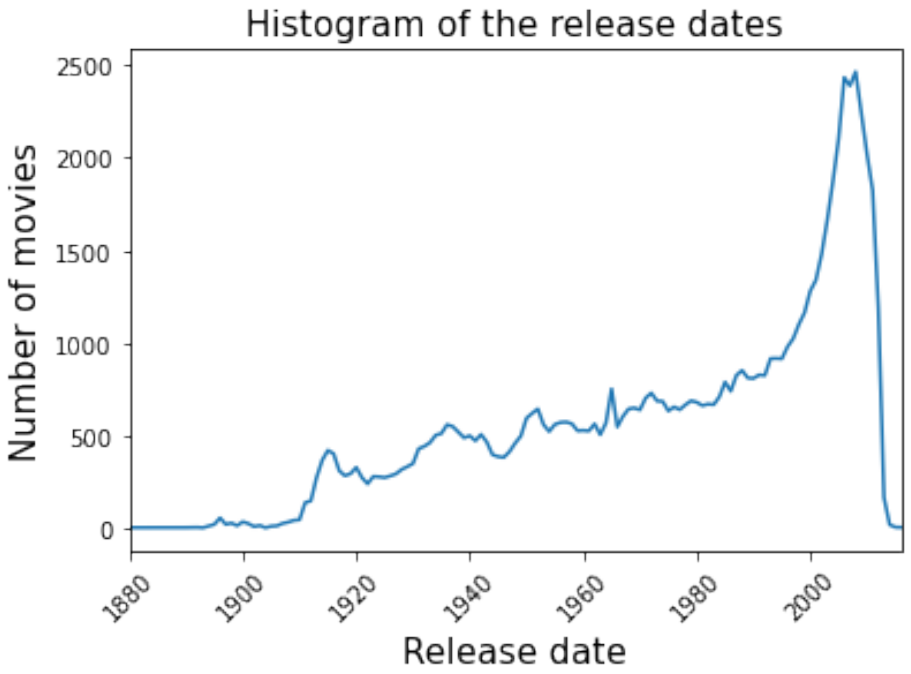
I/ Temporal analysis

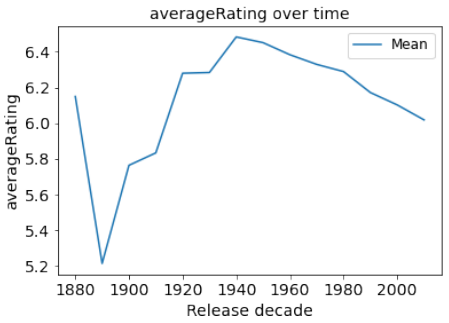
1. Exploratory data analysis

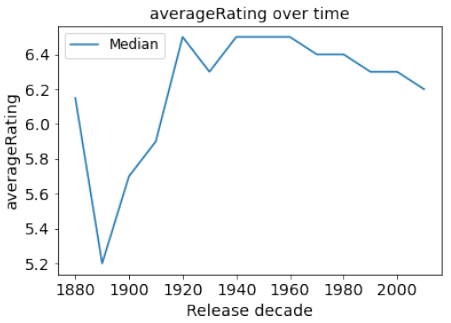
Distribution of the release dates

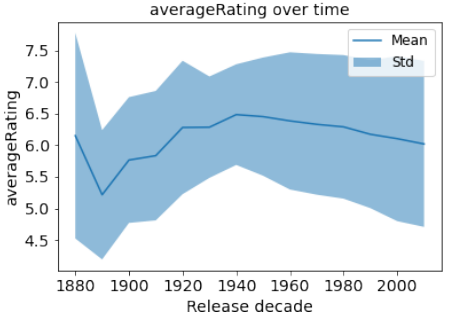


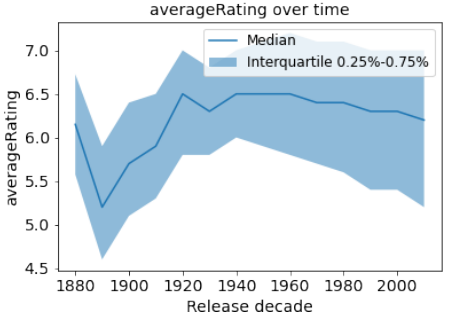
- From the EDA of milestone 2.
- “We can observe the drastic increase in the amount of movies produced in the past 40 years. This is most likely caused by the technological advancements which enabled the movie industry to become ubiquitous. Due to this, it is expected that results from this time period will be a better representation of the real world when conducting a temporal analysis.”

Evolution of variables over time: average rating





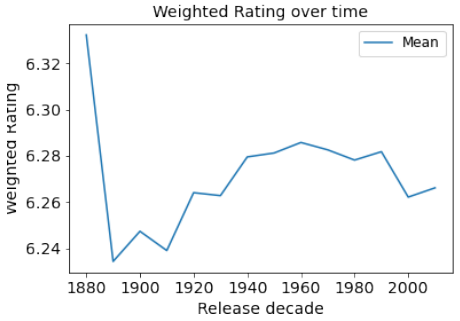


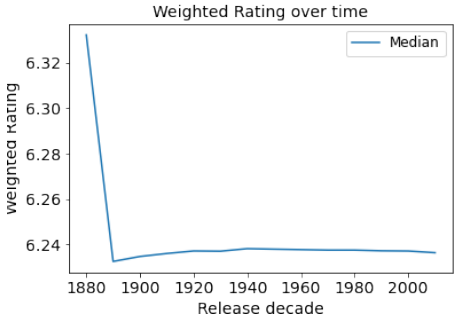


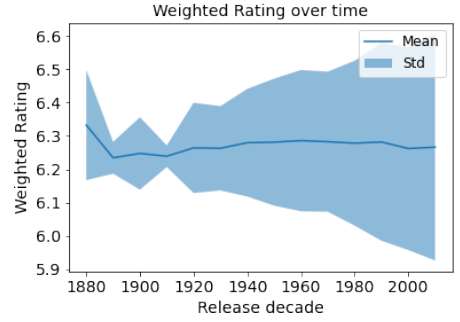
- The average rating has decreased in the recent years.
- This can be due to people becoming more critical as the amount of released movies is increasing, so there are more movies to compare to.

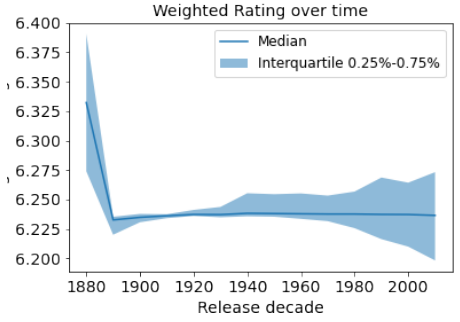
  Or only the most famous and appreciated movies are referenced for the first decades, while now we have statistics about almost every existing movies.

