Data Preparation Process
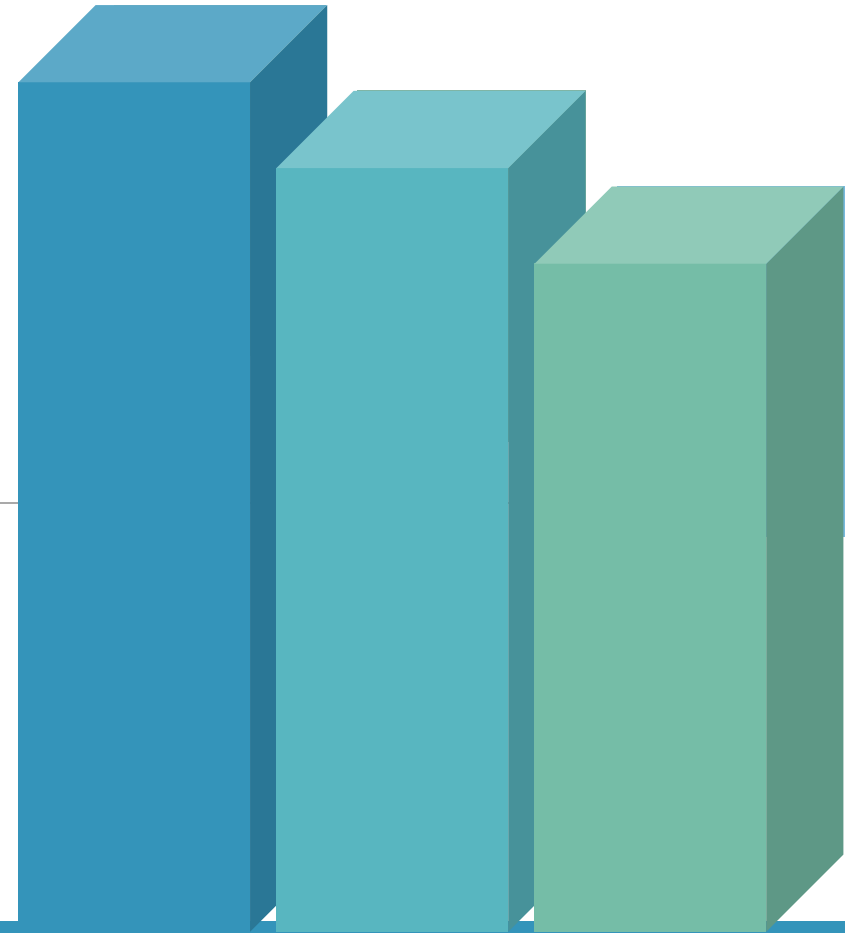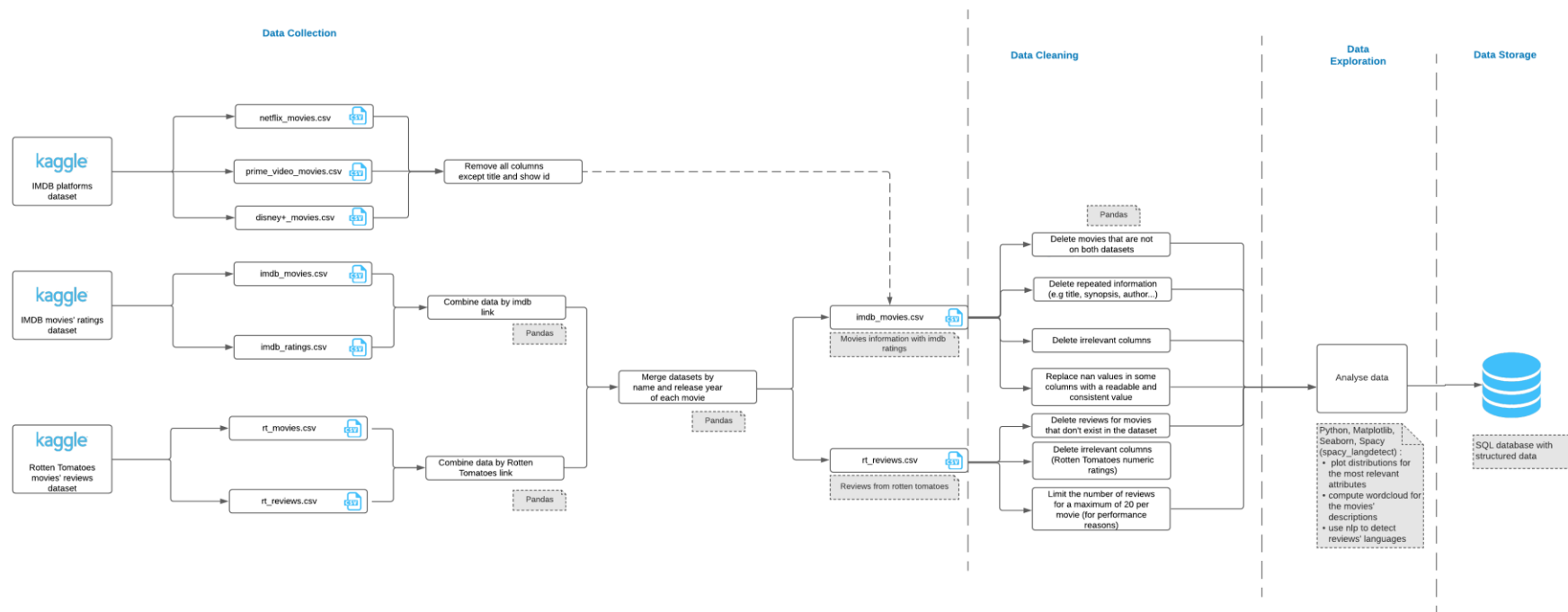## Movies Reviews and Ratings

