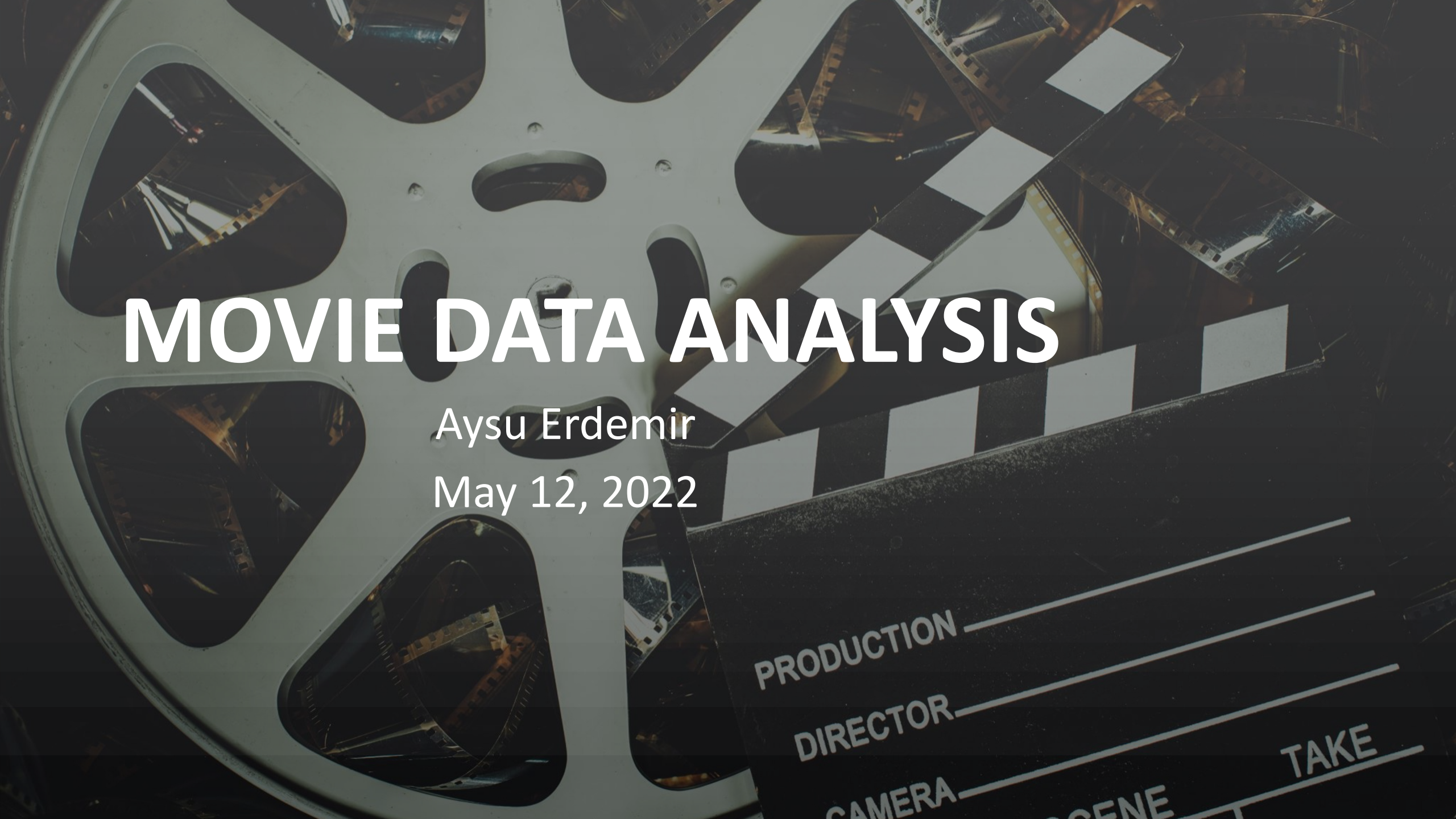


MOVIE DATA ANALYSIS

Aysu Erdemir

May 12, 2022

Business Problem

What type of movies should Microsoft create for their new movie Studio?

- ❑ Find a way to assess movie “profitability”.
- ❑ Explore characteristics of past movies in relation to profitability.
  - What genres of movies to make?
  - Which directors to work with?
  - When to release the movie?
  - Which movie length to focus on?

