



# LIGHTS, CAMERA, ANALYTICS

DOCUMENTATION

John Johannaber

[jwjohannaber@gmail.com](mailto:jwjohannaber@gmail.com)

January 2025

# DASHBOARD OVERVIEW

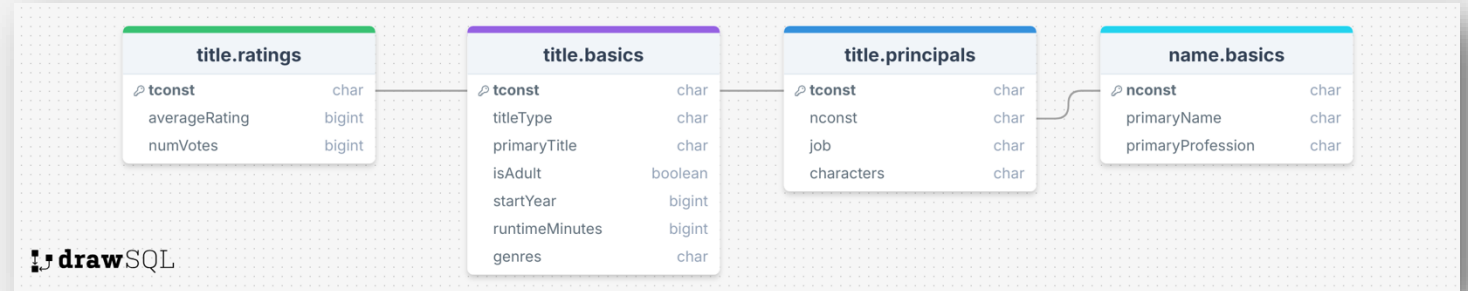
- Data Source
  - IMDb's dataset containing information on movies and TV shows: [IMDb Dataset](#).
- Goal
  - Analyze movie trends over the last 30 years
- Outcome
  - Public dashboard providing insights into movie genres, ratings, and runtime, with an annual trends focus: [Lights, Camera, Analytics Dashboard](#)

## TABLE DETAILS

- *title.basics*
  - Movie/TV title, type, release year, runtime, genres
- *title.principals*
  - Cast/crew associated with the title
- *name.basics*
  - Cast/crew information
- *title.ratings*
  - Movie ratings (average and number of votes)

# SCHEMA DIAGRAM

- *Left join* for title.basics to title.principals
- *Inner join* for title.principals to name.basics
  - Only want actors if name is present
- *Left join* for title.basics to title.ratings

