Analysis and Cleaning of Data

When I first transferred the Data from Open Data Syracuse into my G-Drive, it was an absolute mess. The numbers didn't always line up, there were countless blanks and lots of missing information. There was also a lot of additional information that wasn't necessary to me in my analysis of the data. It took me a long time to figure out how to read it. As I cleaned the data and attempted to sort through it, I found a lot more information on Open Data Syracuse that I knew would be useful. The original data set that I had chosen for this project was entitled "Crime Data 2022 - Part 1 Offenses (With Lat / Long Info)." What I didn't realize was that this was only half of the data for the given year. In the end, I had to go in manually and find both Part 1 and 2 for a given year and combine them into one google sheet. This meant downloading both parts, combining the google sheets, and then cleaning up the data so that they matched each other and could be used for further analysis.

After finding the Part 1 and 2 for the data from the last four years and combining them into one spreadsheet per year, I was able to get a grasp on the data and how to sort through it. I began with creating pivot tables of the addresses/blocks and the count of crimes reported at each one. I did this for 2019, 2020, 2021, and 2022. Then, because my data had anywhere from 8,000 to 14,000 rows in a given year, I chose to just look at the top five blocks with the highest number of crime reports. This led me to the discovery that Destiny USA has held the number one spot for most calls reported year after year.

With this discovery, I created an additional pivot table for the years 2019 and 2022 that had a filter of just the top five addresses. This made it much easier to transfer the data into datawrapper. It also helped me to create a graph that was much easier to read. I did some additional research and sorting on larceny codes on the google sheet, but this didn't make it into the article.