CAIO NOGUEIRA (UP201806218)

CARLOS LOUSADA (UP201806302)

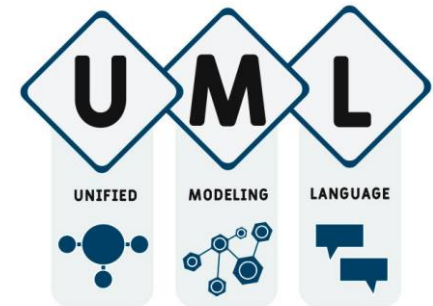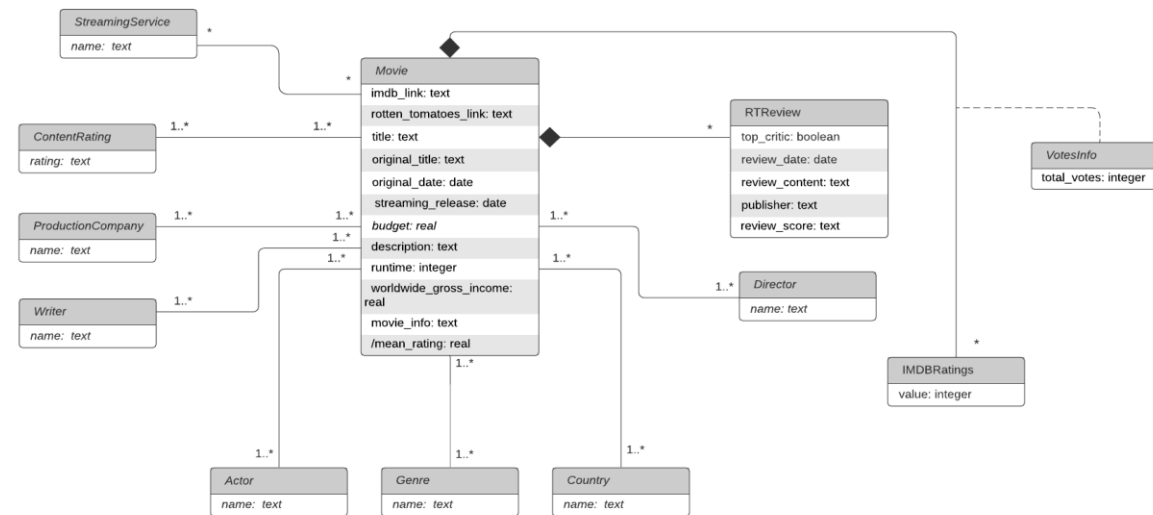TOMÁS MENDES (UP201806522)

