Summary

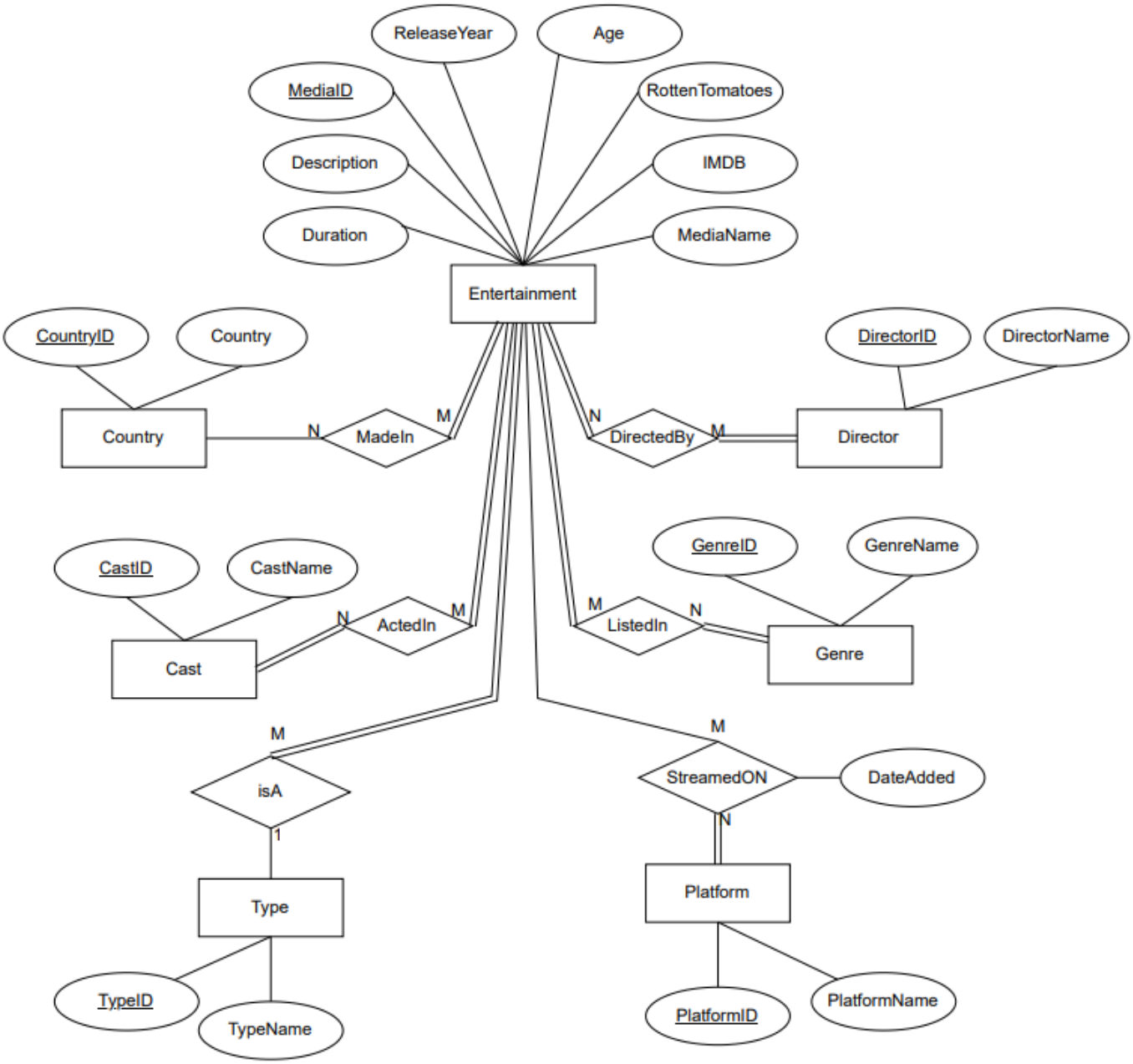
Part 1: Data Discovery:

Given that the data set had to come from a reliable open source, we began our search keeping in mind that we required a sizable data set. To make the dataset more interesting, we chose a number of common datasets. Our search for an intriguing dataset led us to the datasets of movies and TV series available on various streaming services on Kaggle, which is an open source website for datasets.

Part 2: Dataset

Movies and TV shows from various streaming platforms make up the dataset. There are about 15,000 rows in our main table. It offers a variety of details, including the title, streaming services, cast and crew, directors, plot summary, IMDB and RottenTomatoes ratings, and much more. With the help of this database project, it will be possible to find out a lot of information on a certain title simply by typing it into a search bar.

