ETL Project

By: David Lu, Lizeth Orozco, Stephanie Tang

**Extraction**

We used 4 datasets from the public platform Kaggle to compare ratings from the online streaming platforms Disney+ and Netflix. All four datasets were provided in a CSV format.

* Disney+ Shows
* Netflix Shows
* IMDb Movies
* IMDb Ratings

**Transformation**

To clean up the datasets, we used Pandas functions in Jupyter Notebook to load the CSVs. We reviewed the datasets to determine relevant information. We dropped unnecessary columns and duplicate rows using the dropna and drop\_duplicates functions. Renaming of columns was also performed.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Load**

The newly transformed datasets were then loaded into the Postgres database.