**Final Report**

**Extraction:**

The sources for our database were three different data sets obtained from Kaggle.com. This data sets were:

* <https://www.kaggle.com/unanimad/disney-plus-shows>
* <https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset>
* <https://www.kaggle.com/ruchi798/movies-on-netflix-prime-video-hulu-and-disney>

This data was downloaded in a csv format.

**Transformation:**

We used pandas to transform the csv´s into data frames and then cleaned by dropping the unrequired columns. We focused on the following criteria:

"title", "director", "country", "year", "genre", "avg vote" and consolidated the three data frames in to one single data frame using the append method and dropping the duplicate rows.

**Loading:**

We exported the tables to csv´s and to SQL.

We used SQL because we worked with a simple relational database that was better suited for querying our data.

SQL databases are used to store structured data while NoSQL databases like MongoDB are used to save unstructured data. MongoDB is used to save unstructured data in JSON format.