Data Extraction and Preparation

Domain Conceptual Model
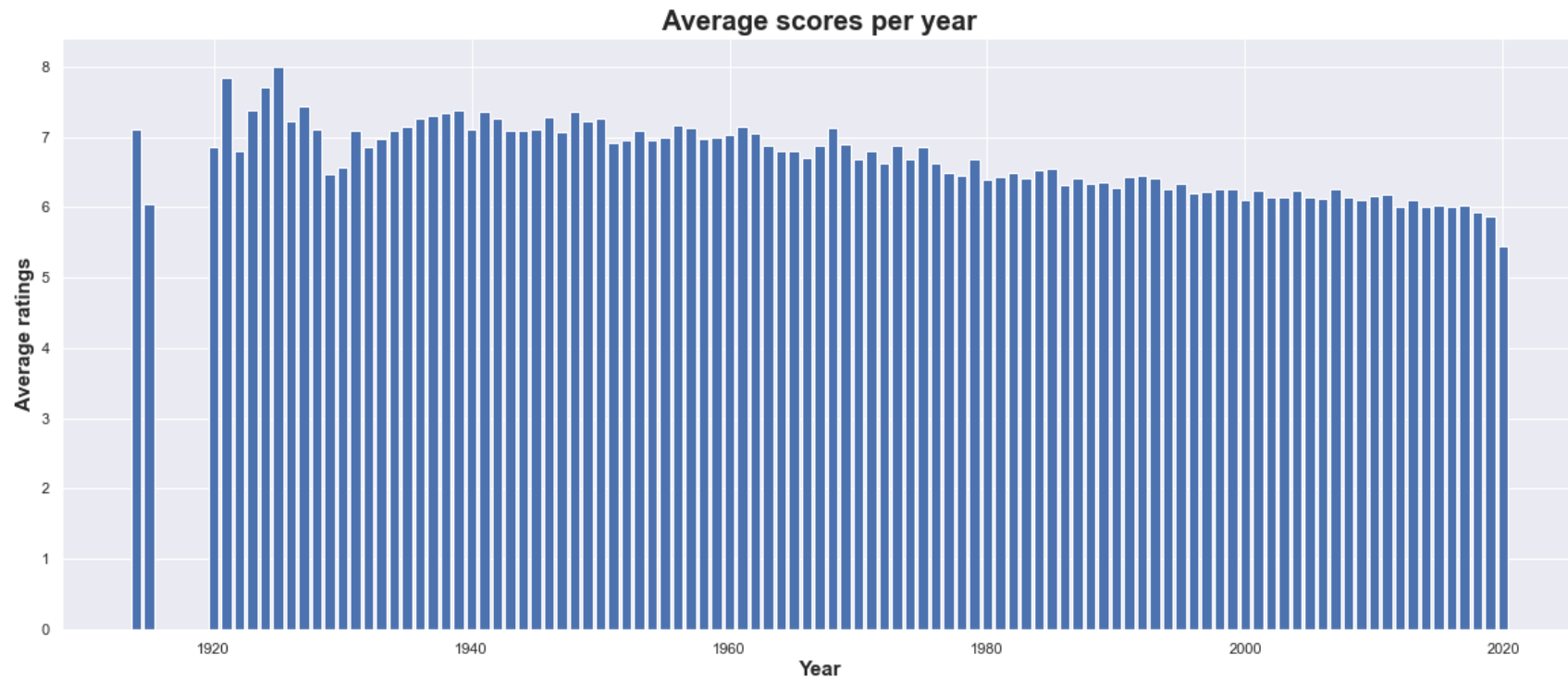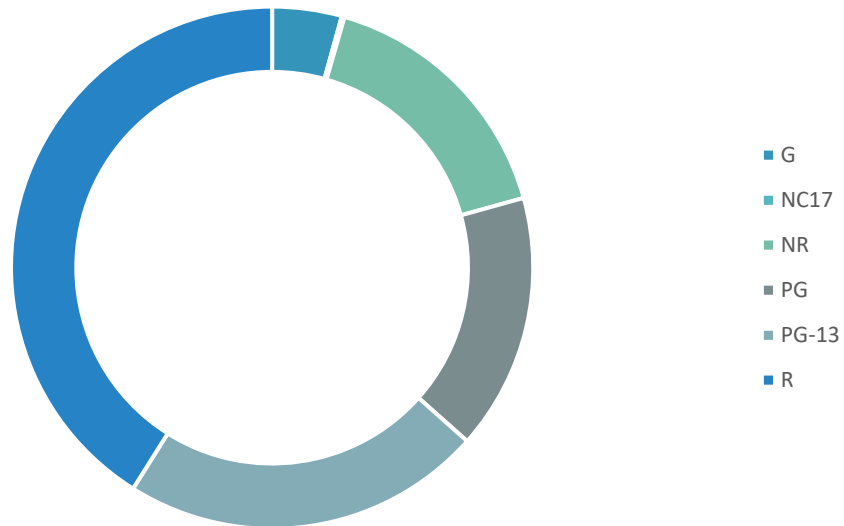
Data Characterization

