Group 6: ETL Project

# Data Sources:

Our data sources include a collection of movies from a CSV file found on Kaggle, which explores film, film ratings, directors, current movie streaming on various platforms and more. In addition, our team decided to perform API calls on TMDB to obtain financial information based on available movies from our CSV file.

# Extraction:

The CSV file was imported as a DataFrame with pandas. To create the movie financials dataset the API calls to our TMDB we pulled a list of films from our CSV file and created API calls based on title name and the film year.

# Transformation:

The transformation portion of our project was to utilize a the financial information pulled from our API and perform a join with the CSV data set, in order to create a new table called “movie\_data”, which would later be referenced through our flask application.

# Loading:

The loading process was done through jupyter lab, which sets up the code to create the tables for the user through postgresSQL. Once the jupyter lab file is run, the user is able to run the flask application, which has our final table results and also has access to the raw data sets.

# Final Database:

Our final database is a join on our films that have available financial information such as revenue, budget and estimate profit. The final database illustrates film financials along with which movies are streaming on different platforms like Disney+, Netflix, Hulu, and others.

# Tables in Database:

There are a total of three tables created and are labeled as the following: “movie\_data”, “movie\_financials”, “movie\_streaming”. The movie financials table includes film revenue, budge, and a calculated estimated profit. The movie streaming table includes director, ratings, which platforms the movie is streaming and more. The final database as mentioned above illustrates a combination of both for users to perform interesting analysis.

# Hypothetical Usage for Database:

There are couple of interesting analysis that can be performed with our final database. The main one that we wanted our users to explore is what movie streaming services stream the most profitable, largest budget, and highest revenue films. Which streaming platform includes the widest range of movies.