



IMDb data preparation

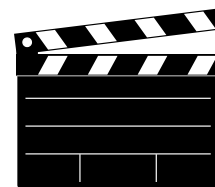
Original Datasets

Downloaded from IMDb datasets



title.akas.tsv.gz

33M entries
8 features



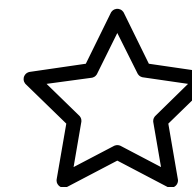
title.basics.tsv.gz

9M entries
9 features



title.crew.tsv.gz

9M entries
3 features



title.ratings.tsv.gz

1M entries
3 features



name.basics.tsv.gz

11M entries
6 features

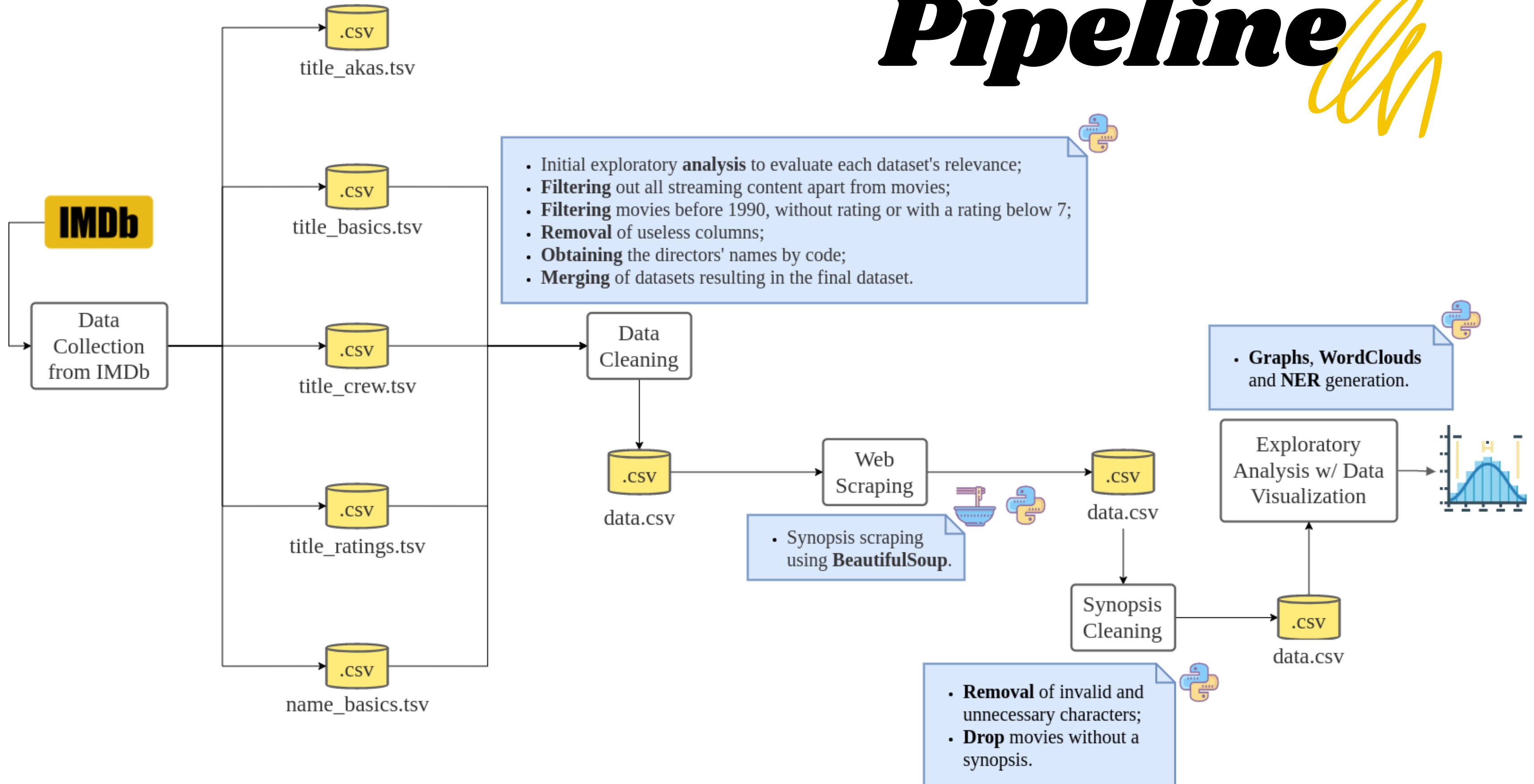
Final Dataset

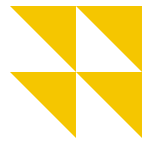