Evolution of variables over time: weighted rating

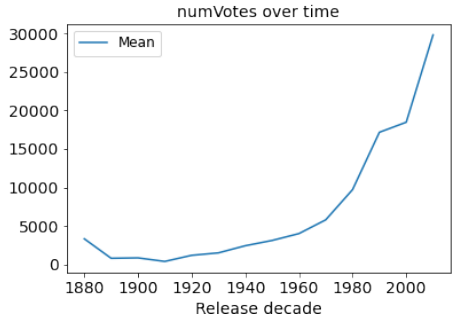


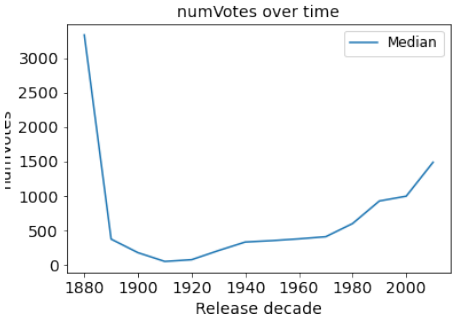


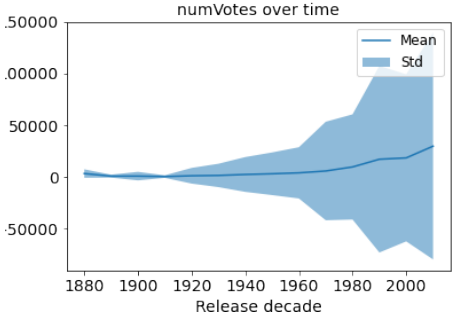


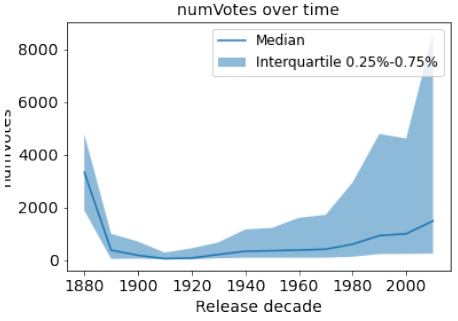


Evolution of variables over time: number of votes



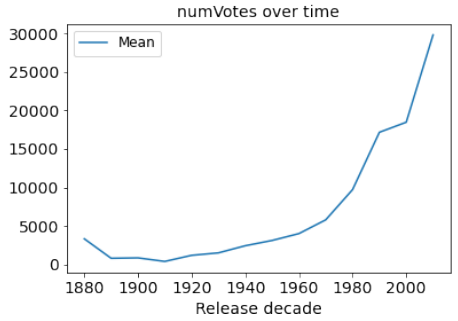


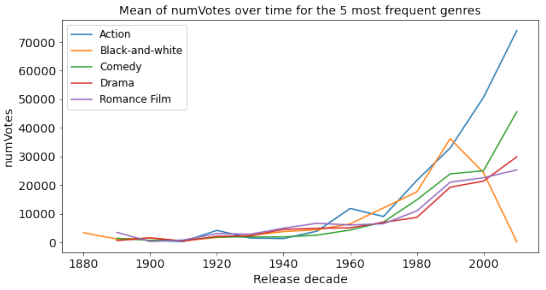


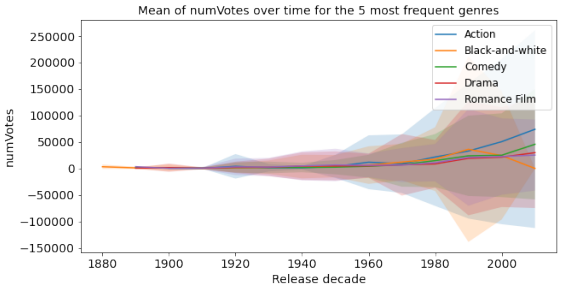


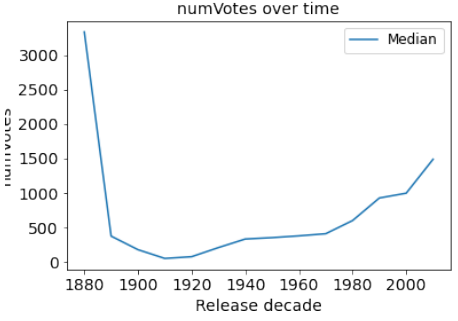
- The number of votes is increasing.
- This can be due to the creation of internet that facilitates voting.

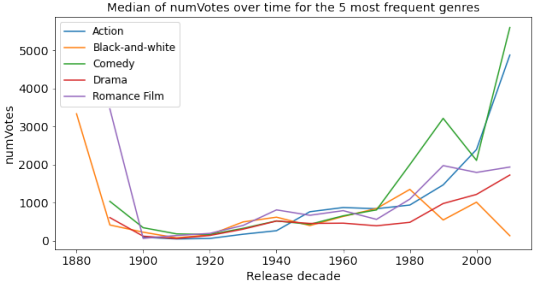
Evolution of variables over time, depending on the genre: number of votes