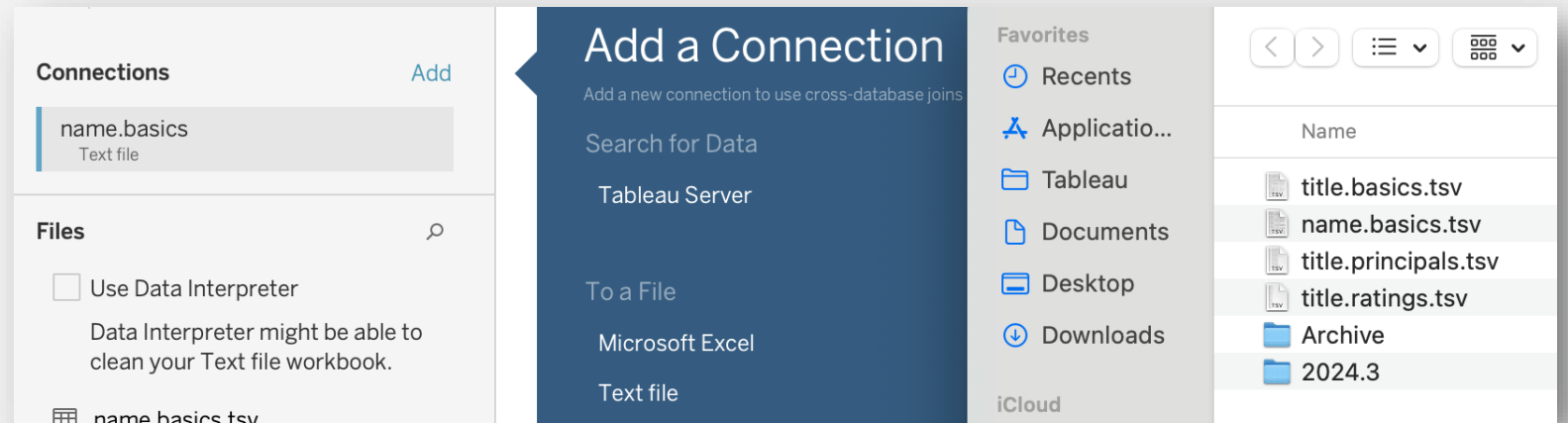


Movies - Actors is made of 4 tables. ⓘ

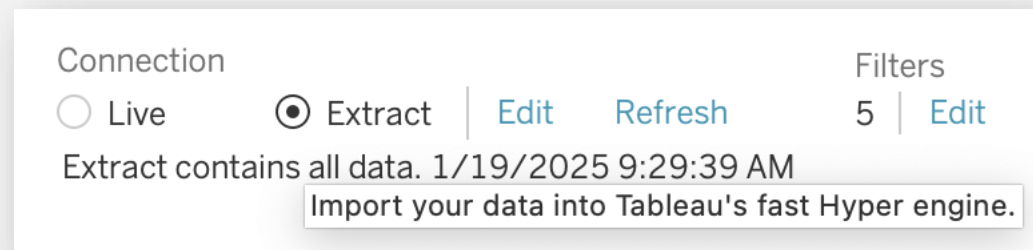


# DATA INGESTION

- TSV text files were downloaded locally
- Imported to Tableau
  - *Data Source tab > Add > Text file > select text file*

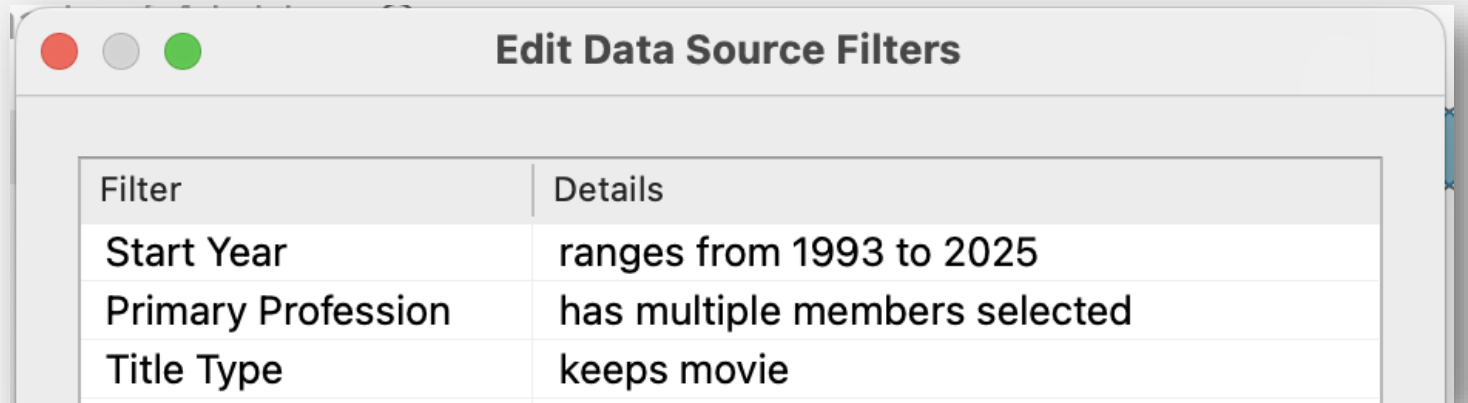


- An extract was created, after data modeling



# INITIAL FILTERING

- Filtered the dataset to only relevant records
  - Only looking at the past 30 full years of data
    - 2024 may not be fully closed out
  - Only keep actors and actresses of the Principals table
  - Keep only movies

