

# Data Analysis of **MOVIES AND TV SHOWS**

NETFLIX

## Content:

The used dataset contains

**6.234\* Movies** and **TV-Shows**

added on Netflix from **2011** till **2019**.

The dataset consists of 12 columns like id, title, type, directors, cast, country, release year, duration, genre.

## Tasks:

1. Average of main actors in movies
2. Hype of tv shows?
  - Increase/decrease of movies and tv shows since 2010
3. Duration of movies since 2000

## Technologies used:

☒ Excel 

☒ Python (Pandas)  (  )

☒ Tableau 

\*As I have seen, the dataset does not contain an entire video library of Netflix.

# 1. Average of main actors in movies

- Clean data (Python)
  - Filter: type = 'Movie'
  - Delete 'null'-Values in column 'cast'
- Bring column 'cast' in a usable form:
  - stacking- & unstacking-operations to get a row for each single actor in a movie

show_id	title	type	cast
70125931	The Karate kid	Movie	Jaden Smith, Jackie Chan, Taraji P. Henson, ...
80186796	Dancing with the Birds	Movie	null



show_id	title	type	cast
70125931	The Karate kid	Movie	Jaden Smith
70125931	The Karate kid	Movie	Jackie Chan
70125931	The Karate kid	Movie	Taraji P. Henson
80186796	Dancing with the Birds	Movie	null

- **Result:**

- **29722** main roles in **3905** different movies
  - Average per cast: ~7,6 main roles
  - Actor with most main roles: Anupam Kher & Shah Rukh Khan (29 main roles)

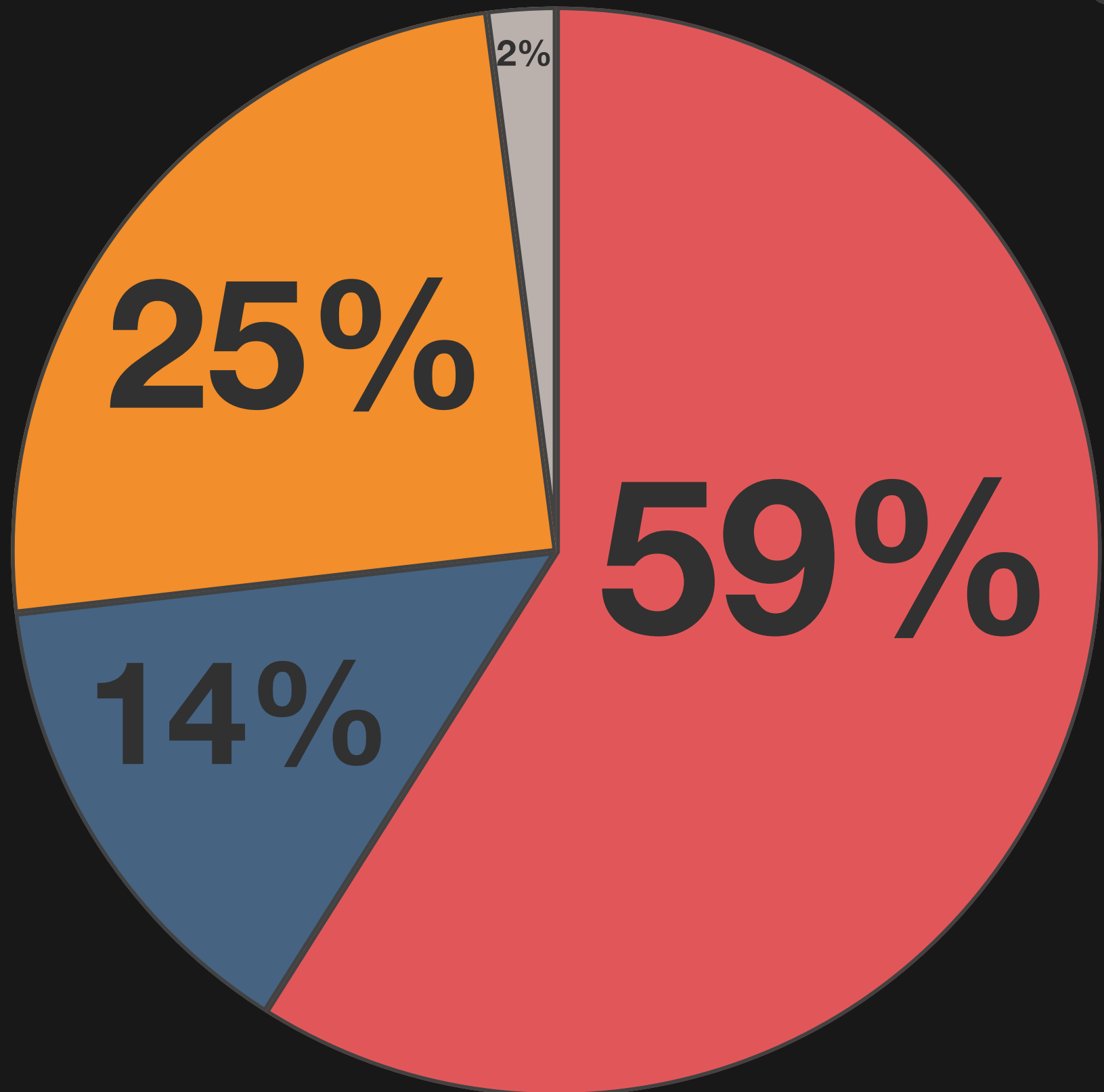
[illegible]