Data

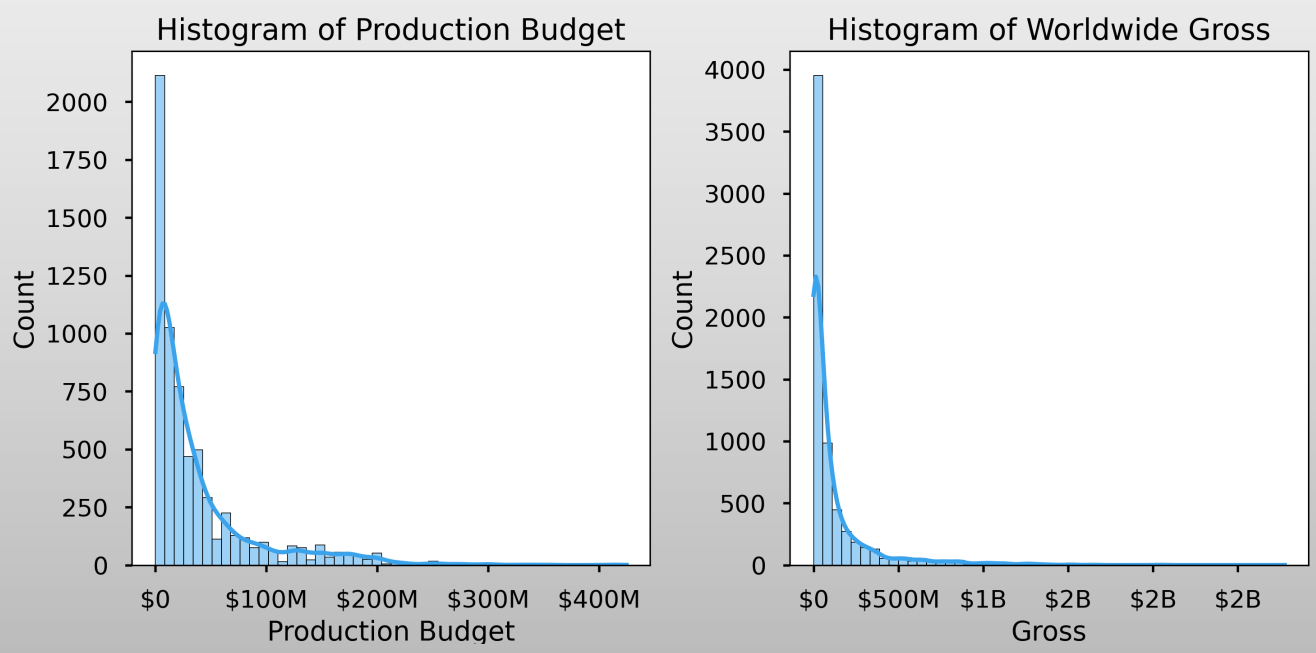
IMDb dataset:

- 140416 movies
- includes **genre**, **release date, runtime, director** and **ratings***

* See appendix I for why we do NOT rely on ratings.

The Numbers dataset:

