**Project Reflection**

Using the *NYC Open Database*, I was able to find various datasets that piqued my interest. Since I enjoy learning about the environment and the growth of our species, it did not take long to find a suitable dataset. I decided to choose the *Water Consumption in the City of New York* dataset. Originally my project idea focused on the increase in population in New York City and the increase in water consumption. However, because of the large number of technical difficulties I had with that data, I settled on viewing the increase in the city’s population. Thankfully, this dataset was still viable since it included many years spanning over decades and the population for each year. This allowed me to focus on the problem that is interpreting the data into a viewable graph.

There were many parts of the development process that were found to be challenging. Due to the use of external python libraries, often errors would be found within the library itself due to my inexperience with the library and python. This created a massive challenge of being unable to find solutions in the terminal, the code, and online sources as well. While this occurred many times after I decided to change what the project would be about, it also occurred many more times before that. Overall, the most challenging portion of the development process was just trying to find examples that were not beyond my scope of experience. I found various examples online however they continued to either be barely related to the issues I was experiencing or too complicated to be able to include in the project due to my lack of knowledge. Finding the dataset used in this project was a lot easier than I expected however the number of available datasets on the *NYC Open Database* slowed down my decision-making speed drastically as I fumbled through the site looking for an appropriate choice.

As mentioned previously, I had to drastically change what my project idea was due to the number of issues implementing the idea as well as the lack of answers found online. Originally, my project idea was to use a dataset that focused on the water consumption in New York City over decades of time and the amount of people taking part in said consumption (i.e., population). Sadly, this idea had to be abandoned due to the incoming due date of the task as well as other required work. Instead of using the original project idea, the *Water Consumption in the City of New York* dataset was retroactively used to complete the new project idea which focused on population increase over a set amount of time.

If there were more time to work on this project, there would be many changes and additions towards it. Ideally, I would revert to my original project idea however if that was no longer feasible, I would have liked to expand the range of data even further as well as properly stylizing the graph itself to make it more presentable. If the dataset used were larger, my chosen hypothesis would further be supported.

My main takeaways from this experience revolved around the complexity of data and the importance of its usage. Regarding the complexity of data, there are many much larger datasets than the one I used for this project. However, even with a dataset at this minimal size, there were many problems developing this project. As for the importance of usage of data, it is clear to me that it is very easy to create visually professional datasets or graphs while avoiding any amount of information from the original dataset. This creates another form of misinformation potential and its relatively easy to learn format creates an even more dangerous weapon for malicious individuals. As mentioned previously, there are many things I would do differently starting from the dataset used and the project idea itself to the type of graph that was displayed.