The project required significant data munging to find out if there was a correlation between popular movies and television shows of a certain year and the popular baby names of the same period. The information was found using API and BeautifulSoup. The information was then saved into a DataFrame through Pandas. Data.drop was used to remove extra information unnecessary to the overall analysis. This type of information included criteria such as “Adult films” or “genres”. The movies pulled from the DataFrames were for a certain time range and consisted of only the top 10 movies to provide information that was simpler to manipulate. An interesting aspect to this process was that there were movies without characters listed online, so they were dropped from the data set. This information was then graphed using Matplotlib and Plotly.

The project was an interesting process that required a lot of manipulation and joining of data set. An interesting aspect of the project was pulling the specific Forloops to provide specific datasets from the Dataframes built pulling the API’s and BeautifulSoup. The loops took significant time, which is why the information had to be reduced to specific criteria. The time spent waiting for the results of queries shows the power of the tools used and the magnitude of the data the tools run in. It also showed the importance of functions like Groupby or Sortby when consolidating information. While I have used these function before, I have not used the use.explode method. This method was sused to make each character name its own row.