| tconst | primaryTitle | originalTitle | isAdult | year |
|-----------|---------------------------|---------------------------|---------|------|
| tt0038086 | Shiva und die Galgenblume | Shiva und die Galgenblume | 0 | 1993 |

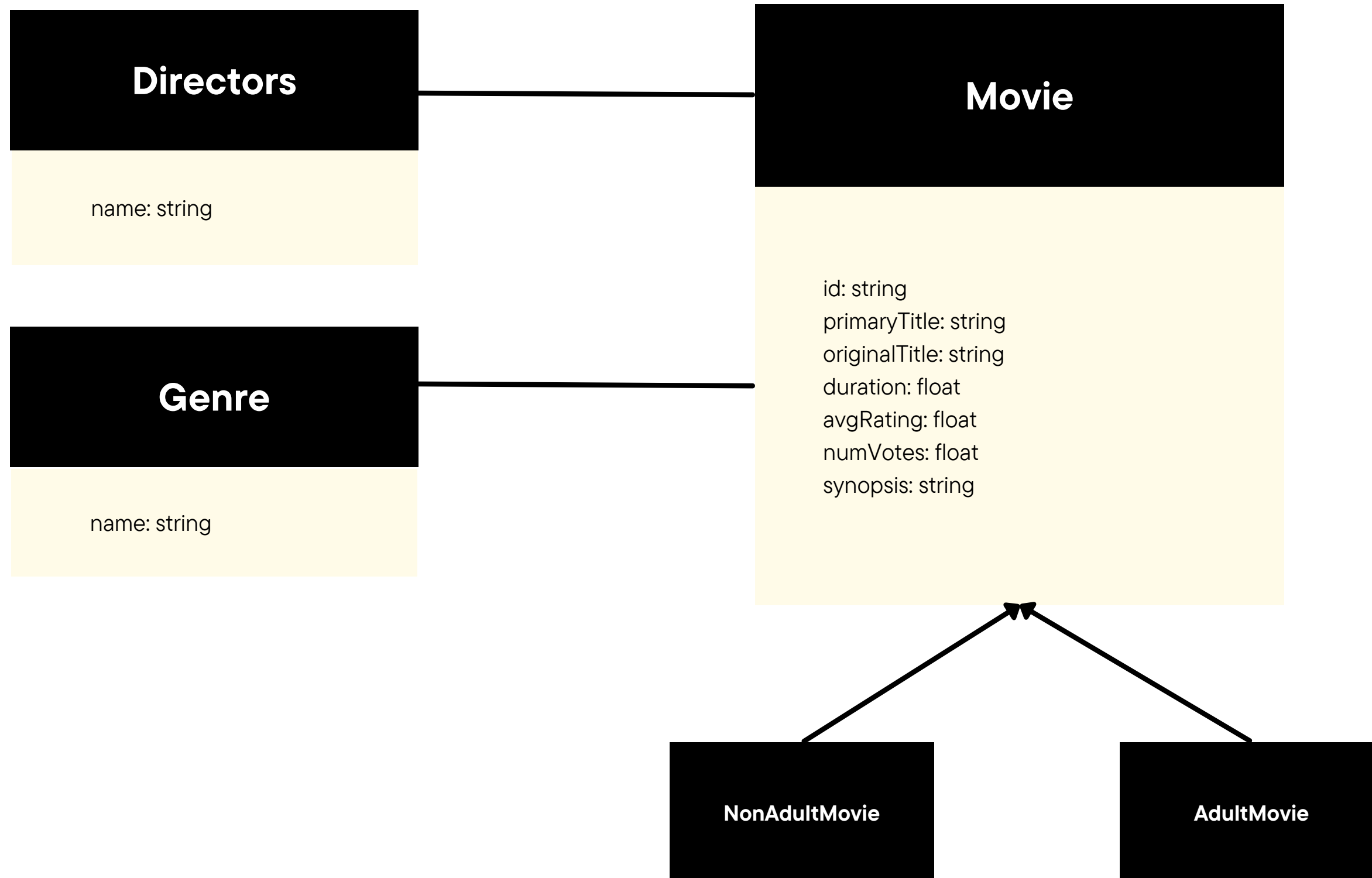
| duration | genres | avgRating | directors | synopsis |
|----------|-----------------|-----------|-------------------------------|--|
| 140 | Thriller, Drama | 7.4 | Hans Steinhoff, Tim Savage | Detective Dongen, known as Shiva, exposes a Russian artist... |