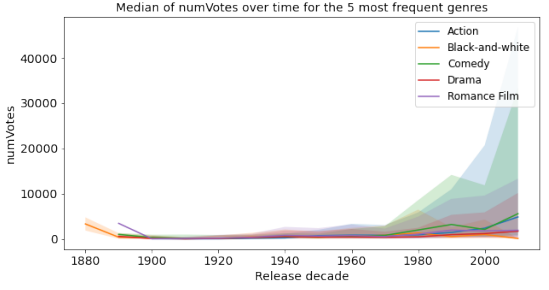




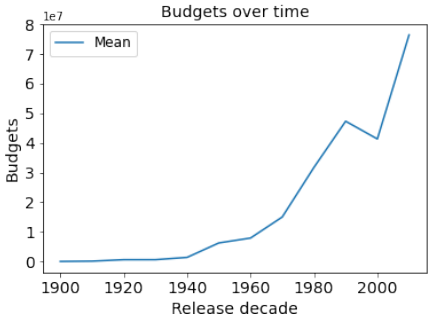


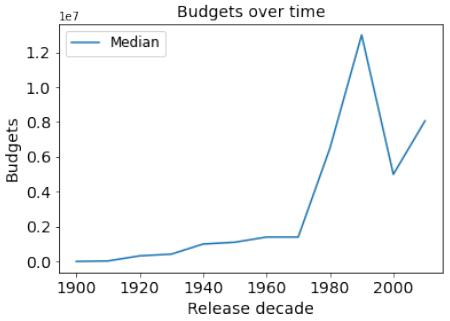


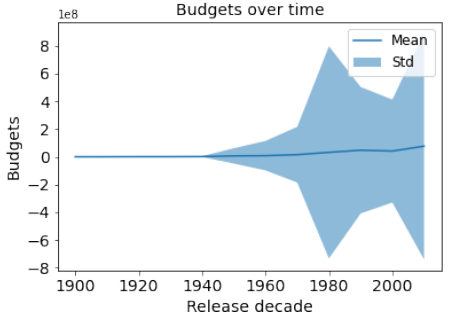


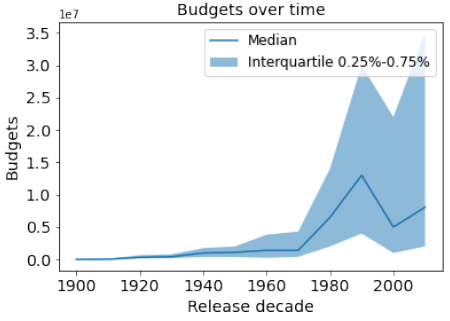


Evolution of variables over time: budget

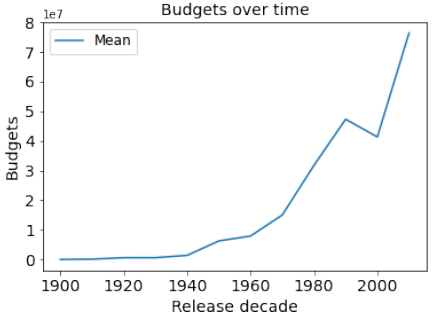


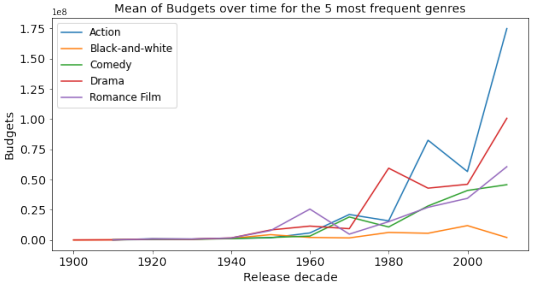


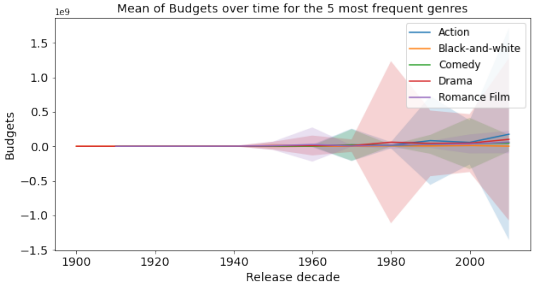


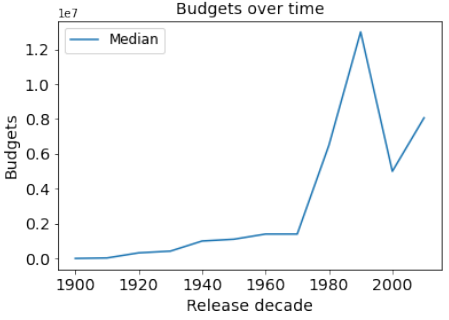


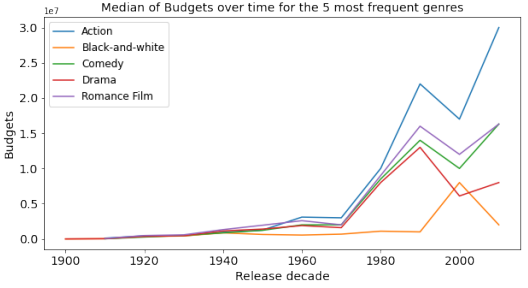
Evolution of variables over time, depending on the genre: budget

