Theodore Koby-Hercsky

November 11, 2021

DSC540-T301

**Milestone Five**

Throughout this semester I was tasked with a major project that involved the gathering, cleaning, and manipulating of data from all different sources such as a csv file, website data, and an API. I focused my report on the current epidemic caused by Covid-19 and to see what the data behind the virus can unravel. Throughout this report I learned how to import my data many different was, format my raw data to replace headers and use locate to identify and update values within my data set. While also learning how to identify outliers and bad data using shape functions and “isfinite” to get rid of outliers and bad data. On another note, I was introduced to Fuzzy matching which opened my eyes to a new way of formatting and cleansing bad data. As the use of Fuzzy Wuzzy aloud me to close matches and typos to values such as Alaska and “Alaskaa” with a typo. I was also introduced to SQLite which made the creation of databases a reality which I have never done before this class. As I was able to create a connection to import my dataframes using sqlite\_table. I also learned how to open my SQLite database back up on a new Jupyter workbook using the sqlite\_connection function. While also learning the ethical implications of cleansing data I thought about when making data cleaning and analysis decisions to think to myself if what I am changing in the data and or removing was justifiable. Which meant to me that the visualizations I am creating are telling a true story and are something that I could stand for and know the data I am presenting is true and accurate to the best of my ability. In the end I felt as if this project was an amazing opportunity for me to learn and build on my coding knowledge as I can say I learned new tips and tricks that I will use in the years to come.