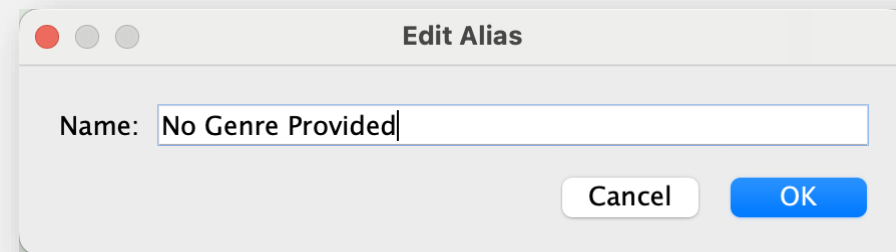
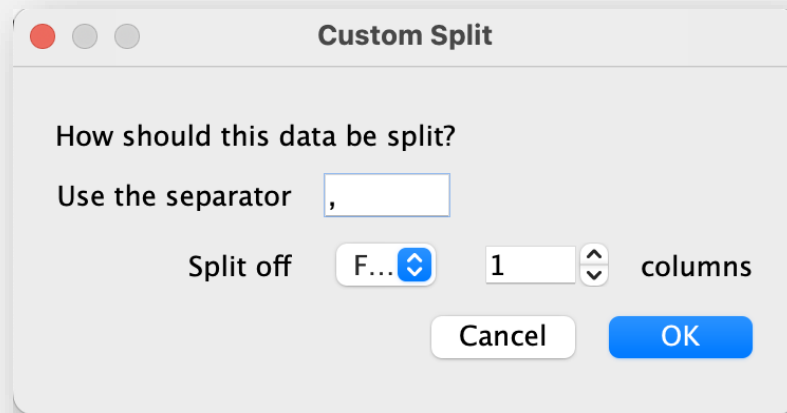


A screenshot of a software window titled "Edit Data Source Filters". The window contains a table with two columns: "Filter" and "Details". The table lists three filters: "Start Year" with details "ranges from 1993 to 2025", "Primary Profession" with details "has multiple members selected", and "Title Type" with details "keeps movie".

| Filter             | Details                       |
|--------------------|-------------------------------|
| Start Year         | ranges from 1993 to 2025      |
| Primary Profession | has multiple members selected |
| Title Type         | keeps movie                   |

# TRANSFORMATIONS: TITLE.BASICS.GENRES

- Custom split with comma separator
  - To split off the primary genre of the film
- Added an alias for \N to be *No Genre Provided*



# TRANSFORMATIONS: TITLE.PRINCIPALS.CHARACTERS

Calculated column to remove brackets, quotation marks, and \N

Character

×

```
replace(replace(replace([Characters (Title.Principals.Tsv1)], "\N", ""), "]", ""), "[", "")
```



# TRANSFORMATIONS:

## TITLE.BASICS.RUNTIMEMINUTES

Calculated column to convert the minutes to hours and minutes

Runtime



```
STR(INT([Runtime Minutes]/60)) + " hr " + STR([Runtime Minutes]%60) + " min"
```

# METRIC CALCULATIONS

Number of Movies

Number of Movies

`COUNTD([Tconst])`

Number of Movies as a Percent of Total

☰ Rows

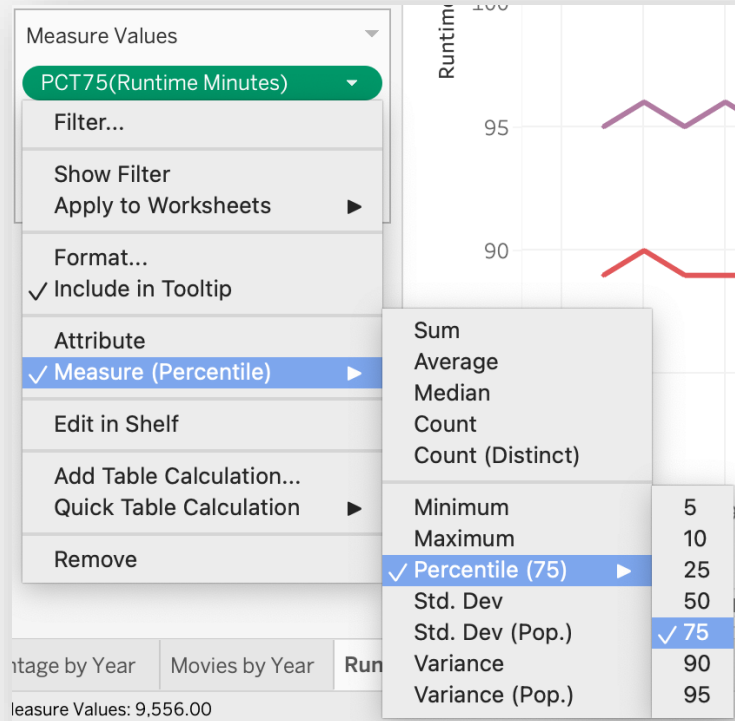
`[Number of Movies] / TOTAL([Number of Movies])`