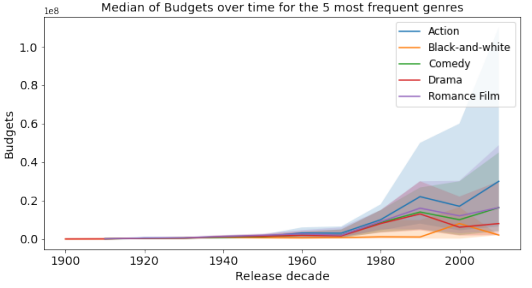




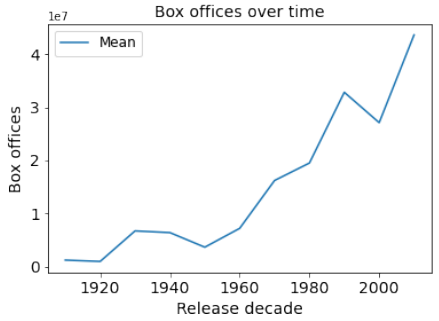


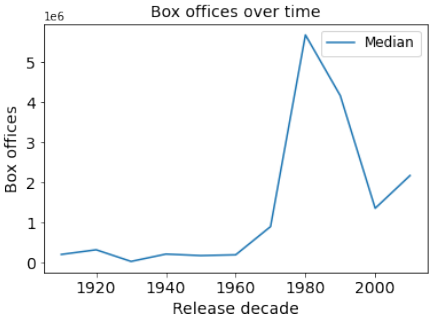


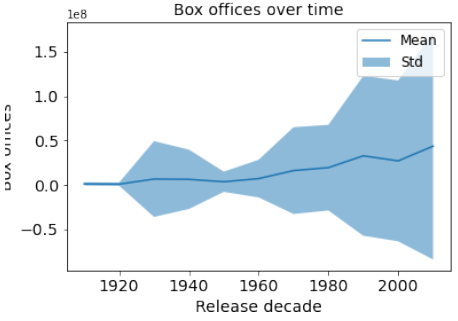


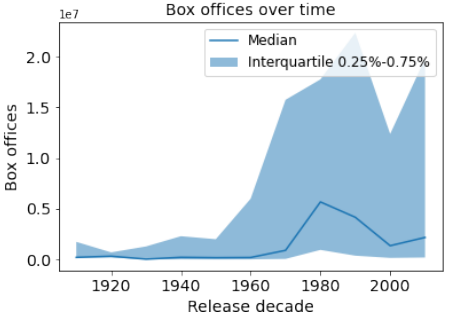


Evolution of variables over time: box office



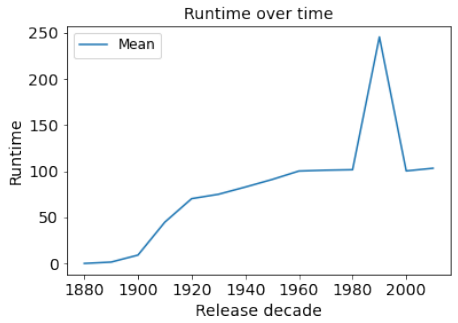


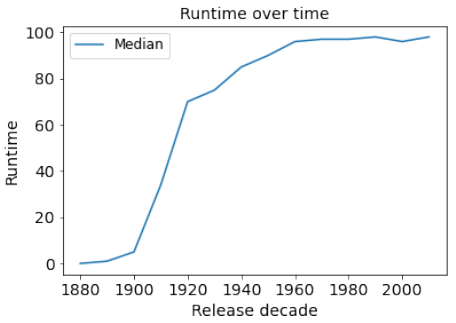


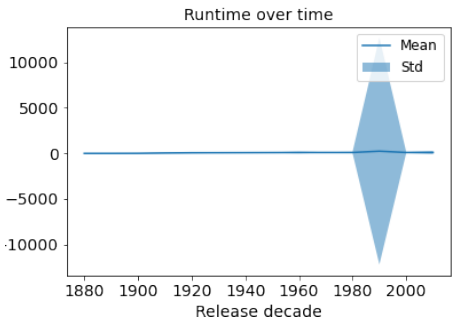


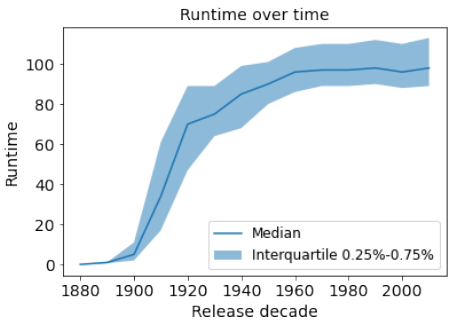
- The box offices are increasing, but the budgets as well.

Evolution of variables over time: runtime









- When movies started to be released, they were shorter on average.
- Nowadays, the runtime stabilized around 100 minutes.

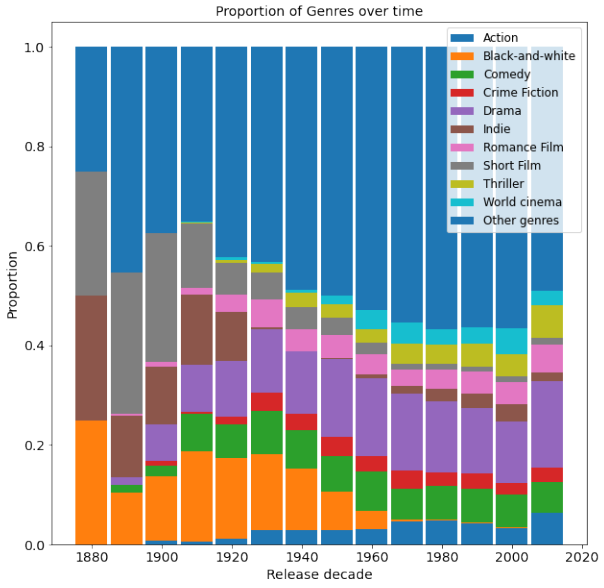
Evolution of variables over time (cont.)

- We first analyzed how the average of the variables evolved over time. We used the standard deviation as the error bar.

  But the resulting error bars were huge, making the trends not visible.

  This could be due to standard deviation not being robust to outliers/extreme values.
- For this reason, we decided to use the median, with the interquartile range between 0.25% and 0.75% as the error bar.

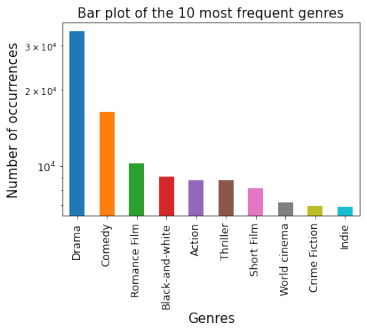
Evolution of genre distribution over time

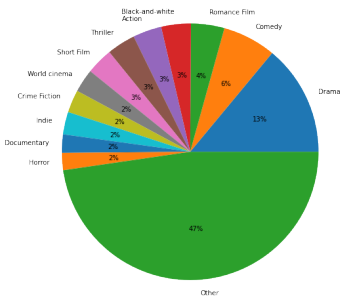


- Black-and-white movies are less frequent.
- Action movies are more frequent.
- Since 1920, drama is the most frequent genre.

2. Evolution over time, of the important features for success

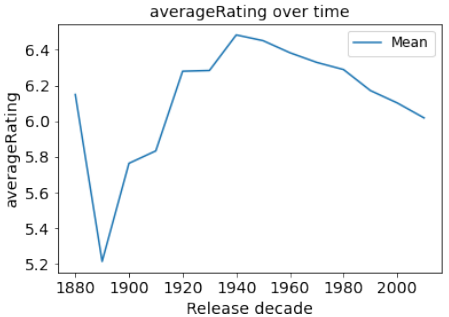
5 most frequent genres in the dataset

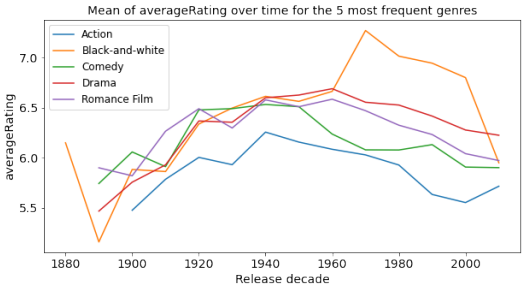


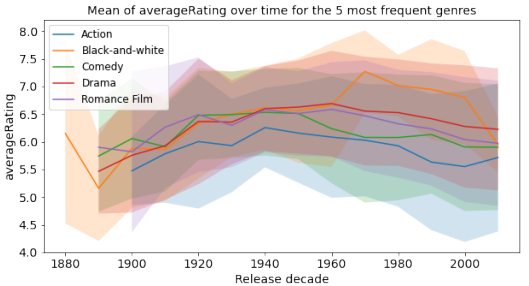


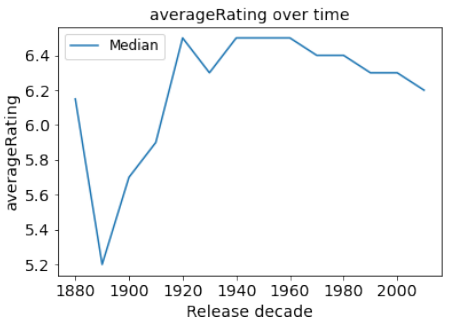
- From the EDA of milestone 2.
- The 5 most frequent genres are:
  - Drama
  - Comedy
  - Romance
  - Black and white
  - Action

