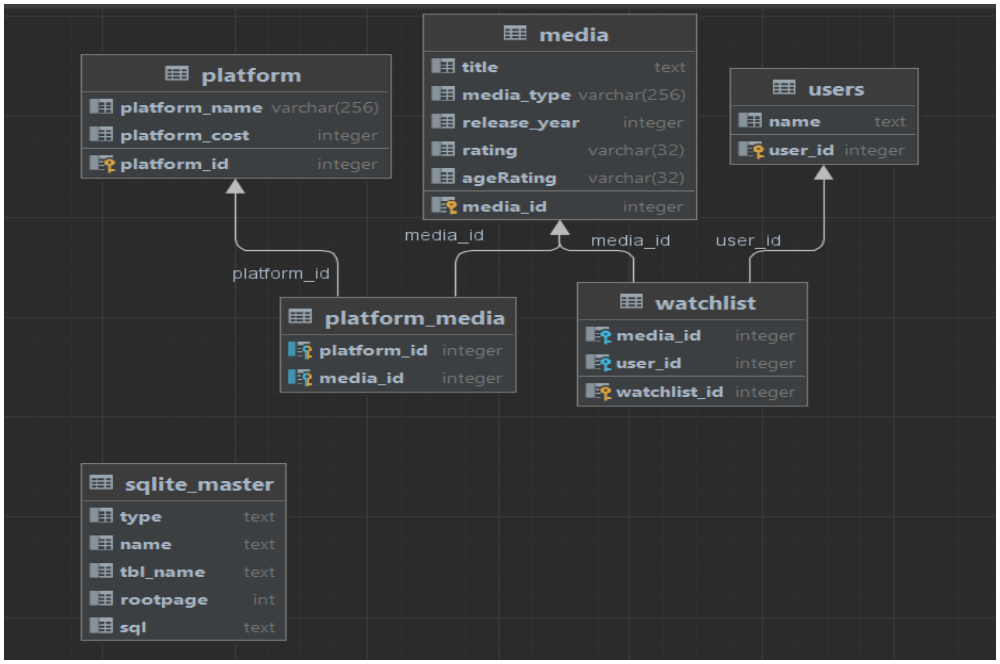
Final Report Platform Database

Marisa Mini, Patrick Polcuch, and Pelin Top

TV streaming has become an increasingly popular way for people to watch their favorite shows and movies. With the rise of platforms like Netflix, Prime Video, Hulu, and Disney+, it is now possible to access a vast library of content from the comfort of your own home. However, with so many options available, it can be challenging to find out which platform a particular TV show or movie is available on. The database we created is a solution to help ease the process of finding the movie or TV show the consumer is looking for, below are the steps and processes we took.

To make it easier for users to find and track their favorite shows and movies, we have created a database that stores information on TV shows and movies, as well as the platforms on which they are available. Our site allows users to browse through the titles and add them to their watchlist, making it easy to keep track of what they want to watch. With our database, users no longer have to spend time searching multiple platforms to find their favorite shows and movies – all of the information they need is right at their fingertips. Below is the picture of our schema and how we were able to get our five tables into the third normal form.



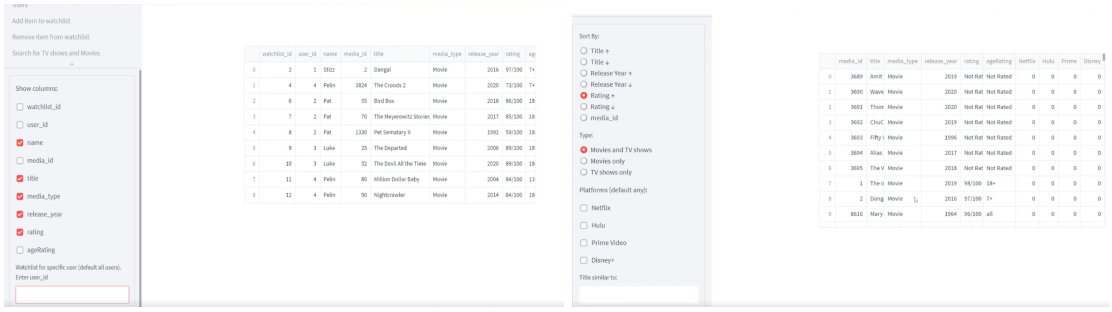
As you can see from the picture of our schema above, we have five tables for: platform, media, users, platform_media and watchlist. Our primary key is platform_id which is an integer in the platform table. We split our media table into two composite tables which are platform_media and watchlist. Our primary key in our media table is media_id which is our composite/foreign key and in our platform_media and watchlist tables. Our primary key is user_id in our user table. This is how we were able to get our tables into 3rd form.

To gather the data for our database, we used a variety of sources, including Kaggle. Kaggle is a website that hosts a wide range of data sets that can be used for research and development purposes. We were able to find a comprehensive dataset on TV shows and movies from the four platforms we included (Netflix, Prime Video, Hulu, and Disney+). This dataset included information such as the title, release date, genre, and rating for each title, as well as which platform each title was available on. We found Kaggle to be a tremendously helpful resource, as it provided us with over 1400 columns of data that we could use to populate our

database. By using Kaggle, we were able to gather a large amount of data quickly and efficiently, which helped us to create a comprehensive and up-to-date database for our users.

Once we had gathered all of the necessary data on TV shows and movies from the various streaming platforms, we used SQLite to store and organize the information. SQLite is a popular and widely-used database management system that allows us to easily create and maintain a searchable database. With SQLite, we were able to efficiently store and organize the data in a way that made it easy for users to browse and filter through the content. By using SQLite, we were able to create a user-friendly database that allows users to quickly find and access the TV shows and movies they are interested in. So now we have our 3 data sources, we had to download the data and write some python code to add the data we needed to our tables. In order to do this, we needed to parse the data from the downloaded data, then write the sql commands to add the data, then write a loop in python to add in all of the data.

Our database is designed to be user-friendly and easy to use. We used StreamLit to create a simple search bar that allows users to quickly find the TV show or movie they are looking for and see which platforms it is available on. Streamlit is an open source python library and the platform information is clearly displayed for each title, making it easy for users to find out where they can stream their favorite shows and movies. Additionally, our database allows users to filter their search results by platform, genre, and other criteria, making it even easier to find exactly what they are looking for.



As you can see from the screenshot above on the left, you can filter and select the columns you want to see, which is an easy way to filter out the data. In the upper right corner on the side, you can also see there is a selection if you want to add or remove items(movies or tv-shows) from the watchlist. When you click on the users tab, you can add the name of the individual in which you want to create a watchlist for. You are also able to add as many individuals as you want to which is extremely convenient because on most platforms, you can only add up to five different users.

You can also filter by movies only, TV shows only and then both of them together as shown in the screenshot on the right. Below that, you can click which platform you want to use which are either: Netflix, Hulu, Prime Video or Disney Plus. Below that, we included a feature in which you can search by a title name that is similar to another title name. For example, if you searched up the cast, the database would return you with the titles and movies: King in Grass Castles, The Outcast and many more that have the word cast in it. You are also able to sort by order, which could be: title, release year, rating or media ID.

Overall, our TV streaming dataset and database make it easy for users to find and track their favorite shows and movies, and help them stay up to date on the latest releases and recommendations. With the increasing popularity of streaming platforms, we believe that this

type of database will become an essential tool for anyone looking to keep track of the vast array of content available to them.