Pipeline



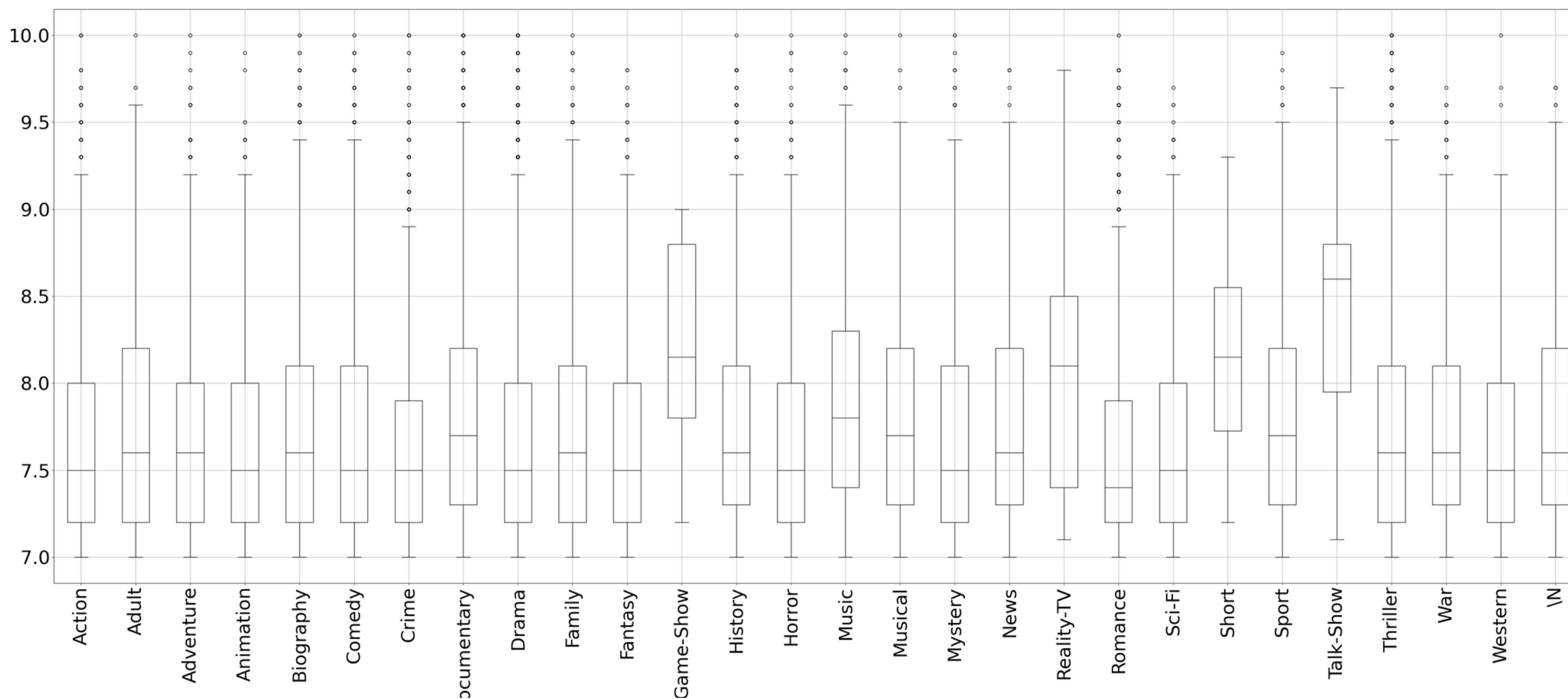


Conceptual Model



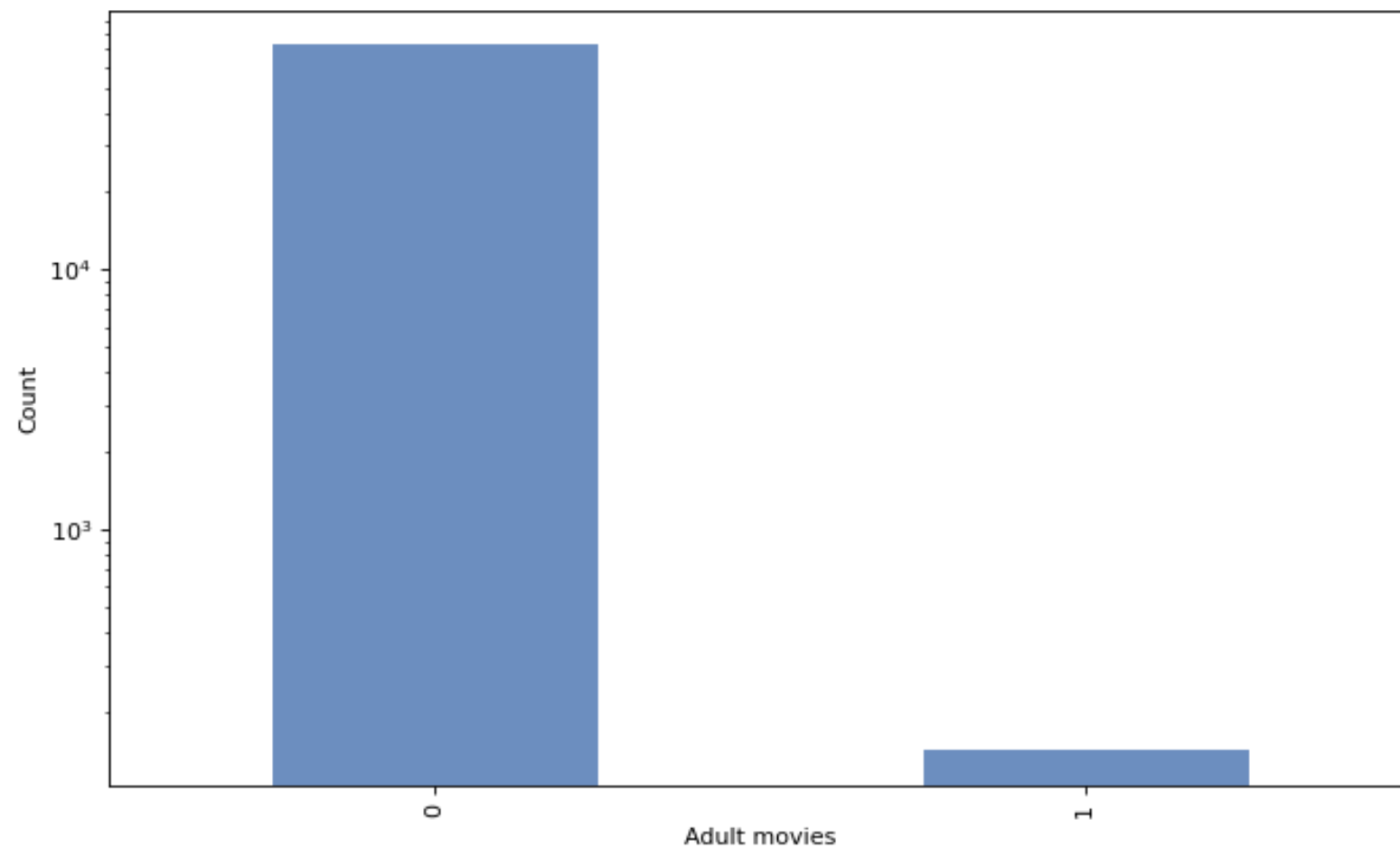
