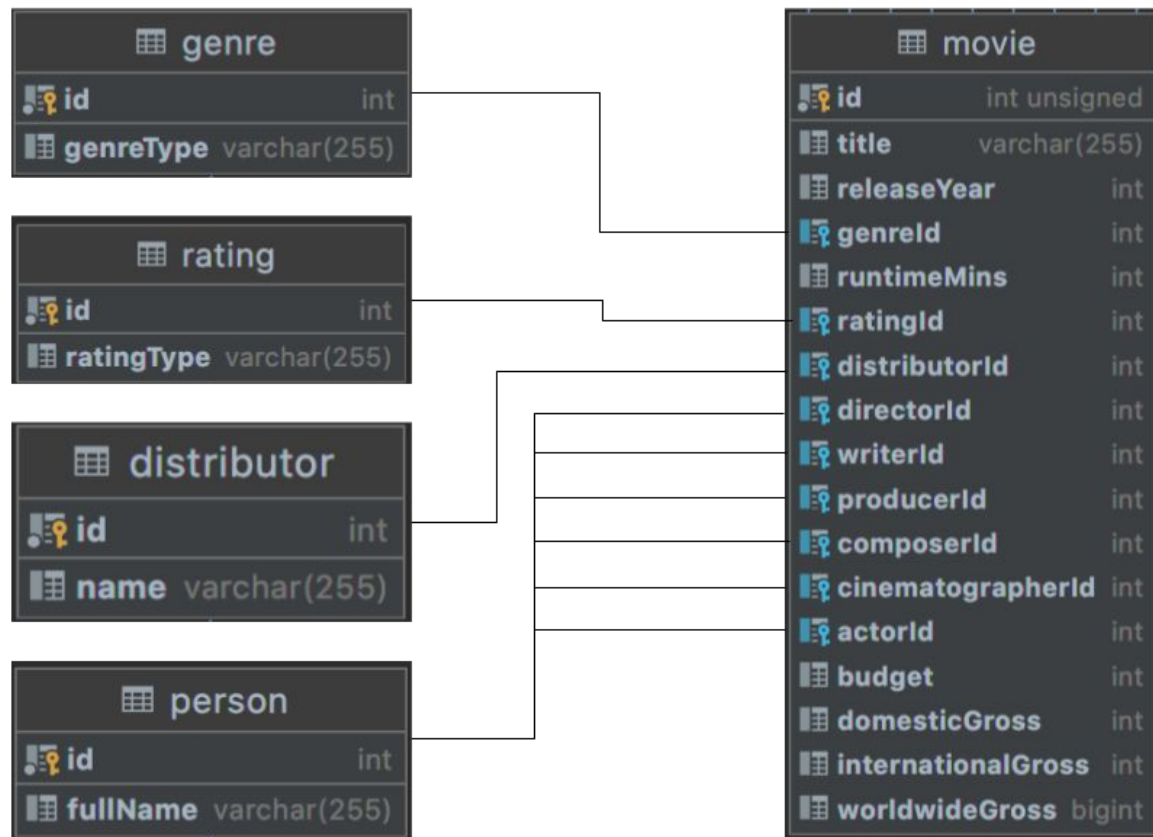# About the data

- Data from Kaggle: boxofficemojo dataset
  - Movies released from 1990 to 2020
  - Used 2018-2020 data

- I broke up the single table into 5 tables for my database:
  - movie
  - person
  - distributor
  - genre
  - rating

- The most interesting part about my data is that I have a 'person' table which connects to multiple columns in the 'movie' table

Sasha Kenkre
Uniqname: skenkre

ERD

Challenge inserting data

- 'id' column would not start from 1 when importing data for 'movie' table
  - Two queries helped:
    - truncate table movie;
    - alter table movie AUTO_INCREMENT=1;

Favorite question from letter & what I'd do differently

- What are the top 3 genres and how many films have been released for those 3 genres?

- Solution:
  SELECT g.genreType, count(1)
  FROM genre g
      JOIN movie m ON g.id = m.genreId
  GROUP BY m.genreId
  ORDER BY count(1) DESC
  LIMIT 3;

	genreType	`count(1)`
1	Action	81
2	Comedy	38
3	Adventure	32

- What I'd do differently:
  - I would see if there was a way to create a table for the 'roles' or combine that with the 'person' table in a way that made sense with the rest of the data.