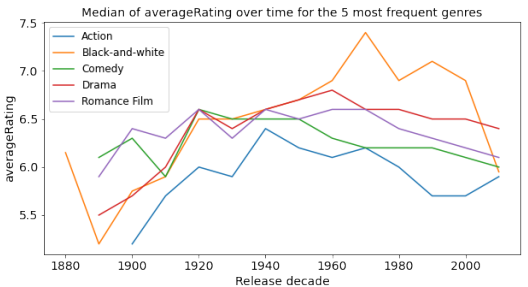
Evolution of variables over time, depending on the genre: average rating

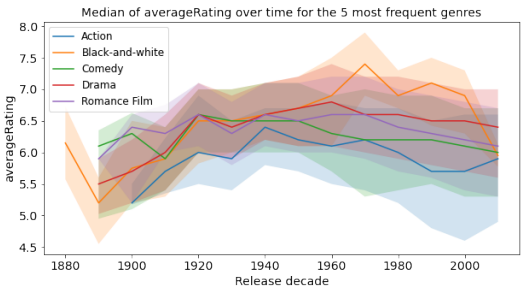








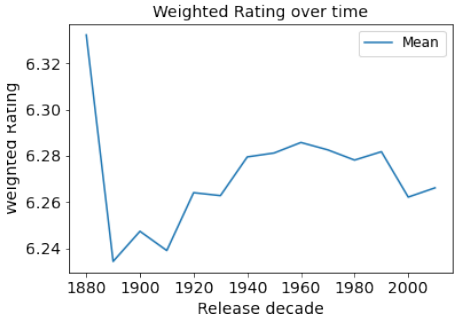


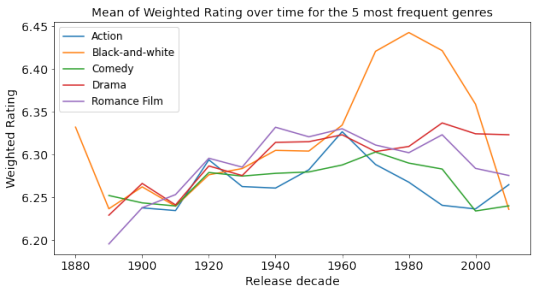


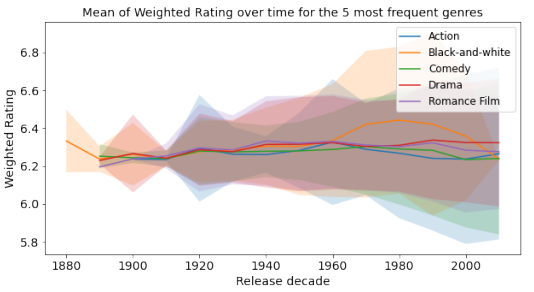
Evolution of variables over time, depending on the genre: average rating (cont.)

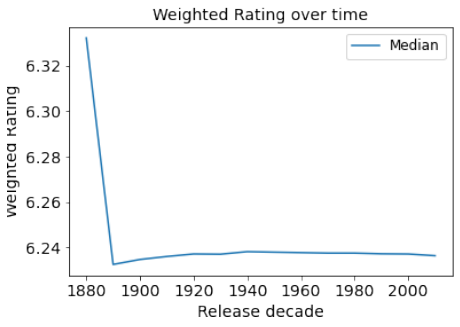
- We saw that the overall ratings are changing over time.
- The ratings for the 5 most frequent genres are following the general trend. But we can still observe some particular behaviors:
  - Black-and-white movies were rated higher around 1970-1980, but they are receiving lower ratings nowadays.
  - Action movies are rated higher nowadays.
- But with the error bar, can we still say something?

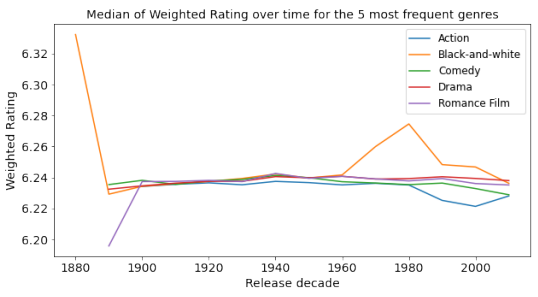
Evolution of variables over time, depending on the genre: weighted rating

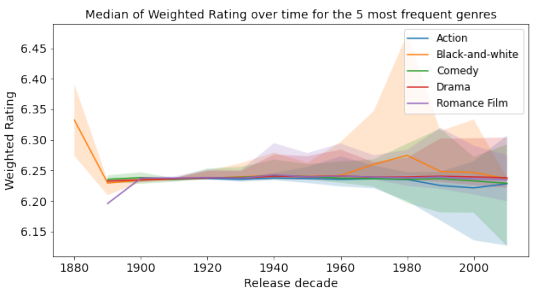








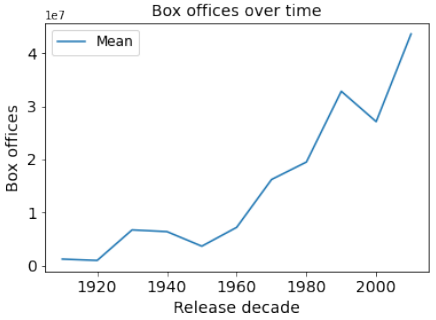


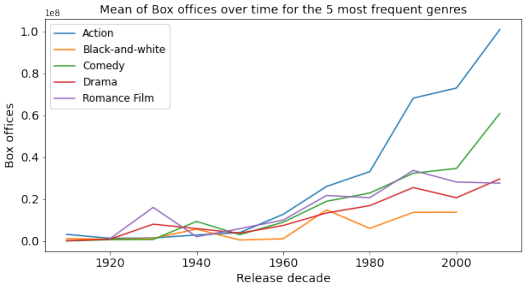


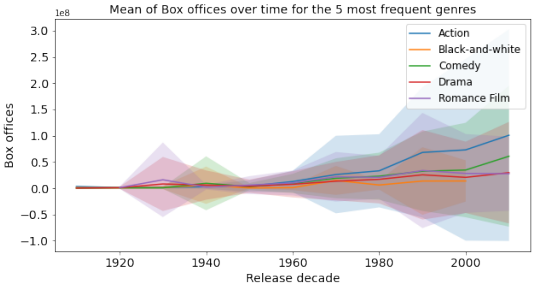
Evolution of variables over time, depending on the genre: weighted rating (cont.)