Datasets Characterisation

Boxplot per movie genre

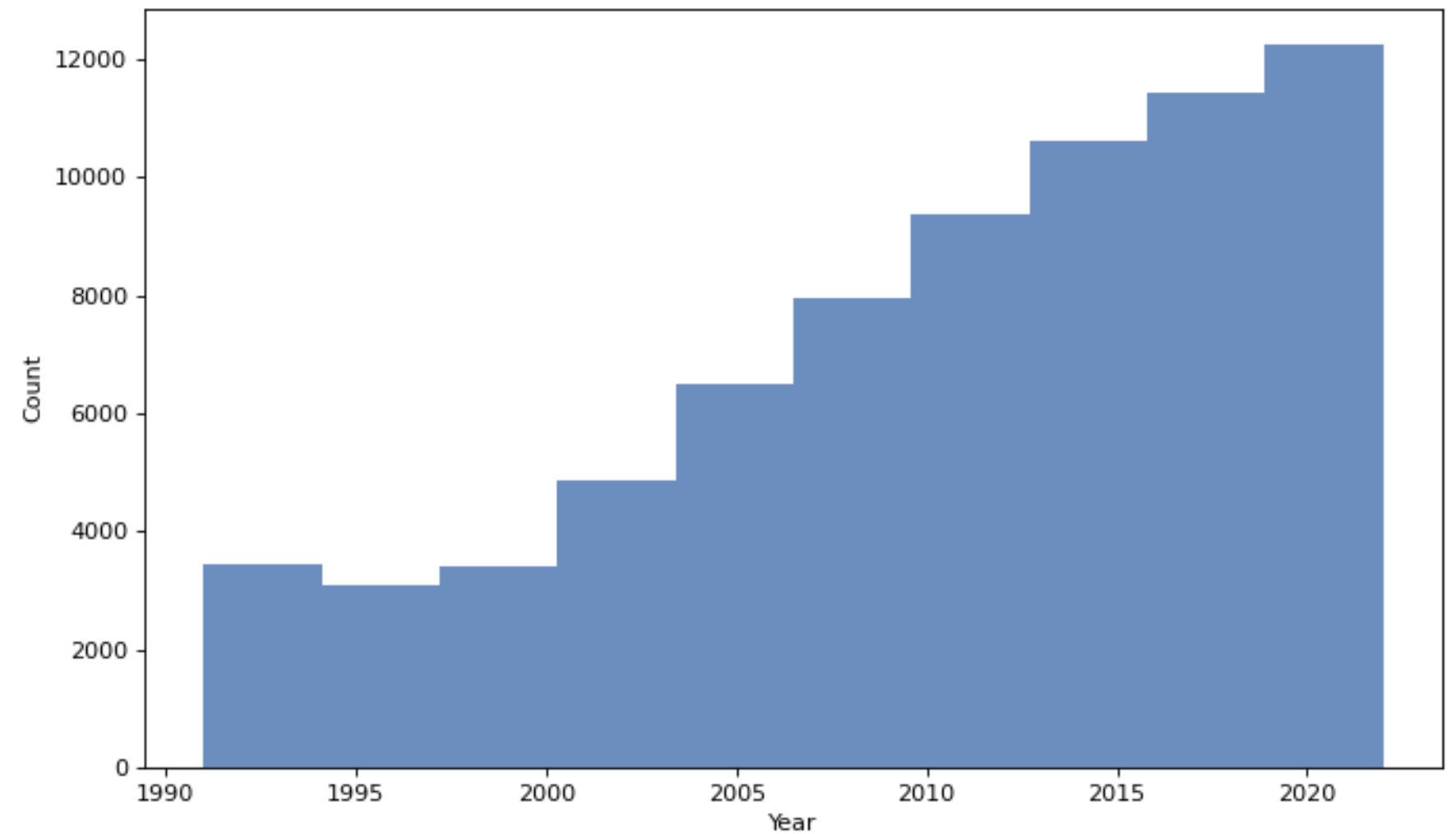


Datasets Characterisation