Average scores per year
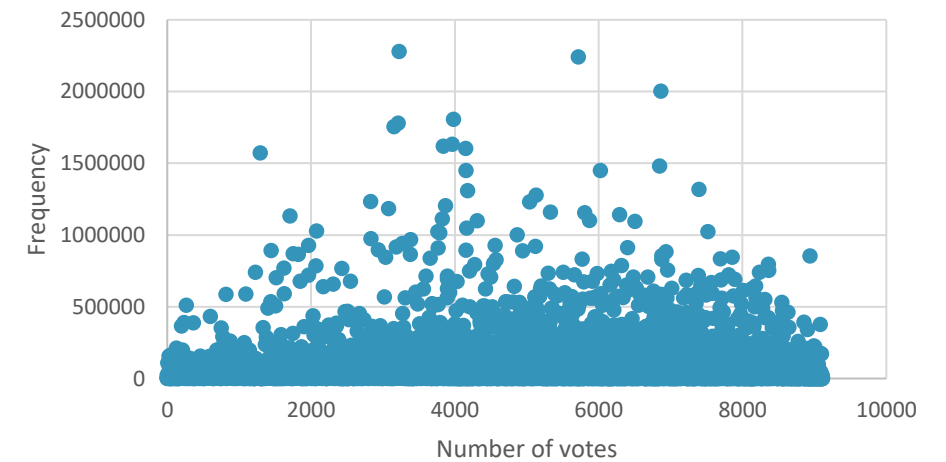
Data Characterization