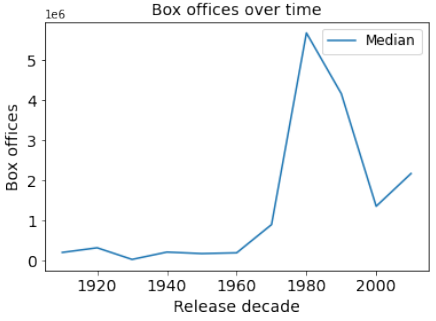
- Same observations as the one made with average rating.

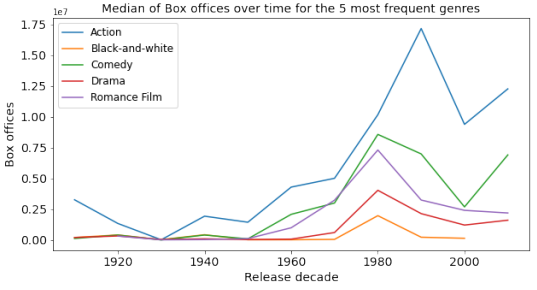
Evolution of variables over time, depending on the genre: box office

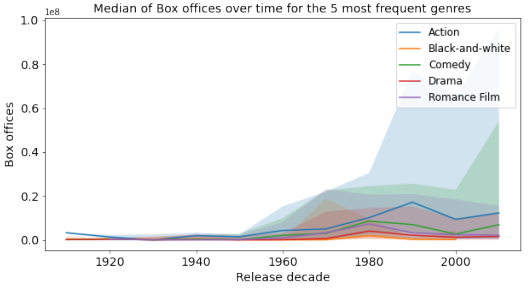




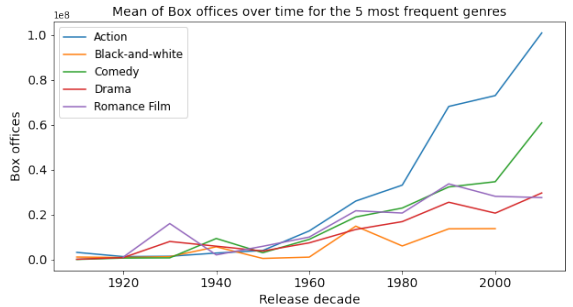


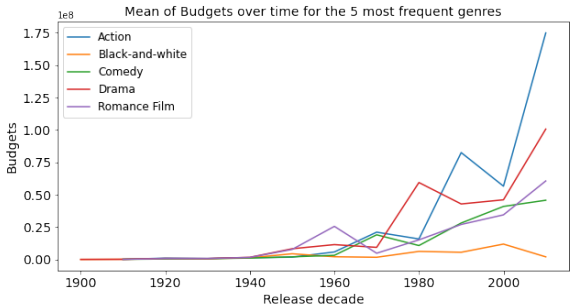






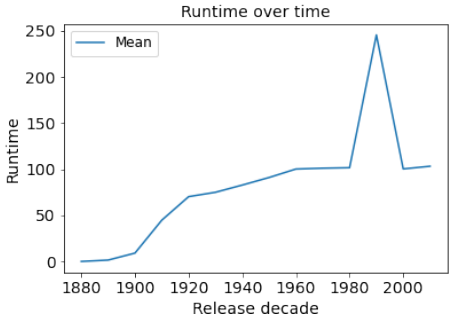
Evolution of variables over time, depending on the genre: box office and budget

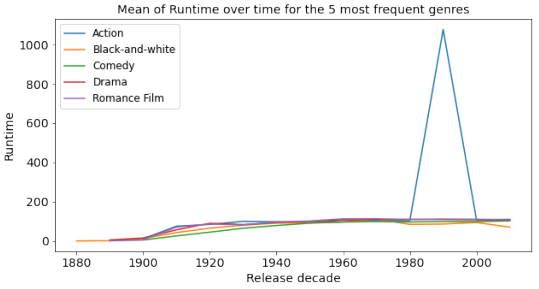


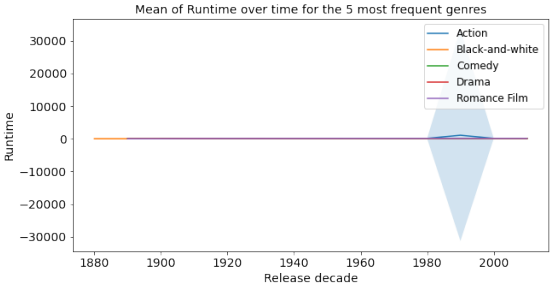


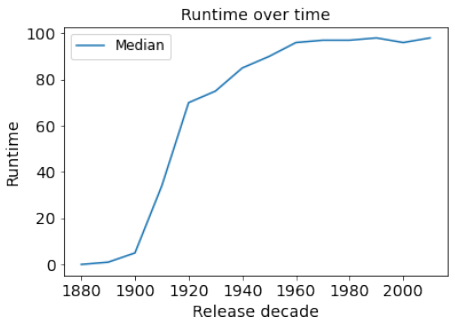
- Action movies are becoming more profitable, but the allocated budget is also higher.
- On the contrary, the profit for drama seems to have increased less than the increased of the allocated budget.

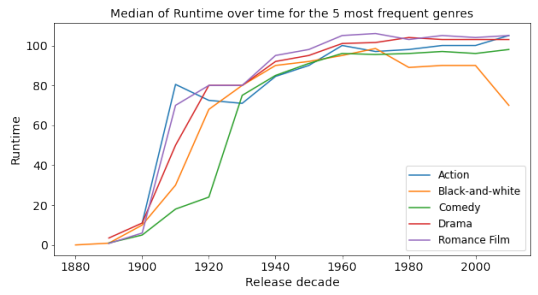
Evolution of variables over time, depending on the genre: runtime

