

Data Analysis exercise for ourstreaming.com*

Cesare Ressa

*Ourstreaming.com is an imaginary company!

Let's start with the approach

The "Exercise: Insights" has been split in two parts:

Requests

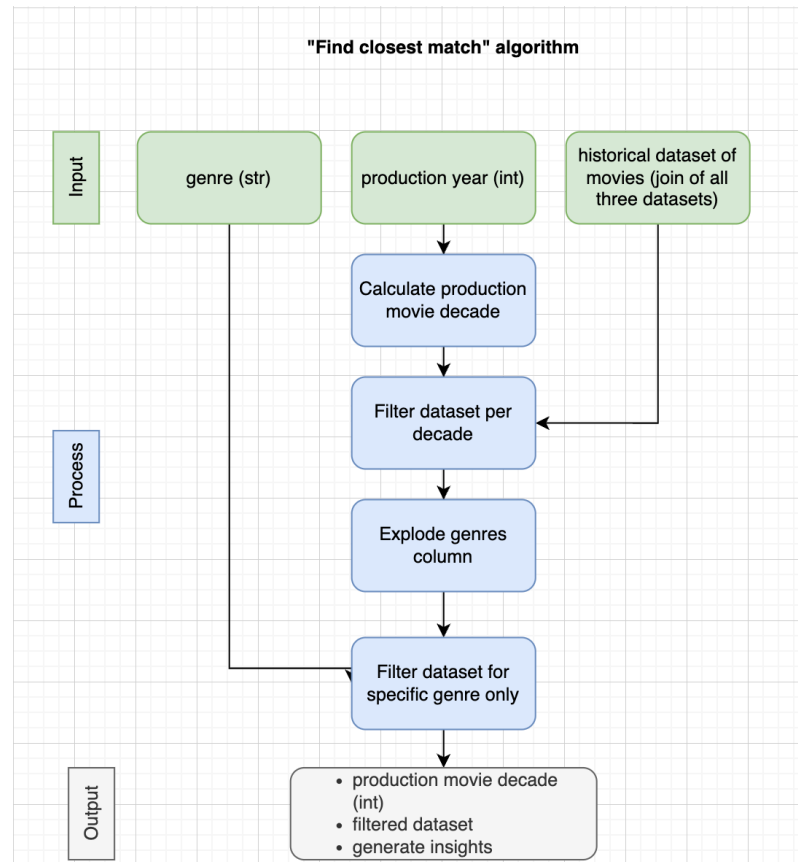
1. Help the procurement team making data-driven decisions on what to sign next
2. Provide valuable insights for the Studio



Solution

1. Development of an algorithm that generates insights and new metrics (e.g., total views) based on historical data of similar movies produced in the same decade and same genre
2. Generated insights, e.g.:
 - Views over time by genre,
 - Most watched movies in a defined time interval,
 - Most watched genres
 - Movie saga (or sequels) analysis

1. Algorithm for the procurement team



- What it does

The “find closest match” algorithm compares a proposed new movie with similar movies, based on genre and production year. It returns the decade in which the proposed movie was produced, a filtered dataframe of movies that match the genre and decade criteria, and generates some insights.

- Why needed

The procurement team can be in this scenario: they have received a proposal for a movie (e.g., a drama movie produced in 2013). They don't have any data about it yet but they can find out how a similar movie behaved.

- **Caveat**

The algorithm takes only one genre as input.

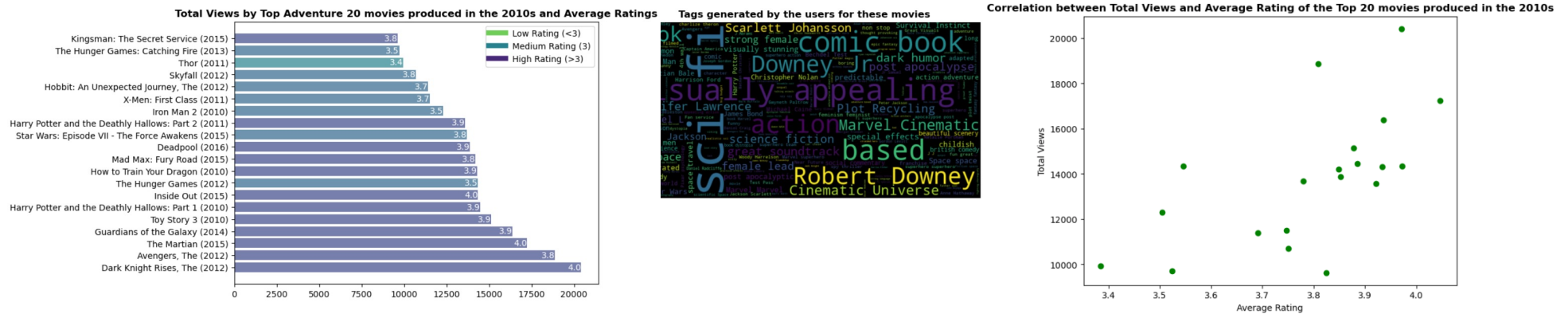
Example on the next slide

Use case of the algorithm for the procurement team

Scenario: The procurement team received a proposal for a movie produced in 2017. We do not know much about it except that it is an Adventure movie.