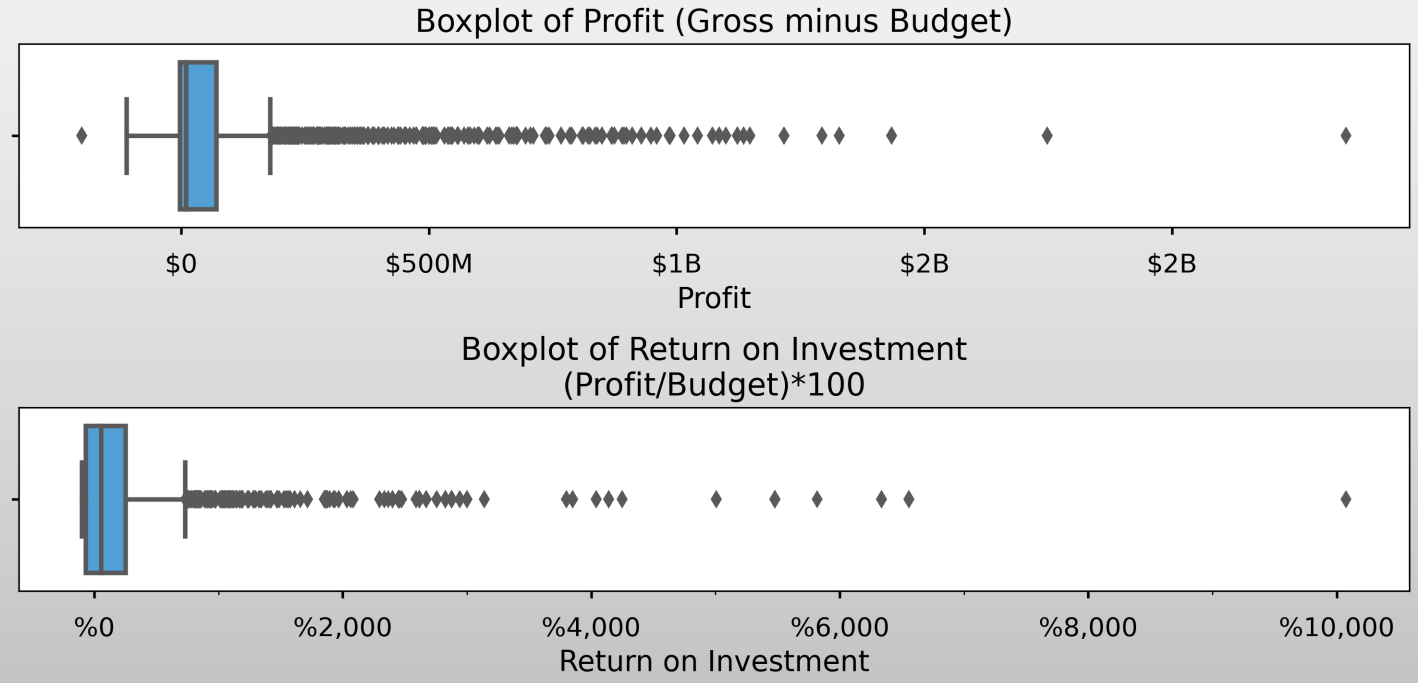
- 5698 movies
- includes **budget** and **gross**



Methods:

Gross - Budget = Profit

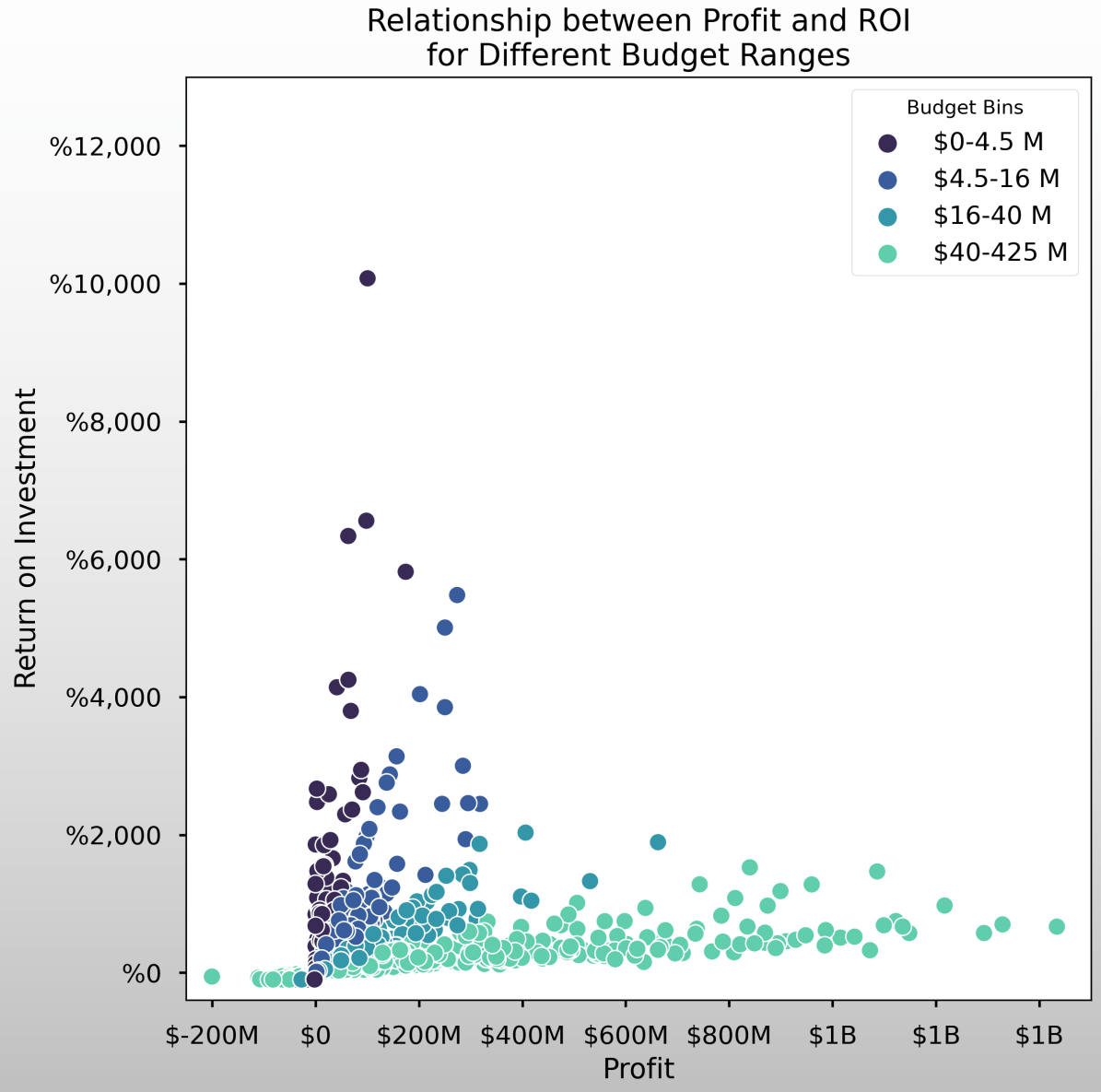
Profit / Budget *100 = ROI



- Derive **Profit** and **Return on Investment** to assess *profitability*.
- Use **MEDIAN** as a measure of central tendency.**

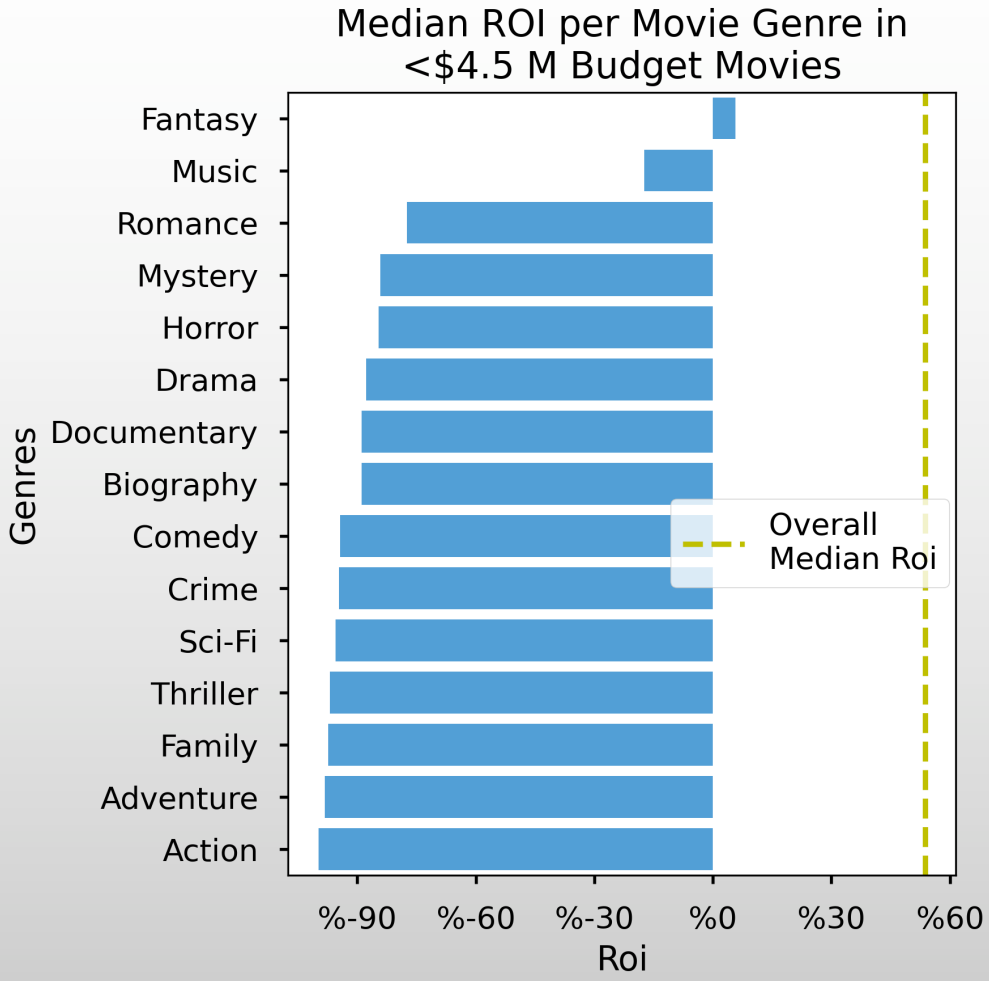
** See appendix II for why we should NOT rely on mean values.