Number of adult movies

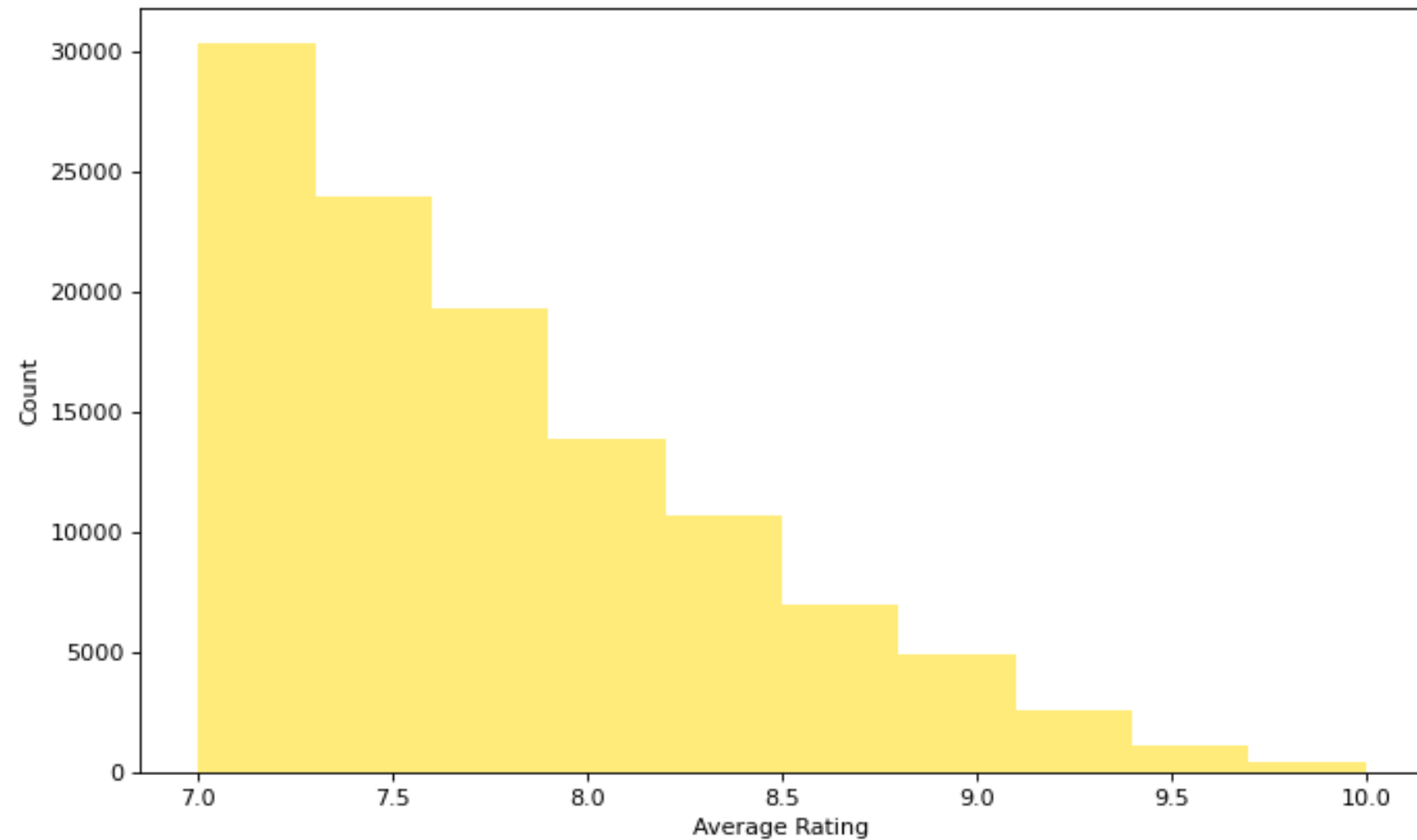


Number of movies per year