Below is the ER model of the dataset and the final relational model after normalization:


ReleaseYear
Age
MediaID
RottenTomatoes
Description
IMDB
Duration
MediaName
Entertainment
CountryID
Country
Country
N
MadeIn
M
N
DirectedBy
M
Director
DirectorID
DirectorName
CastID
CastName
Cast
N
ActedIn
M
GenreID
GenreName
M
ListedIn
N
Genre
M
isA
1
TypeID
TypeName
Type
M
StreamedON
N
DateAdded
Platform
PlatformID
PlatformName

Normalization:

0NF :
Entertainment (mediaID, name, releaseYear, age, rottenTomatoes, IMDB, duration, description, dirID, director, castID, cast, countryID, country, platformID, netflix, hulu, amazon, disneyPlus, typeID, type, genreID, genre, dateAddedToPlatform)

1NF :

Entertainment(mediaID, name, releaseYear, age, type, typeID, rottenTomatoes, IMDB, duration, description)
Director (dirID, dirName)
Genre (genreID, genreName)
Cast (castID, castName)
Country(countryID, countryName)
Platform(platformID, platformName)
StreamedON(mediaID, platformID, dateAddedToPlatform)
DirectedBy (mediaID, dirID)
MadeIN (countryID, mediaID)
InvolvedCast (castID, mediaID)
MediaGenre(mediaID, genreID)

- **Here we broke down multi valued attributes with its keys into their own table with new relationships that maintain the connections.**
- **Multivalued attributes in our table - Director, Genre, Cast, Platform and Country**

2NF:

- **Since we do not have any partial dependencies in any of the above relational schema, we are done with 2NF. So it is same as 1NF.**

3NF:

Entertainment (mediaID , name ,releaseYear, age, Rotten Tomatoes, IMDB, duration, description, typeID)
Director (dirID, dirName)
CountryID (countryID, countryName)
Genre (genreID, genreName)

Cast (castID, castName)
Platform(platformID, platformName)
Type (typeID, typeName)
isA(mediaID, typeID)
StreamedON(mediaID, platformID, dateAdded)
DirectedBy (mediaID, dirID)
MadeIN (countryID, mediaID)
InvolvedCast (castID, mediaID)
MediaGenre(mediaID, genreID)

- **Since in entertainment, mediaID determines typeID, and typeID determines type, we would breakup typeID and type into a new table with its new relationship isA(mediaID, typeID)**

BCNF:

- **Since all determinants are superkey, we are in BCNF form.**

Post normalization relational model :
Entertainment (mediaID , name ,releaseYear, age, Rotten Tomatoes, IMDB, Duration, Description, typeID)
Director (dirID, dirName)
CountryID (countryID, countryName)
Genre (genreID, genreName)
Cast (castID, castName)
Platform(platformID, platformName)
Type (typeID, typeName)
isA(mediaID, typeID)
StreamedON(mediaID, platformID, dateAdded)
DirectedBy (mediaID, dirID)
MadeIN (countryID, mediaID)
InvolvedCast (CastID, MediaID)
MediaGenre(mediaID, genreID)

Justification of cardinality and constraints of relationship:

- **Titles have to be made in at least one country, but it's not necessary for all countries to have a title, hence it has total participation from one side and partial from other. Even animated/comic titles are made in some countries. Also one country can have many titles and one title could be a result of production of many countries, hence many to many relationships.**
- **Titles must have at least one director, and director cannot exist in a database without a title, hence it is a total participation. Also a title can have more than one director and a director could have directed more than one title, hence it is a many to many relationship.**
- **Titles must have at least one cast member, and it is a must for a Title to have at least one cast. Even in animated/comic titles, there are voiceover casts. Hence it is a total participation. Also there could be more than one cast in a title, and an actor could be in more than one title. Hence it is a many to many relationship.**
- **We are considering all titles in our dataset has a genre, and all genres have a title related to it, which makes it a total participation. A title could have more than one genre and a genre could have more than one title, hence it is a many to many relationship.**
- **We are considering, our titles are either TV shows or Movies only, hence it is a total participation from one side, however, our database could not include all Movies or Tv shows that exists, hence there is a partial participation. Since there could be more than one TV show or Movie, but a title could only be either one of them, it is a one to many relationship.**
- **Platforms have to have at least one title to be considered as a streaming service, but a title doesn’t have to be in on any of the streaming services, hence it is a partial relationship. A title could be streamed on more than one streaming service, and a streaming service could stream more than one title, hence it is a many to many relationship.**

REFERENCES

Amazon Prime Movies and TV Shows. (2021, October 12). Kaggle. Retrieved October 15, 2022, from https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows

Disney+ Movies and TV Shows. (2021, November 29). Kaggle. Retrieved October 15, 2022, from https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows

Hulu Movies and TV Shows. (2021, October 25). Kaggle. Retrieved October 15, 2022, from https://www.kaggle.com/datasets/shivamb/hulu-movies-and-tv-shows

Netflix Movies and TV Shows. (2021, September 27). Kaggle. Retrieved October 15, 2022, from https://www.kaggle.com/datasets/shivamb/netflix-shows

***Movies on Netflix, Prime Video, Hulu and Disney+*. (2021, December 16). Kaggle. Retrieved October 16, 2022, from https://www.kaggle.com/datasets/ruchi798/movies-on-netflix-prime-video-hulu-and-disney**

***Netflix Disney+ Prime Video Hulu Shows Collection*. (2022, September 13). Kaggle. Retrieved October 16, 2022, from**

https://www.kaggle.com/datasets/whenamancodes/netflix-prime-video-disney-hulu?resource=download

Bhatia, R. (Ed.). (2022, September 13). ***Netflix Disney+ Prime Video Hulu Shows Collection*****. Kaggle. Retrieved October 15, 2022, from https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows**