Project Report

For this project we are searching/scraping different file which will provide us with certain information about movies and the platforms they are streamed in, the idea is to transform this information in order to create 3 datasets that will be used to analyze different aspects between the movies and the platforms.

Extract: we used five .CSV files downloaded from Kaggle and the OMDB API’s which are:

http://www.omdbapi.com/

AllMoviesDetailsCleaned.csv

IMDb\_movies.csv

Movies\_metadata.csv

Netflix\_titles.csv

Rotten\_tomatoes\_top\_movies\_2019-01-15.csv

Transform: the data was transformed by cleaning it up from NaN values and columns that we did not consider relevant, joining or merging different files to complete the datasets, separate values from columns and aggregating them into new ones.

We decided to create 4 catalogue tables which are: cat\_genres, cat\_original\_language, cat\_production\_companies, and movies\_details and a single detailed table called movies\_db with the tool PostgreSQL.