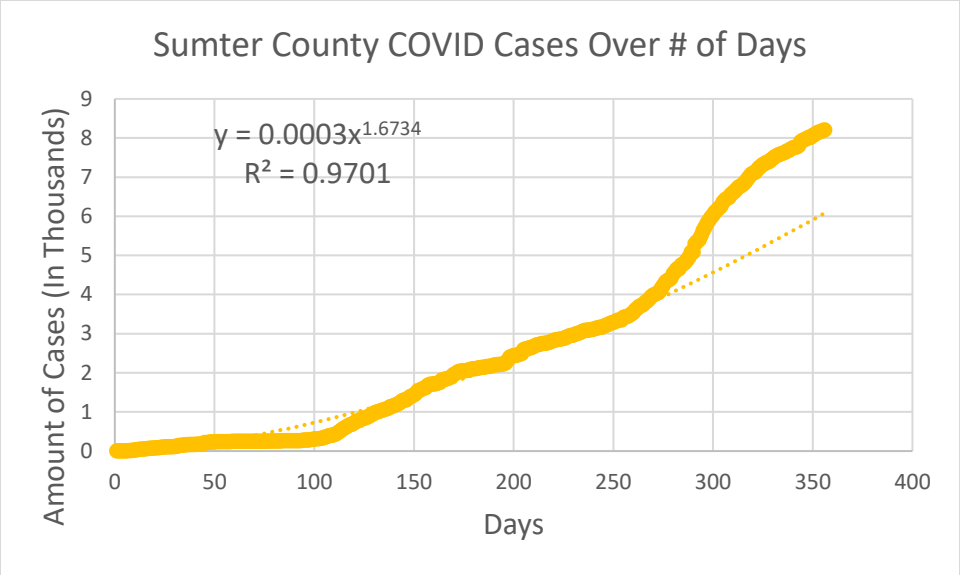

Figure 1.5 (Sumter County)

In the graphs, I was able to spot similarities between the counties very early. For instance, I've noticed that the counties with higher populations (Miami-Dade, Orange, and Osceola) have a "bump" between the 100th and 150th day. This corresponds to the dates between April and June of 2020. According to Kathy Katella, the author of "Our Pandemic Year – A COVID-19 Timeline," between these months several states have started to begin their "reopening" after the United States COVID-19 lockdown. This led to an outburst of covid cases and a huge bump in the graphs. Looking at the Sumter and Flagler counties, the "bump" is very minor than that of higher population counties. This may be because of less population leading to fewer interactions with people. Another similarity between all counties is the small bump at the top of the graph. According to the "CDC COVID-19 Timeline," the increase in cases at the top of the graph correlates to the holidays and colder temperatures, causing people to gather and celebrate fall festivities indoors. Another similarity between the graphs is the model appeared higher than the actual data between the 50th and 100th day. This marks the COVID-19 lockdown mandated by the United States, so it makes sense that the number of cases is lower than that of the model. It is also interesting that the counties all have similar graphs. Every county roughly continues in the same direction of positive cases even though their populations and land sizes are way different. It was surprising to see that Miami-Dade County and Orange County were able to keep the number of cases below the model line at the end of the graph even though their populations and city-like infrastructure would greatly oppose these results. I believe I have found a reason as to why this

may have occurred. In the data, you can see that Sumter and Flagler counties have a super high increase of cases at the end of their graphs. According to "Voter Registration – By County and Party" from the website *dos.myflorida.com*,

	# Of Republican Affiliations	# Of Democratic Affiliations
Flagler County	42,732	26,190
Sumter County	66,703	24,961
Orange County	217,061	360,389
Miami-Dade County	433,113	575,793

Figure 1.6

Highlighted are the political parties with the most affiliations for each county. You can see that Flagler and Sumter counties residents have most of their preferred parties as Republican by a considerable margin. Now, according to "Both Republicans and Democrats cite masks as a negative effect of COVID-19, but for Very Different Reasons" by Patrick Van Kessel and Dennis Quinn, among the study of what Democrats and Republicans thought of masks in the COVID-19 pandemic, "Republicans accounted for 92% of those expressing skepticism or opposition to masks." Although this may be a stretch, the larger number of cases over the given model could be the result of opposition to the precautions mandated by the United States. Notice that this result does jump to a conclusion very early, however, I believe the possibility and correlation of these subjects could lead to interesting insights.

We are left with the use of the model to estimate the growth rates of each county. If $y(t)$ is the number of cases t days from the first reported COVID case, then the instantaneous case growth rate can be estimated by taking its derivative, $y'(t)$.