Impact Score: *MIN()* used to increase performance

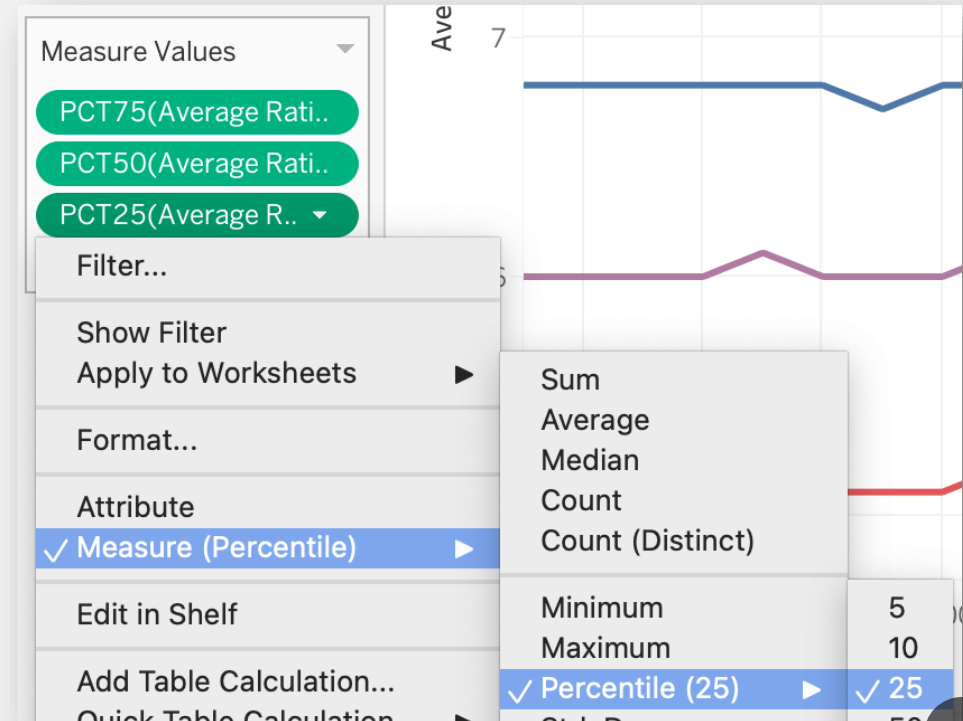
`MIN([Average Rating])*MIN([Num Votes])`

