**ETL Project**

1. Obtained 3 csvs from Kaggle named movies\_metadata, MoviesOnStreamingPlatforms\_updated, and wiki\_movie\_plots\_deduped
2. Copied all movie titles, year released, streaming services, and IMDB rating from MoviesOnStreamingPlatforms\_updated into a new Movie table and assigned a unique ID number to each record.
3. Used advanced filter on wiki\_movie\_plots\_deduped file to remove all movies that did not appear on a streaming service based off the movie title and year released.
4. Used this new table to gather director and cast data. Some records in these columns need to be manually changed to ensure that all data was in the same format for the next step of the process.
5. Used text to columns on director column to separate movies with more than one director.
6. Imported data into pandas and used wide\_to\_long function to place all directors in one column based on movie id number. This table became Join Table 1.
7. Took the directors column from Join Table 1 and created a new csv called Director Table. Removed all duplicate directors and assigned each director a unique ID number.
8. Repeated steps 5-8 for actors instead of directors and created Join Table 2.
9. Final step was to perform an index match from movies\_metadata to the Movie Table to gather budget and revenue data based off of the movie title and the year released.