For Milestone 5, I converted the DataFrames I developed from the csv, web scraping, and API data sources to csv files using: **df.to\_csv(‘newcsvfile.csv’, index = False).** (The index=False was to get rid of the default index in each table, as I did not need them). By doing this within the Jupyter Notebook, it saved the new csv file in the same folder as the DataFrame.

I made a few minor changes to the tables at this point, including standardizing the name of the column that contains the movie titles to be ‘Film’ for all three, and fabricating a few Oscar awards in order to have more movies in common among the three data tables (to avoid sparse data visualizations later in the project).

Then I created a SQLite3 database using DB Browser for SQLite and created three tables within it by importing the three csv files. I set the primary key for each data table in SQLite using ‘Modify Table.’ I added some more common data using ‘New Record’ to add rows to the body count and awards databases, using movie titles from the TMDb/popularity database so that my relationship among the three datasets (which is movie title) would retrieve more than FIVE MOVIES when merging the three!

At this point I needed to decide where to merge the tables and did it upon importing the database tables into the Jupyter Notebook via Python. This proved more of a challenge than I anticipated, as I learned that SQLite does not support full outer joins. Once I completed the join, I created a scatterplot using variables from two of the original sources.

I completed the four other required visualizations for this project using data imported into Power BI from the DB Browser SQLite database (that I connected to in order to pull this data into Jupyter Notebook for Python wrangling). They are saved as a pdf as requested.