# METRIC CALCULATIONS

Runtime: *Percentiles 25/50/75*



Ratings: *Percentiles 25/50/75*



## FUTURE WORK

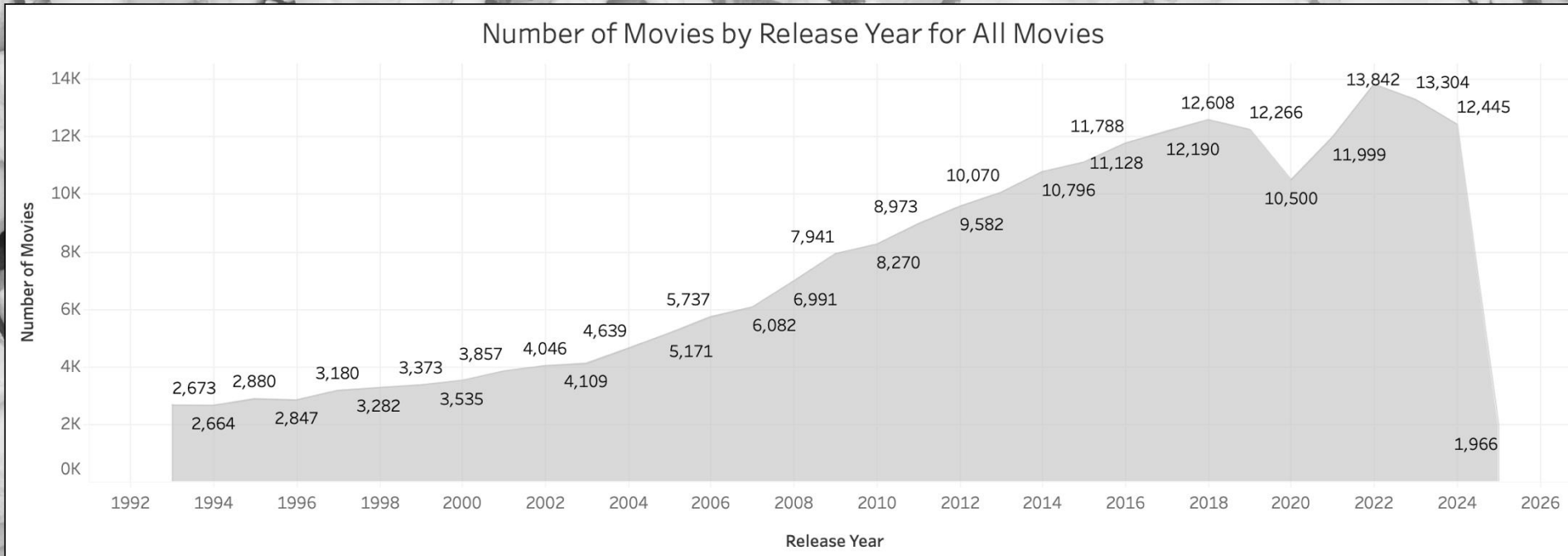
- Automated the data pipeline
  - Apply filtering and transformations outside of Tableau
- Build a dashboard for TV Shows
- Obtain the full IMDB data
  - Includes box office performance for deeper insights
  - Requires a fee to download

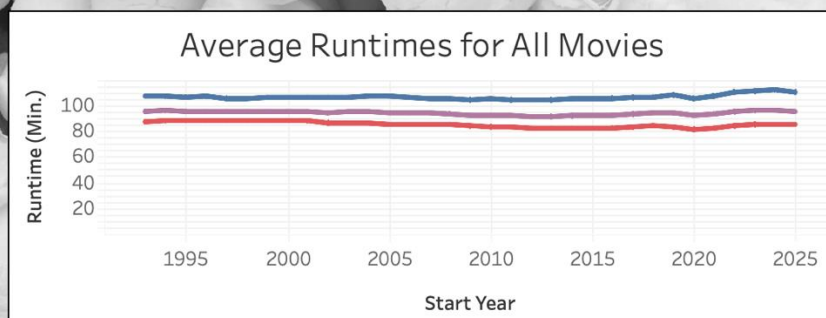
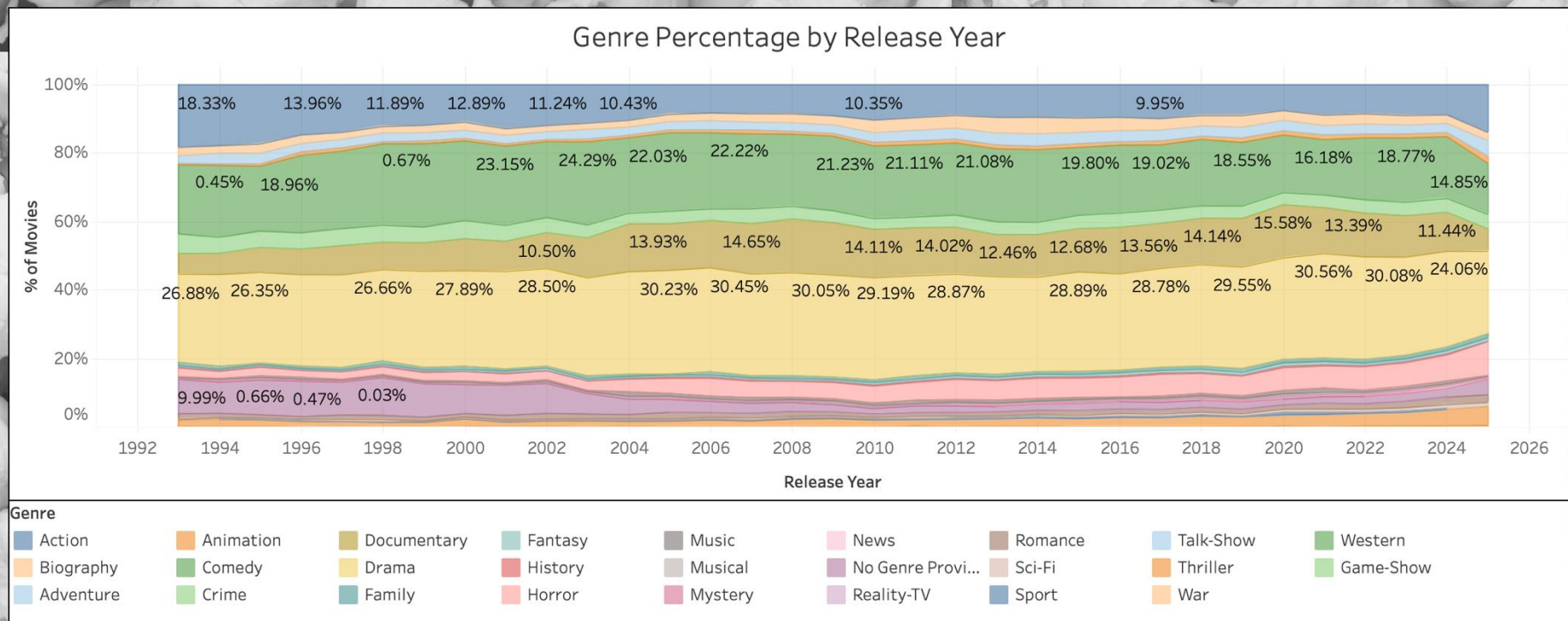
## APPENDIX: DASHBOARD SCREENSHOTS