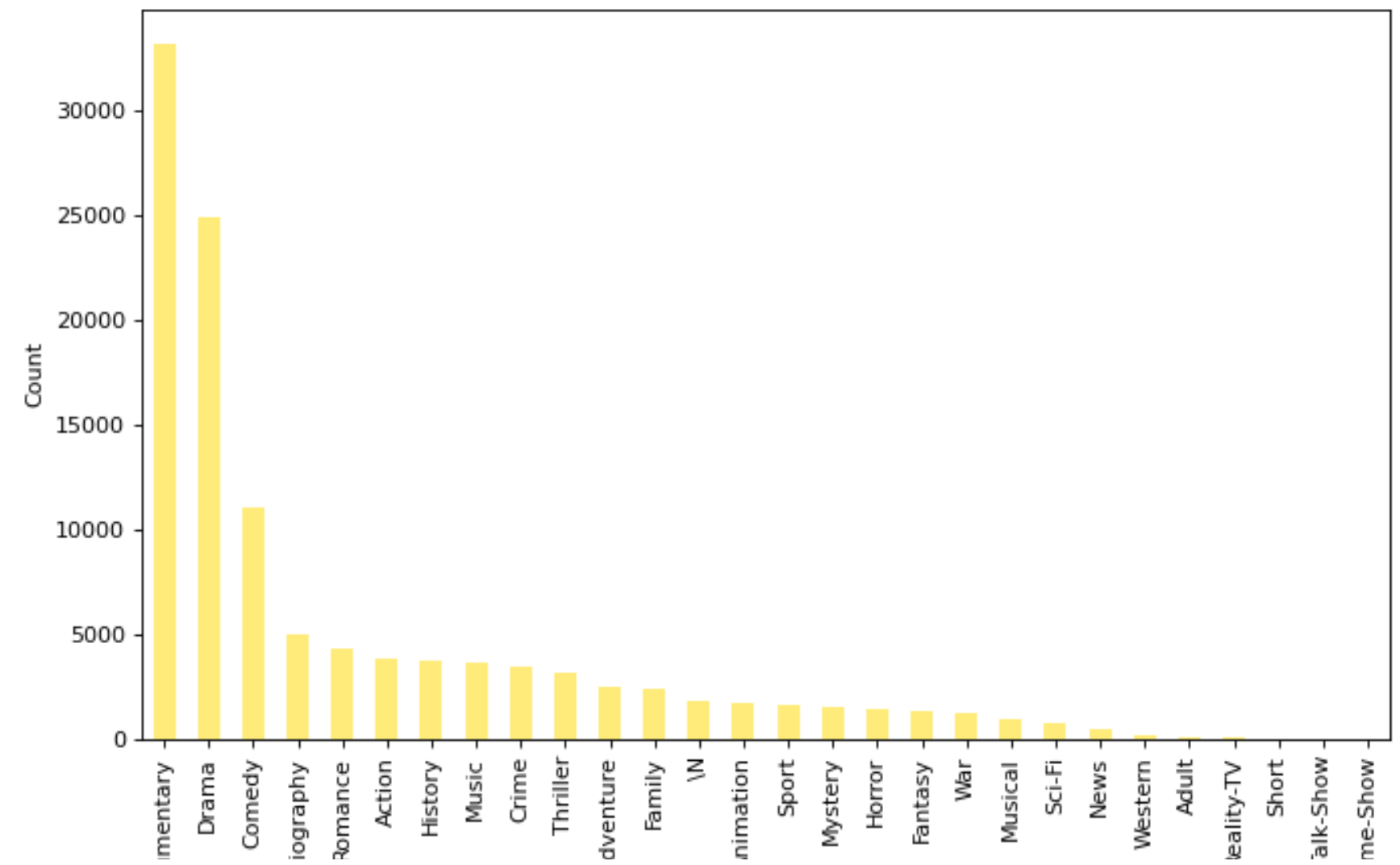


Datasets Characterisation

Number of movies per average rating

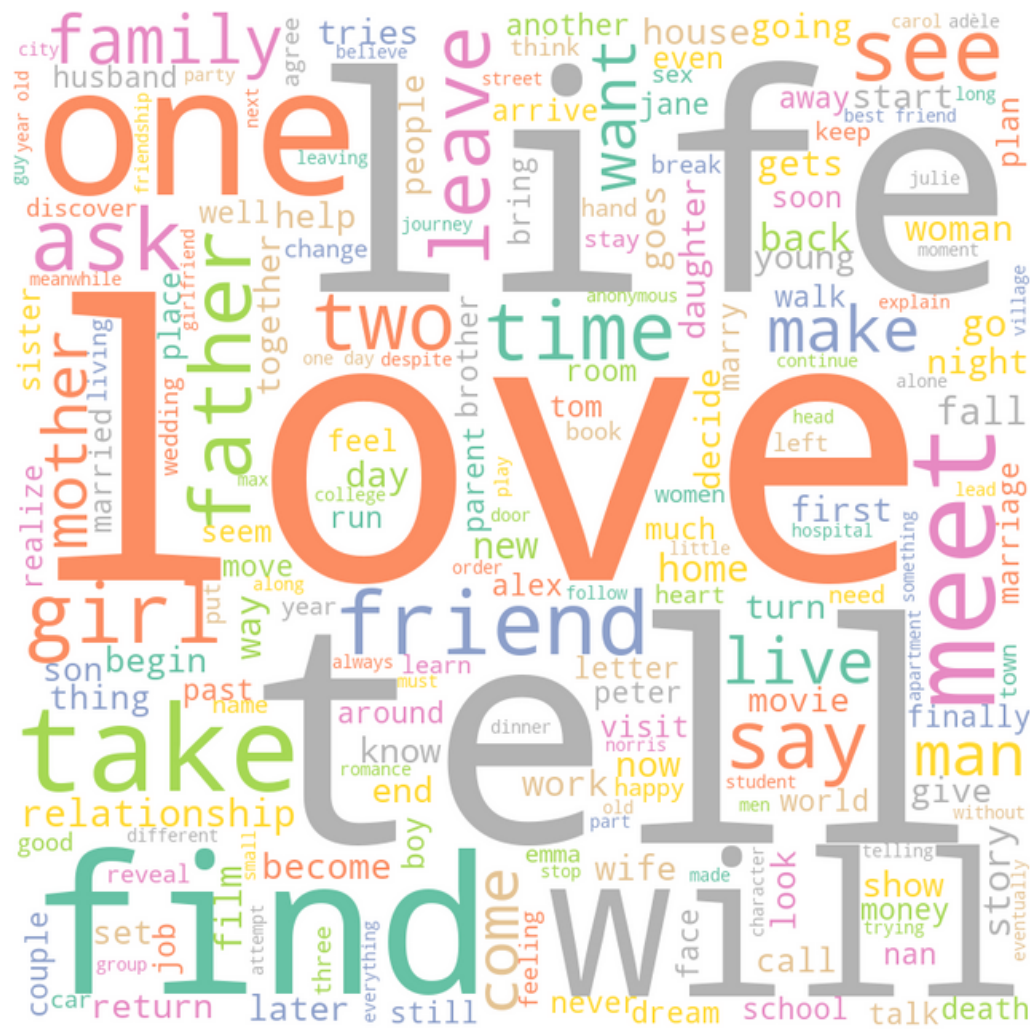


Number of movies per genre



Datasets Characterisation

Wordclouds per genre



Search tasks

- Family Movies
- Movies about the space
- Romance teen movies
- Movies to watch in your twenties
- Life-changing movies
- ...