Flagler County	$y'(t) = 0.0003357t^{0.6789}$
Miami-Dade County	$y'(t) = 0.005012t^{1.0884}$
Orange County	$y'(t) = 0.000862t^{1.1561}$
Osceola County	$y'(t) = 0.0004128t^{1.064}$
Sumter County	$y'(t) = 0.0003x^{0.6734}$

Figure 1.7 (Growth Rates)

Next, I analyzed the same counties using an exponential model. This model would work better under the assumption that no safety precautions were set into effect by the state of Florida. Thus, the mask requirement and six feet social distancing requirement were never set into effect.

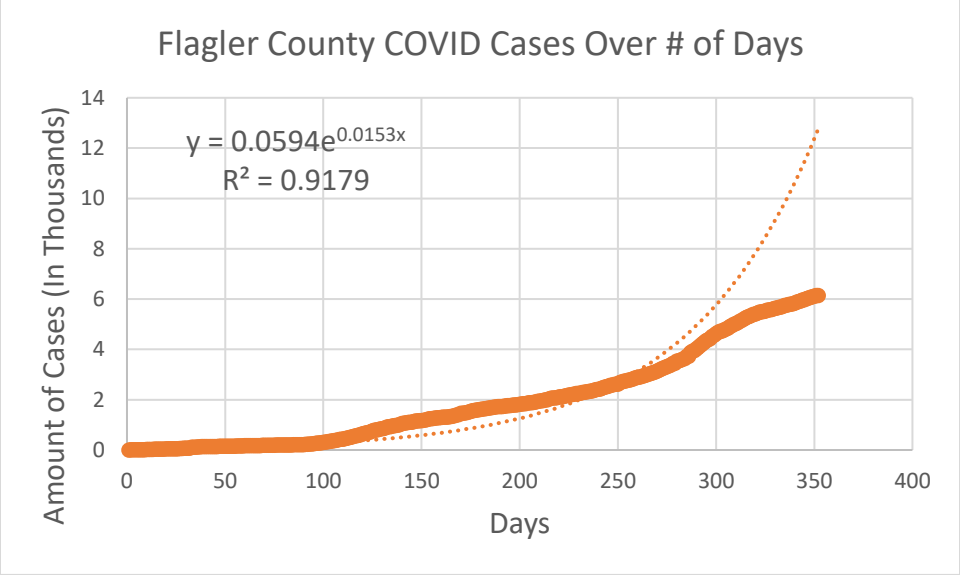

Figure 2.1 (Flagler County)

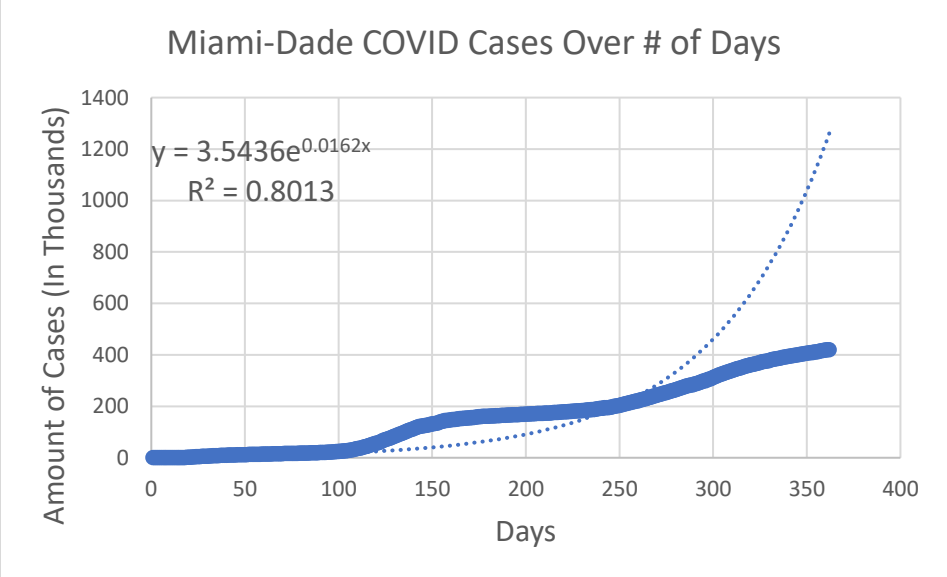

Figure 2.2 (Miami-Dade County)

