**Project Documentation**

For this project, I decided to utilize the *NYC Open Database* to find a fitting dataset. Since this project mainly focused on the increasing population in New York City over a long period of time, there were many options available. This project utilized the python programming language as well as the python libraries pandas and matplotlib. The overall goal of this project is to learn whether New York City’s population has increased over a period with or without external factors and display the result in a formatted manner. This is also the hypothesis for this project or more specifically stated; does New York City’s population increase over a long period of time.

I used the *Water Consumption in the City of New York* dataset which can be found using *NYC Open Database*. This dataset primarily focuses on the correlation between the increasing population of New York City (NYC) as well as the increasing water consumption. The dataset also provides a number of years spanning decades to provide more thorough information. Although this data was found on *NYC Open Database*, it was originally uploaded by the Department of Environment Protection (DEP) and is updated annually. The main possible biases that came to mind while developing this project included the external factors influencing the usage of water as well as the population in New York City. For example, perhaps the external factor of the COVID-19 pandemic affected the number of people still living in New York City as well as the amount of water that was consumed due to the city lockdown mandates. Another possibility is the altering of data from individual(s) within the agency that uploaded this dataset due to the continued increase in bipartisan interest that is seen throughout the entirety of the United States of America. Since the project primarily focused on the increase in population in New York City over a few decades, the concerns regarding the data on water consumption are not valid to an extent.

The first step to complete my analysis was finding and viewing the dataset. Following this, I had to ensure that I understood what the data was presenting to me. After this, the data was imported and read using the python pandas library. The data was then formatted to be ready for display which presented itself in the form of a bar graph. This graph showed the increasing population of New York City over the course of a few decades and supported the original hypothesis.

The results of this project are very clear and show that the population in New York City does increase over time. This is shown through the display of the bar graph using the matplotlib python library. For a small dataset such as the one used in this project as well as my limited knowledge surrounding data analysis in python, the outcome was expected but not guaranteed. This is especially the case when considering the external factors mentioned earlier in this document. In the future, this project could easily be expanded out to not only include many more recorded years, but the population from other cities in the United States as well. There are many possibilities surrounding the expansion of a project such as this one.