Content Rating Distribution

- G
- NC17
- NR
- PG
- PG-13
- R

Total votes distribution

Data Characterization

Distribution of Movie's release years

Platforms Distribution
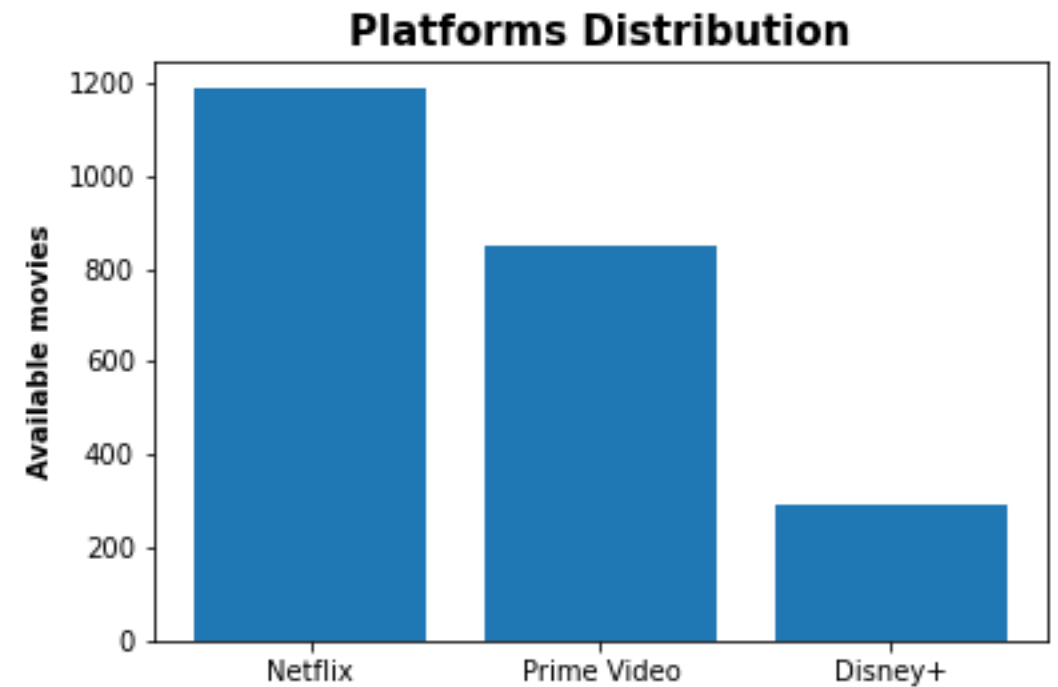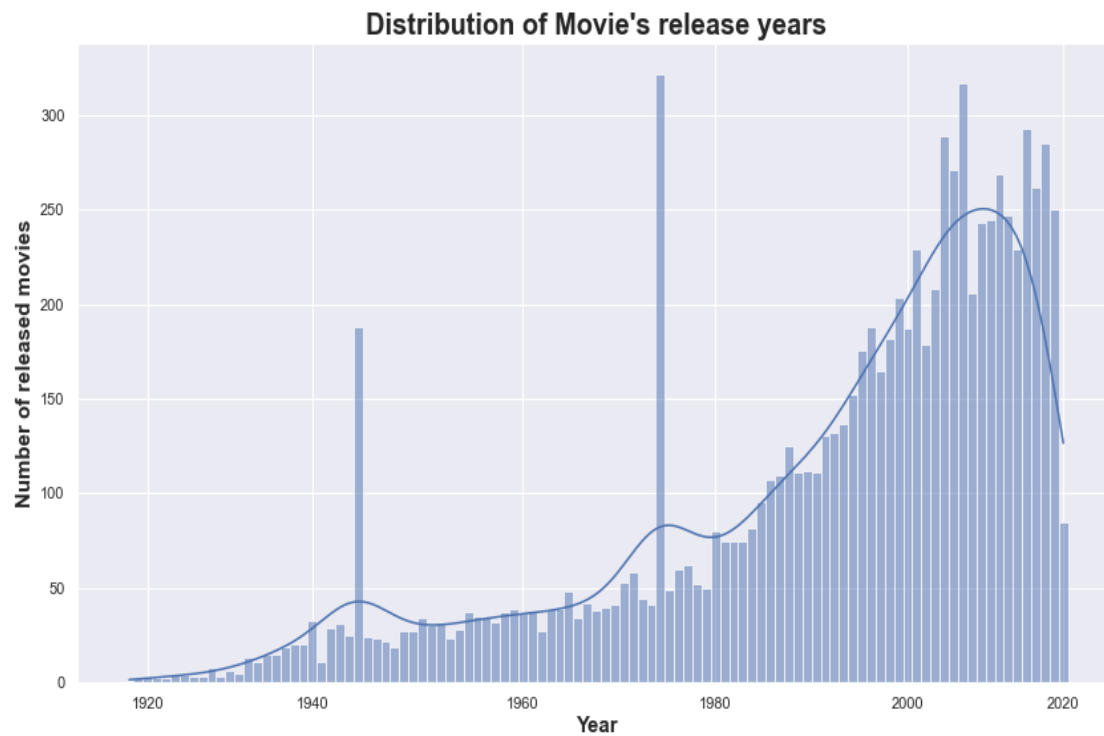
Conclusions and Future Work

Extraction and Preparation of the dataset:

➢Gathering data from different sources;

➢Cleaning the data;

➢Add information to the dataset;

➢Analyze the final dataset.

Future work:

Develop a **search engine** to obtain data from the dataset:
- Search most popular movies in a certain platform;
- Search for movies from a specific year;
- Search the most popular/highest rated movies from a certain genre;
- Search movies from a certain actor.

References

- Imdb movies extensive dataset:
  https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset

- Rotten Tomatoes movies and critic reviews dataset:
  https://www.kaggle.com/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset

- Netflix movies and TV shows:
  https://www.kaggle.com/shivamb/netflix-shows

- Amazon Prime Video movies and TV shows:
  https://www.kaggle.com/shivamb/amazon-prime-movies-and-tv-shows

- Disney Plus movies and TV shows:
  https://www.kaggle.com/shivamb/disney-movies-and-tv-shows