Inputs to the algorithm: Adventure, 2017, dataset.

Output:



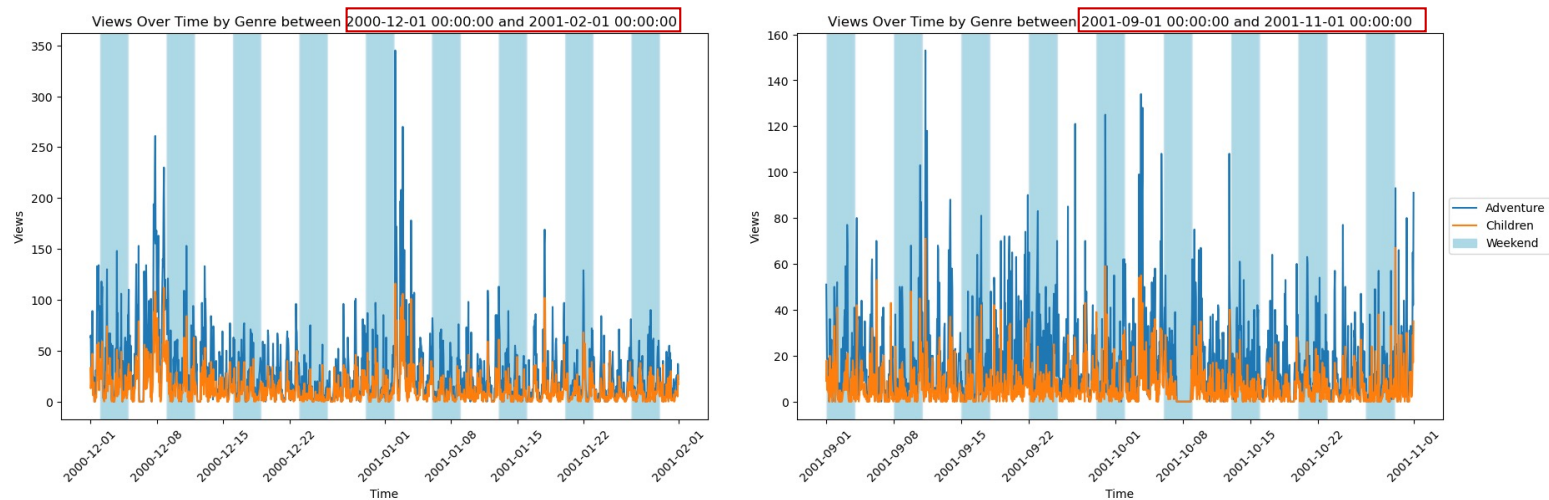
Filtered dataframe with movieId, title, genres, movie_year, total_views, average_rating, tags

2. Valuable insights for the Studio (1)

- **Insight 1: Views over Time by Genre**

Features

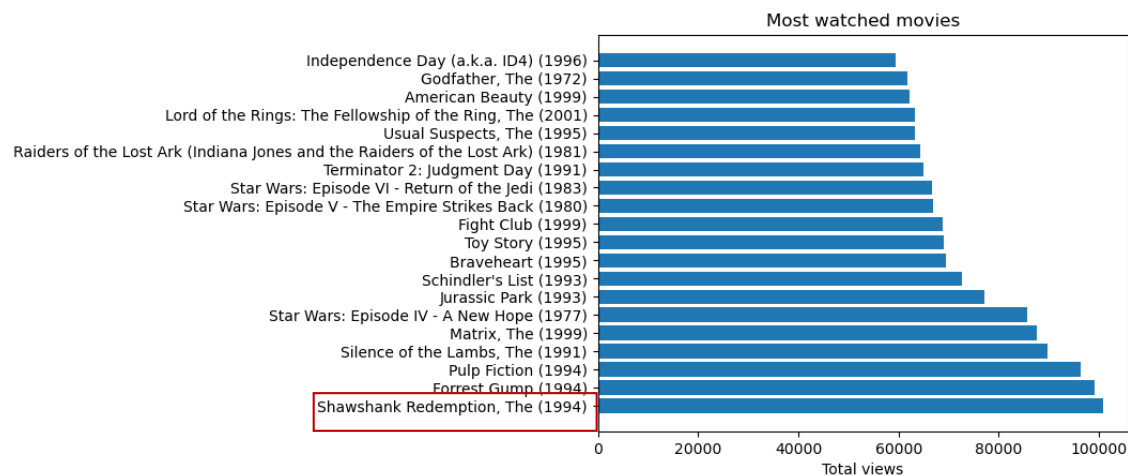
- Multiple genres can be visualized (also all of them)
- A time window and different types of aggregation (e.g., hourly, weekly, monthly) can be defined



The two plots above illustrate the views over time per classic genres watched by Children. The Christmas holidays season is definitely a peak season to broadcast kids movies!