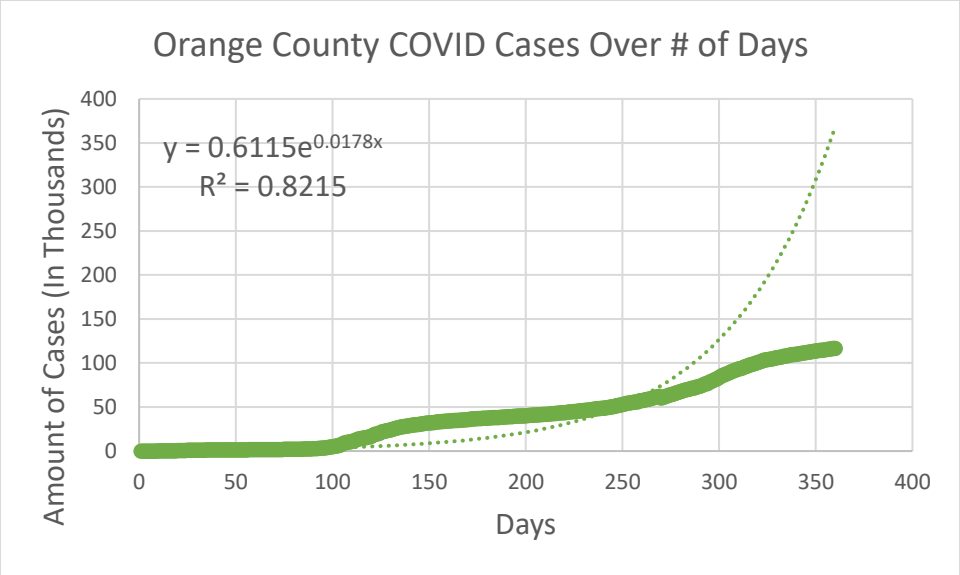

Figure 2.3 (Orange County)

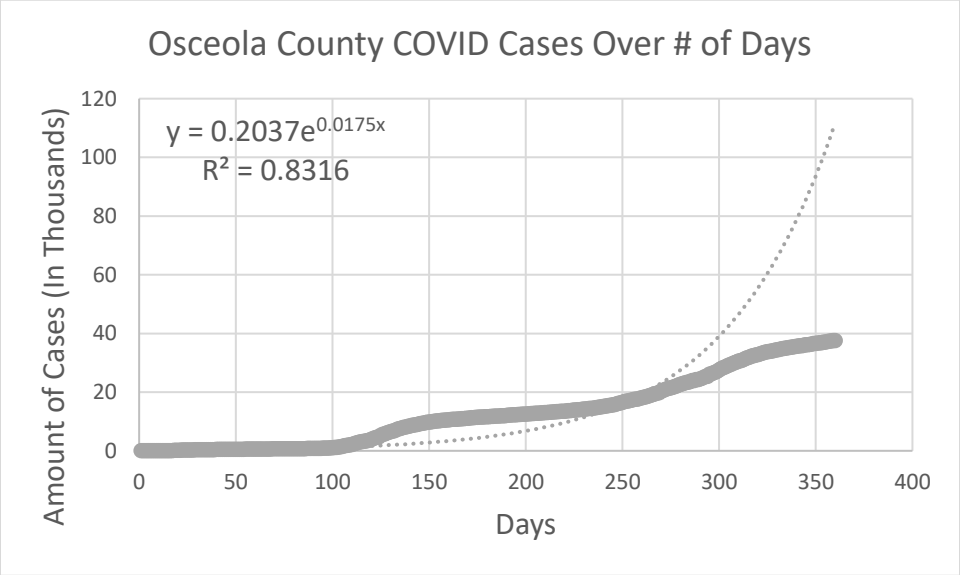

Figure 2.4 (Osceola County)

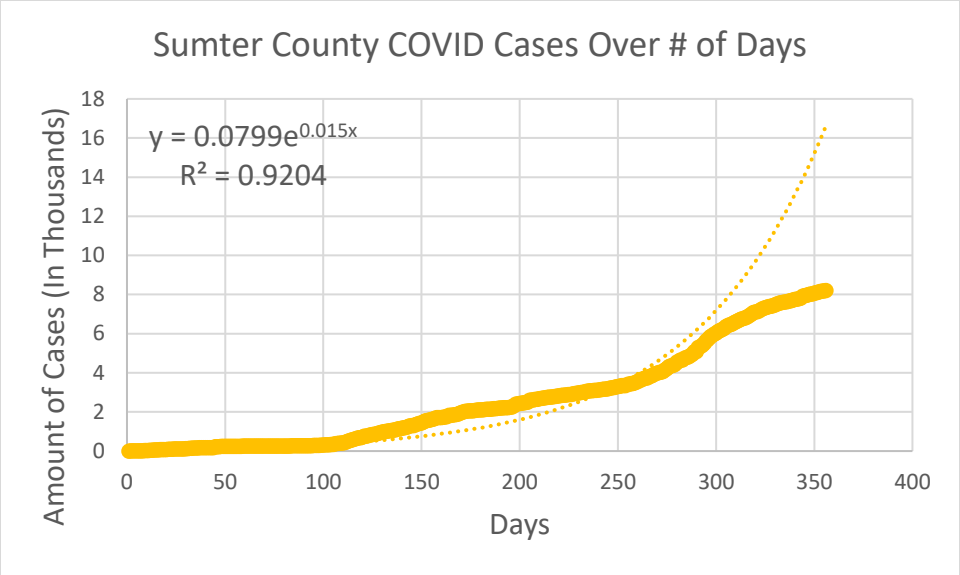

Figure 2.5 (Sumter County)