# 1. Average of main actors in movies

1.1 Search percentage of movies with x main roles

- **Result:**

- 1-5 main roles
- 6-10 main roles
- 11-15 main roles
- 16 + main roles



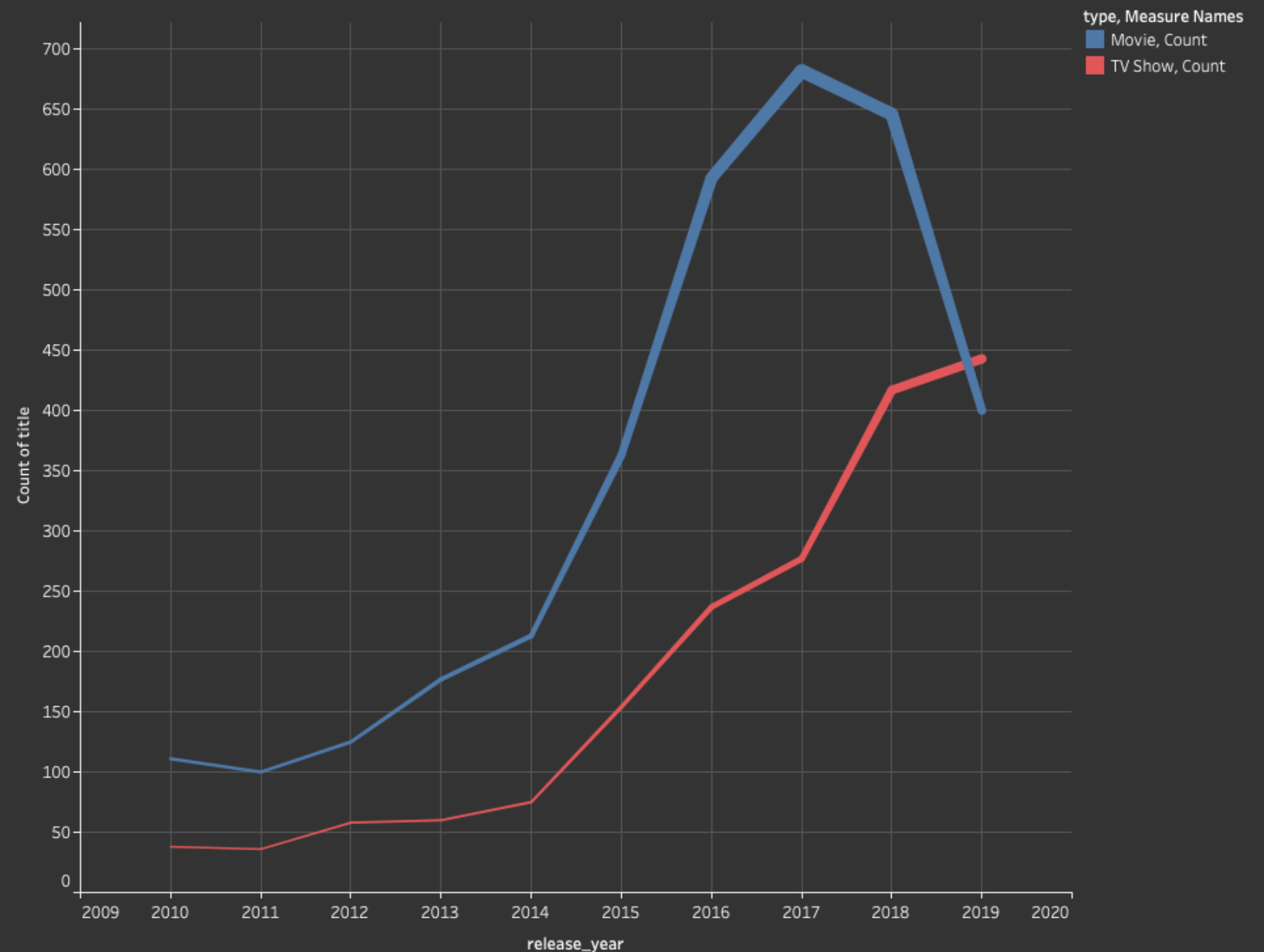
## 2. Increase/decrease of movies & tv shows since 2000