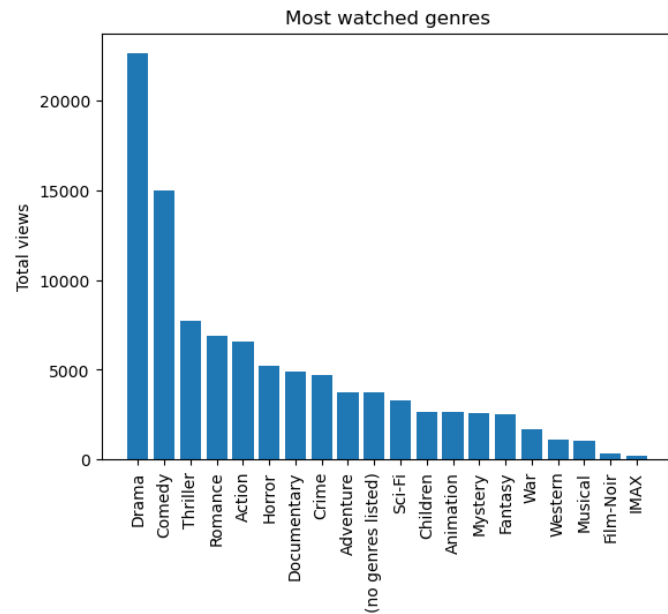
2. Valuable insights for the Studio (2)

- **Insight 2: Most watched movies**
 - Can be filtered per time window and genre(s)



The most-watched movie (**100805 views**) of all time is **The Shawshank Redemption**. The movie has been produced in 1994. It has an average rating of 4.43 and has more than 1000 tags.

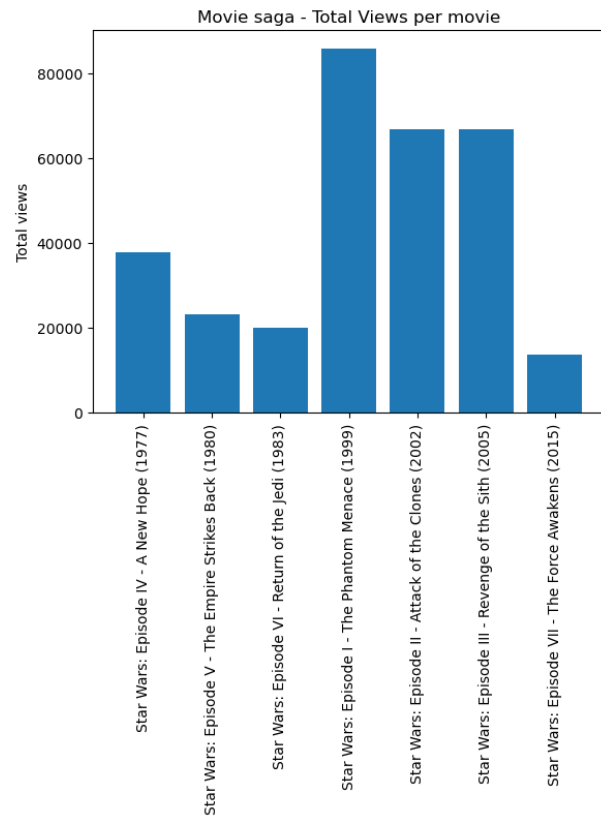
2. Valuable insights for the Studio (3)



- **Insight 3: Most watched genres**
 - Can be filtered per time window (e.g., of all times / year / month)

Drama, Comedy and Thriller are top 3 most-watched genres of all times

2. Valuable insights for the Studio (4)



- **Insight 2: Movie Saga / Movie sequel analysis**
 - Can be filtered per movie saga title (e.g. "Star Wars: Episode")

The successful Star Wars saga shows that the sequels generally had a guaranteed audience and makes easy to launch the movie. The plot shows absolute views per movie. Filtering by launch year can help understand how a sequel movie performed w.r.t. the first movie.

Conclusion

- An algorithm able to compare how a proposed movie would behave based on similar movies has been developed for the procurement team.
- Multiple insights have been generated for the Studio.
- *Other interesting insights can be:*
 - Distribution of ratings per movie
 - Total views per licensing cost – this insight can be very interesting for the procurement team to help select the correct movie at the right price (or justify a higher price), and it can be very helpful for the Studio to define the right price for the movie (based on expected views, forecasts).

Thank you!

Cesare Ressa

- cesare.ressa@gmail.com
- [Link to the GitHub repository](#)
- [LinkedIn profile](#)