The similarities highlighted previously can also be seen in these graphs. The exponential model shows how devastating the "reopening" of the state was since the cases went above the model in all five counties between the 100th and 200th day. Also, the huge difference between the model and the actual data points at the end of the graph can be a direct result of all the precautions mandated by the CDC. If the mask requirement and the 6 feet distance rule weren't placed into effect, the data points would be closer to that of the exponential model. Thus, it is reasonable to suggest that the safety precautions did help with the decrease of the spread of COVID-19.

Using the same process to gather the estimation of growth rates for each county, I calculated the growth rate using the exponential models. When taking the derivative of the model, you can see that the instantaneous growth rate is proportional to the population at time t (that is, $y'(t) \propto y(t)$) where the proportional constant (growth rate constant) is the exponent constant. The growth rate constants can also be used to further analyze the similarities and differences between the counties.

Flagler County	$y'(t) = 0.0153(y(t))$
Miami-Dade County	$y'(t) = 0.0162(y(t))$
Orange County	$y'(t) = 0.0178(y(t))$
Osceola County	$y'(t) = 0.0175(y(t))$
Sumter County	$y'(t) = 0.015(y(t))$

Figure 2.6 (Growth Rates)

Flagler County	0.0153
Miami-Dade County	0.0162
Orange County	0.0178
Osceola County	0.0175
Sumter County	0.015

Figure 2.7 (Growth Rate Constants)

Notice that these growth rate functions are merely an estimation of the growth rates of each county since we are using the formula of the models to compute them. The growth rate constants give us a little insight into the counties. It is reasonable for the growth rate constant of Orange County to be the highest since Orlando is a huge tourist destination with a large number of opportunities for spreading the virus. The reasoning behind Osceola County's growth rate constantly being super close to that of Orange County may be the fact that they are neighboring counties and the distance of travel between them isn't far. Sumter County and Flagler County have the lowest growth rate constants. This is very likely due to the fact that these counties have a very low population compared to Miami-Dade County and Orange County. Overall, this leads to fewer opportunities for interactions with the infected population. Also, notice that the growth rates of each county are similar and not far off. This may be likely due to all the counties having the same precautions and practices the state has done to keep the growth of cases in check. Therefore, the exponential modeling and the associated growth rates further prove the similarities and differences highlighted in the power models.

Conclusion:

In conclusion, the analysis of COVID-19 data for five Florida counties led to interesting insights. I was able to use modeling methods to ponder and research on any causes that may produce the similarities and differences found within the graphs. Also, I used differential techniques to calculate the instantaneous growth rates for each county. This led to the conclusion that the counties had very similar graphs and growth rates despite the difference in population which is a direct cause of precautions and safety procedures in Florida. Overall, this project fortified many concepts tied to math modeling and further improved my knowledge of Microsoft Excel. In reality, there are many factors that can change the outcome of growth rates; however, it is reasonable to use modeling and analysis as a basis for understanding the data. In the case of a new pandemic, it is very clear that the data and analysis tied to COVID-19 cases can be used to predict and to lower the growth rates even further. With that being said, the data analysis conducted in this project proves the importance of data and how it can be used to benefit our society in tough times.

Works Cited

Kessel, Patrick van, and Dennis Quinn. "Both Republicans and Democrats Cite Masks as a Negative Effect of Covid-19, but for Very Different Reasons." *Pew Research Center*, Pew Research Center, 2 Nov. 2020, https://www.pewresearch.org/fact-tank/2020/10/29/both-republicans-and-democrats-cite-masks-as-a-negative-effect-of-covid-19-but-for-very-different-reasons/.

"Voter Registration - by County and Party." *Florida Department of State*, https://dos.myflorida.com/elections/data-statistics/voter-registration-statistics/voter-registration-reports/voter-registration-by-county-and-party/.

Katella, Kathy. "Our Pandemic Year-A Covid-19 Timeline." *Yale Medicine*, Yale Medicine, 9 Mar. 2021, https://www.yalemedicine.org/news/covid-timeline.

"CDC Museum Covid-19 Timeline." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 16 Aug. 2022, https://www.cdc.gov/museum/timeline/covid19.html.