# LIGHTS, CAMERA, ANALYTICS

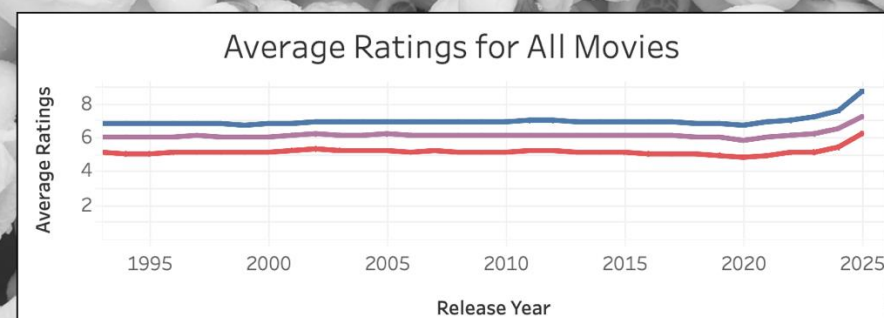
## Unpacking Cinema Trends Through Data

This dashboard explores cinema trends through a detailed analysis of IMDb movie data. It highlights genre popularity, average runtimes, ratings evolution, and dives into the most impactful movies and genres by year. All visualizations are designed to uncover insights into the evolving landscape of film over time.