- Clean data from incomplete years
- Sort by type (movie & tv show)
- Show count of movies and tv shows and years in line graph

### Result:

- Releases of **Movies** were higher than of TV Shows from **2010-2019**
- Releases of **TV Shows** were higher than of Movies in **2019**

Releases



## 2. Increase/decrease of movies & tv shows since 2000

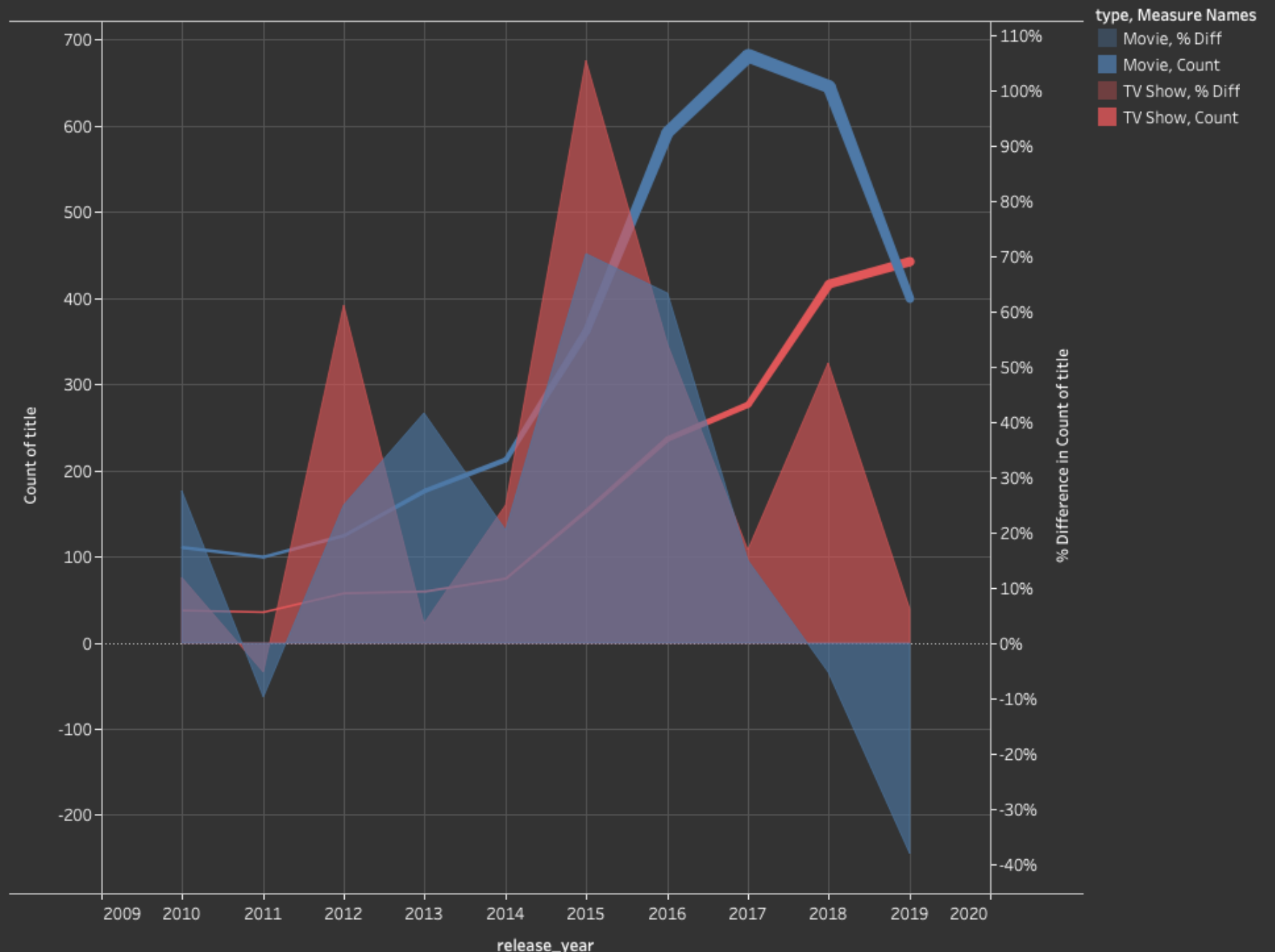
- Clean data from incomplete years
- Sort by type (movie & tv show)
- Show count of movies and tv shows and years in area graph

### Result:

- **Highest increase** of **Movie** and **TV Show** releases in **2015\***
- **Decrease** of **Movie** releases since **2018**
- **Both types** used to have a **similar trend** until **2017\***

\*Compared to increase/decrease of the previous year

Releases & annual increases/decreases



# 3. Duration of movies since 2000

- Clean data (Python)
- Adjust coordinate system and Ganttcharts (Every movie and average)

## Result:

- **Duration of movies**  
decreased **since 2000**

**2000** → **2019**  
**~115min** → **~90min**