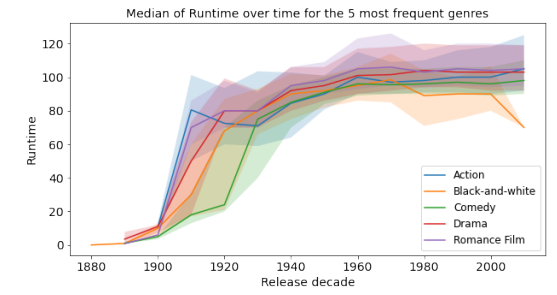




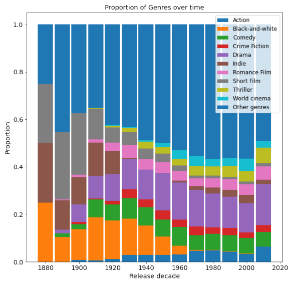




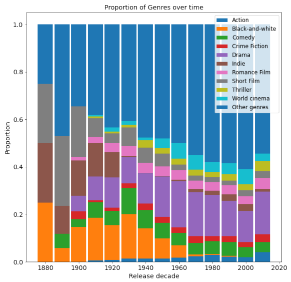




Comparison between all movies and successful movies: genre distribution

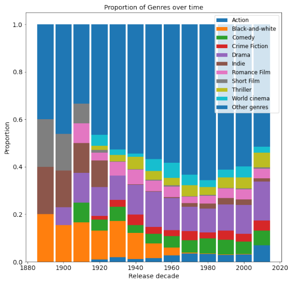


- Complete dataset



- Movies with average rating > 7

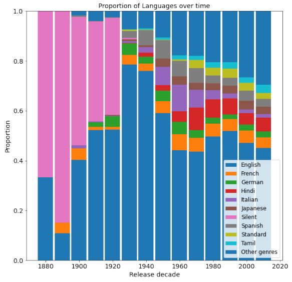
  (16275 movies)

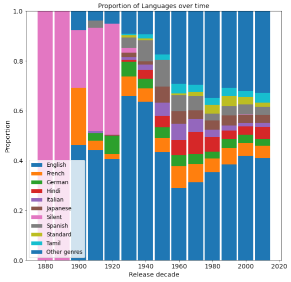


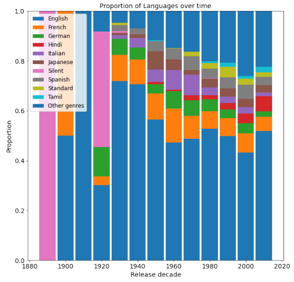
- Movies with weighted rating > 6.5

  (4391 movies)

Comparison between all movies and successful movies: language distribution







- Complete dataset

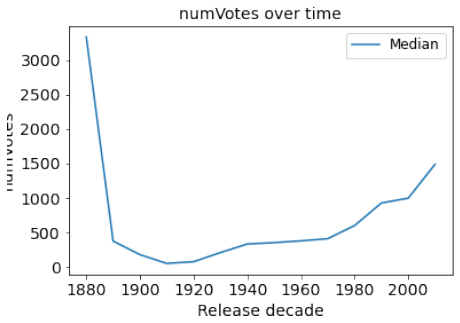
- Movies with average rating > 7

  (16275 movies)

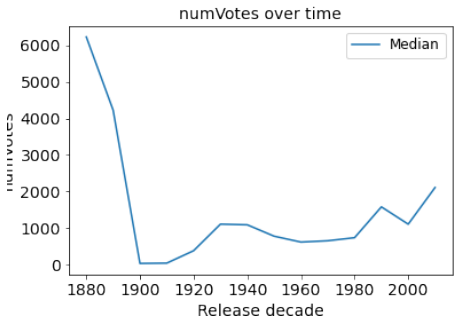
- Movies with weighted rating > 6.5

  (4391 movies)

Comparison between all movies and successful movies: number of votes

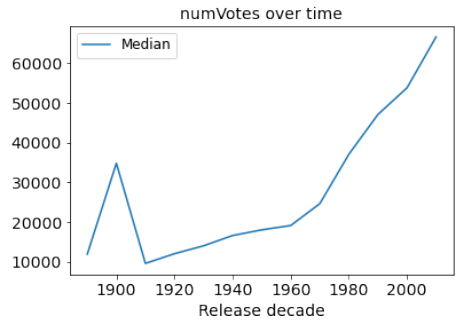


- Complete dataset



- Movies with average rating > 7

  (16275 movies)



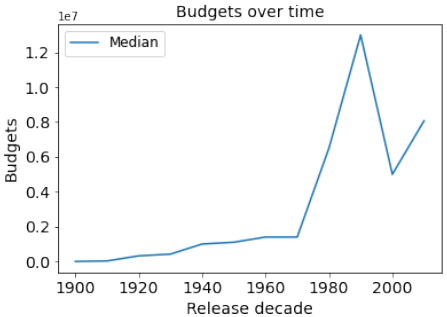
- Movies with weighted rating > 6.5

  (4391 movies)

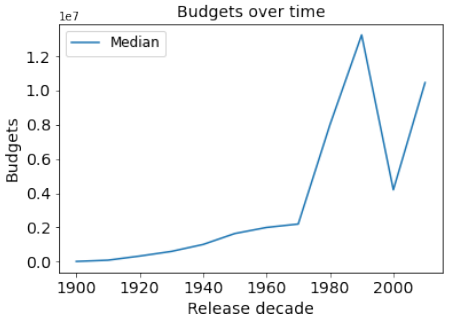
Comparison between all movies and successful movies: number of votes (cont.)

- The number of votes for successful movies is increasing a lot more than in average.

Comparison between all movies and successful movies: budget

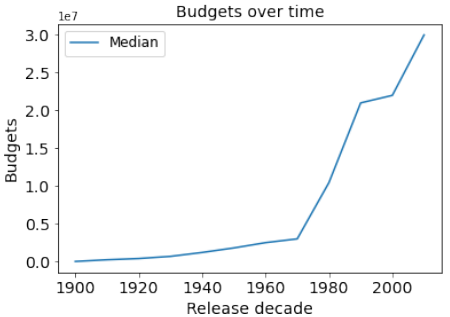


- Complete dataset



- Movies with average rating > 7

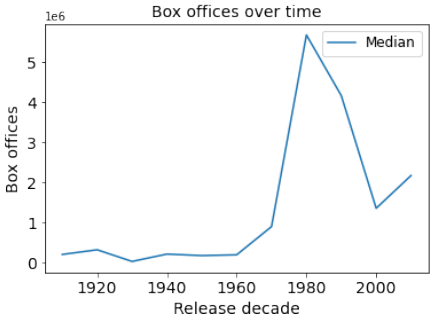
  (16275 movies)

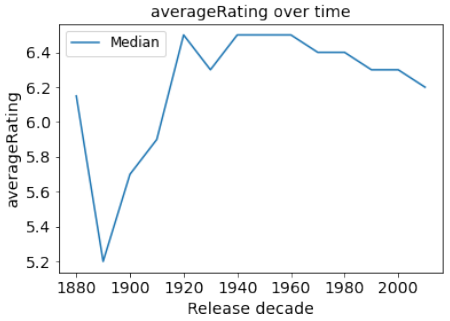


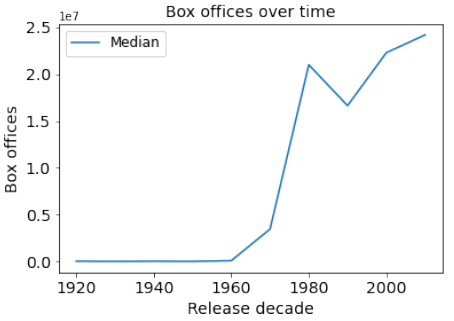
- Movies with weighted rating > 6.5

  (4391 movies)