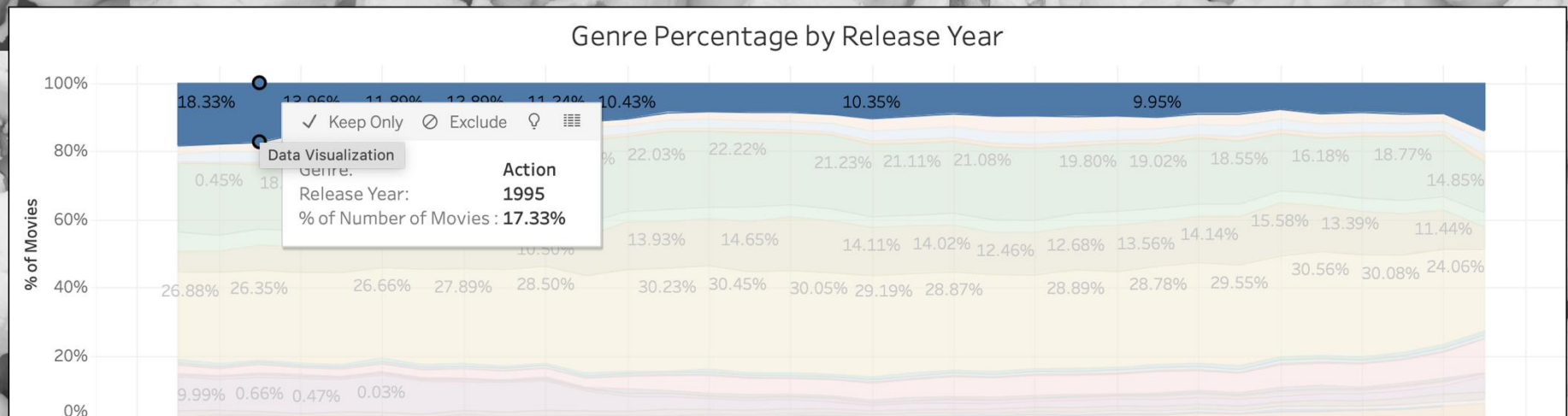
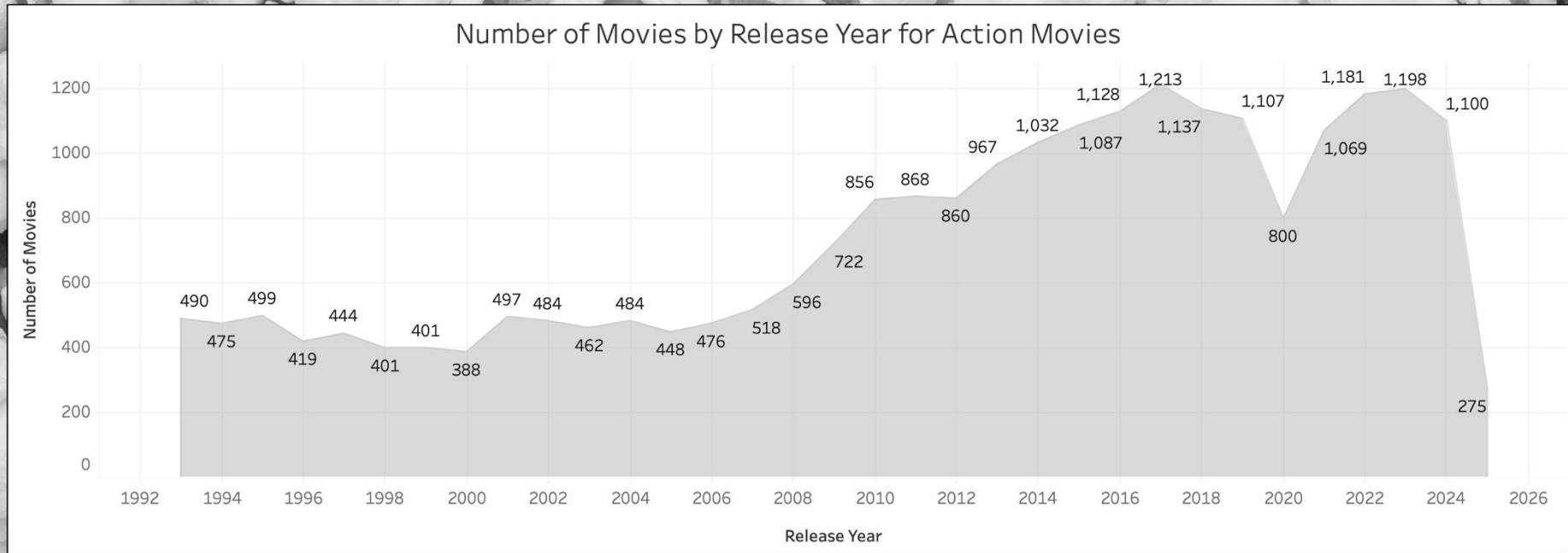


■ Percentile (75) of Runtime Minutes  
 ■ Percentile (50) of Runtime Minutes  
 ■ Percentile (25) of Runtime Minutes



■ Percentile (75) of Average Rating  
 ■ Percentile (50) of Average Rating  
 ■ Percentile (25) of Average Rating







# Zooming In:

## Highlighting Yearly Impact by Movies and Genre

### Impact Score

The Impact Score is calculated by multiplying a movie's average rating by the total number of ratings it has received. This score reflects both the quality of the movie (as indicated by its rating) and its popularity (based on the volume of ratings). A higher Impact Score signifies a movie that resonated with a larger audience.

