Final Report

For this project we wanted to find a data source that had meaning to both group members. After searching for a while on Kaggle, we found a dataset containing information about the Star Wars Universe. Looking through the dataset we found that was split into 5 different csv files, each one detailing its own part of the world. The website is https://www.kaggle.com/jsphyg/star-wars.

Taking the csv files, we preformed some basic maintenance. Making sure there were no blank values by preforming drop.na and updating values to include the correct information. We also removed a few columns of information, some of them were not relevant to what we wanted to show. Removing the designation, average heigh and average lifespan columns from the species CSV, as well as removing similar columns from the planets CSV.

Once that was done, we transferred the data to Postgres. In Postgres we began on creating individual tables for the data to be in. The tables are based on the csv files, allowing us to compile them with the exact columns the file contains. Then the multiple tables were joined on key characteristics including planets, characters and home world. Allowing for a wider array of data to be viewed at the same