Guoyi Chen, Rohit Mathew

Prof. Taoufik Ennoure

CIS 3120

22 May 2020

Project Documentation

The primary purpose of this project was to collect the COVID-19 data provided by New York State and perform different, more in-depth analyses of the data supplied by each county within the state. In this project, Rohit focused primarily on creating data frames, using pandas, and the majority of the data analysis. Guoyi focused more on some parts of the data analysis and all data visualization using matplotlib. This report aims to provide a better insight into our coding, including the inputs, processes, and outputs of our data.

**The input** for this project was the data we got from “health.data.ny.gov”. We tried to organize the data by county and by date.

**The process** for this project was:

1. We installed all the necessary packages including pandas, sodapy, datetime and matplotlib to analyze the data better
2. Since the data the website provided for us as a string, we cleaned the data and made data points into numbers, which are a readable format in python
3. We defined four functions to identify county name, test date, county information on the data frame and the total cases by day in NYS. The purpose of this was to ask user for the specific criteria order to provide them the needed information for the menu.
4. We also created the menu which includes five different topics to give user a variety of options.

**The output** includes

1. The program, which provides a summary of current county statistics, comes with the total number of positive cases, negative cases, the total number of tests, and % of the positive cases in a specific county.
2. The user interested in the COVID-19 stats for a specific date could also access the number of positive cases and negative cases for a specific date in the county they picked.
3. We also included the county which has either the highest or lowest positives cases in NYS
4. The program provides the user with bar charts, which include the number of new cases per day in NYS to give a better visual understanding of the covid-19 in NYS
5. We also provide a summary of the new cases in each county with a bar chart.

For the future, we would like to include the total population from each county and NYS so we could use this to understand better how COVID-19 affected the people who got tested and the communities as a whole.