**The 2020 US COVID-19 Pandemic**

DSC530 Final Project

Trish Girmus

11/21/2020

As the new decade of 2020 was a promise of exciting change, hope and a new chapter for many, this year has proven to be anything of that sort, and quite the opposite. While we are facing uncertainty and our hope is now for a vaccine, many of us struggle with life and the unknown. I wanted to study and analyze the data during this pandemic to learn more about the number of tests, cases, and deaths statewide over the past 10 months (January-October). After receiving feedback from my original hypothesis, I decided to change it. My new hypothesis for this project is: can we assume that the US average amount of cases per month is five million?

In hindsight, my questions and hypothesis were not great, I will be honest. My questions were easily answerable, as I had received feedback on that as well. Unfortunately, I learn my lessons as they say, the hard way. However, this was a great learning experience for me to answer my questions with the analysis I performed – and for that I am grateful. I learned through my EDA and hypothesis that there were not five million cases per month, based on the data I had. However, if I the time to really analyze and perform EDA, I would have done this project completely differently.

If I could reverse time, I would have not used this data set for this project. It was a very large file and contained too many variables that in my opinion, some which seemed like duplicates. I was not sure if the data was entirely accurate or not as it appeared to overestimate the values. Taking the time to educate myself on the data was also a valuable lesson, especially with the state FIP codes and learning through trial and error that the monthly totals were cumulative. I struggled with the PMF and CDF requirements for this assignment and plan on re-reviewing those. My biggest challenge with this assignment was time. I regret not having it as I really had high expectations of how this project would have turned out and quite disappointed with it. I learned that these types of projects need to allow extra time for coding error, the presentation piece and all the details that go into this type of work. I also feel that I have struggled through this semester with the concepts of statistics, and while learning all of this information in the past few weeks, the concepts and ideas have blurred together, making it difficult for me to understand.

Overall, I enjoyed this project and learned a lot by applying what we learned in a real data set.