Comparison between all movies and successful movies: box office





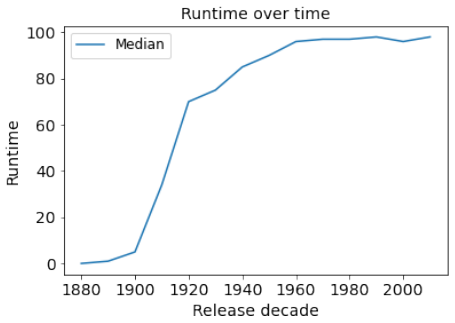


- Complete dataset
- Movies with average rating > 7

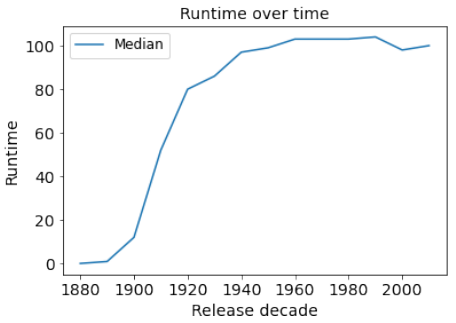
  (16275 movies)
- Movies with weighted rating > 6.5

  (4391 movies)

Comparison between all movies and successful movies: runtime

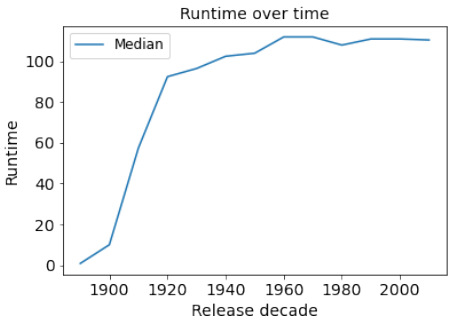


- Complete dataset



- Movies with average rating > 7

  (16275 movies)



- Movies with weighted rating > 6.5

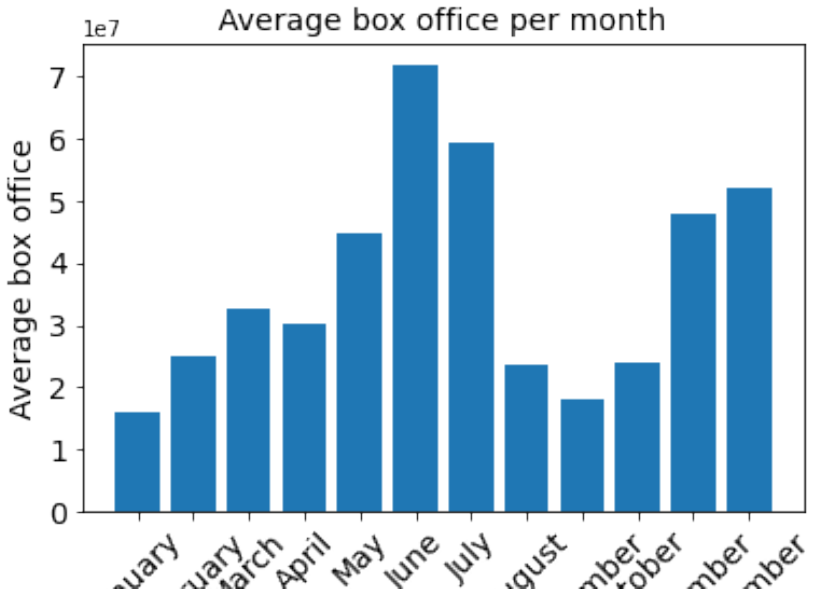
  (4391 movies)

Comparison between all movies and successful movies: runtime (cont.)

- It seems that the runtime of successful movies is a little bit over the average runtime for the movies in general (a bit over 100min).

3. Distribution over the months

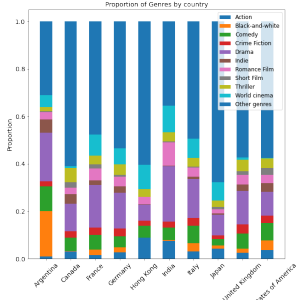
Average box office depending on the release month



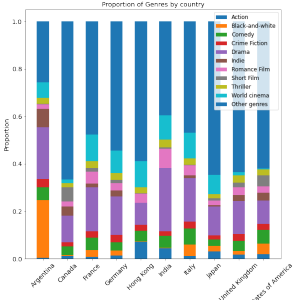
- In average, the box office is the highest for movies released in June and July. It can be due to people going to cinema more often during the summer holidays.

II/ Spatial analysis

Comparison between all movies and successful movies: genre distribution

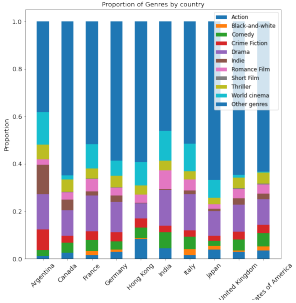


- Complete dataset



- Movies with average rating > 7

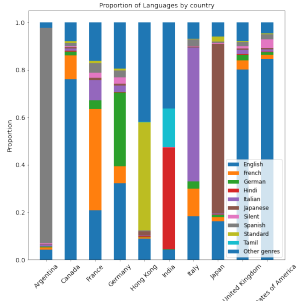
  (16275 movies)



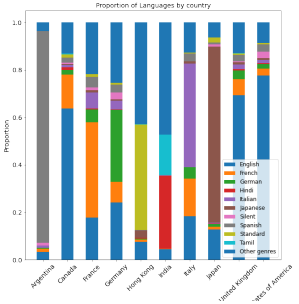
- Movies with weighted rating > 6.5

  (4391 movies)

Comparison between all movies and successful movies: language distribution

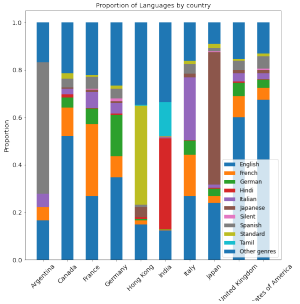


- Complete dataset



- Movies with average rating > 7

  (16275 movies)



- Movies with weighted rating > 6.5

  (4391 movies)

Comparison between all movies and successful movies: language distribution (cont.)

- More successful movies include English then the average.