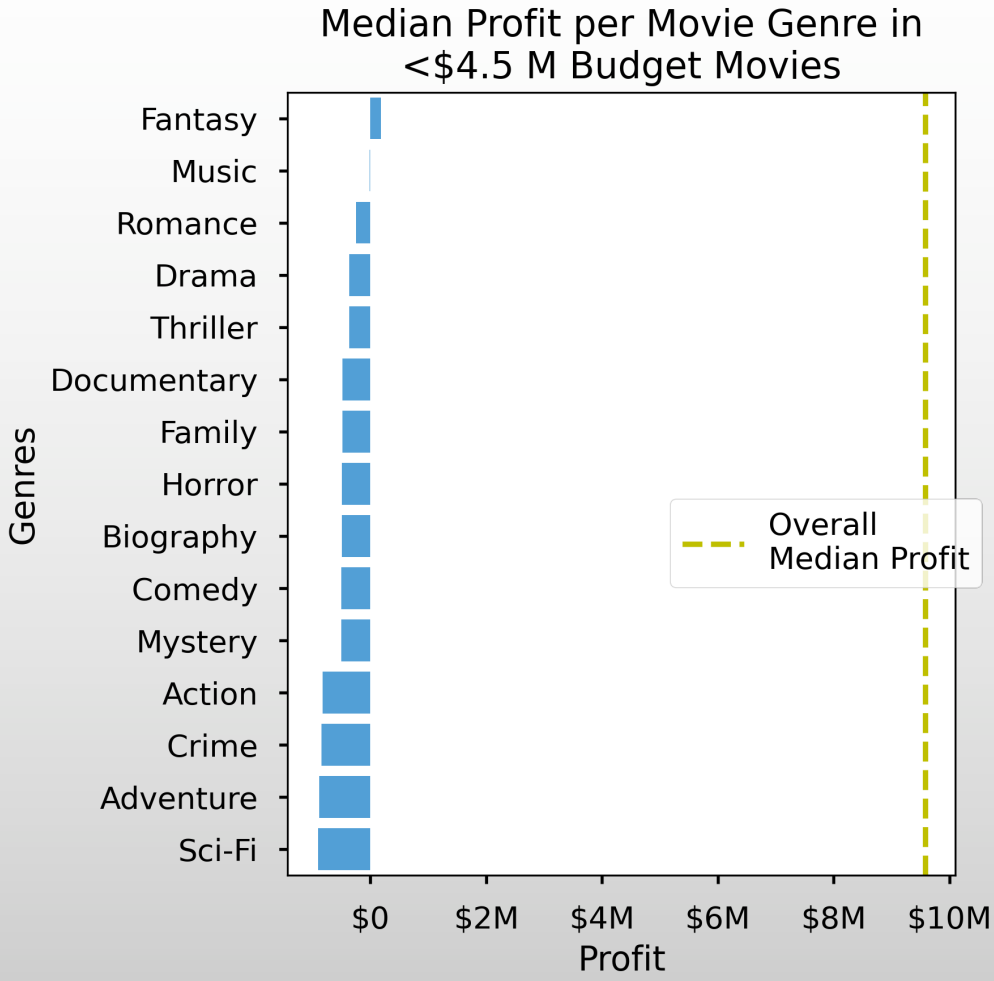
Methods:



- **>$4.5M Budget** : Low profit, high ROI potential.
- **$4.5-40 M Budget** : Moderate profit, moderate ROI potential.
- **>$40M Budget**: High profit, low ROI potential.

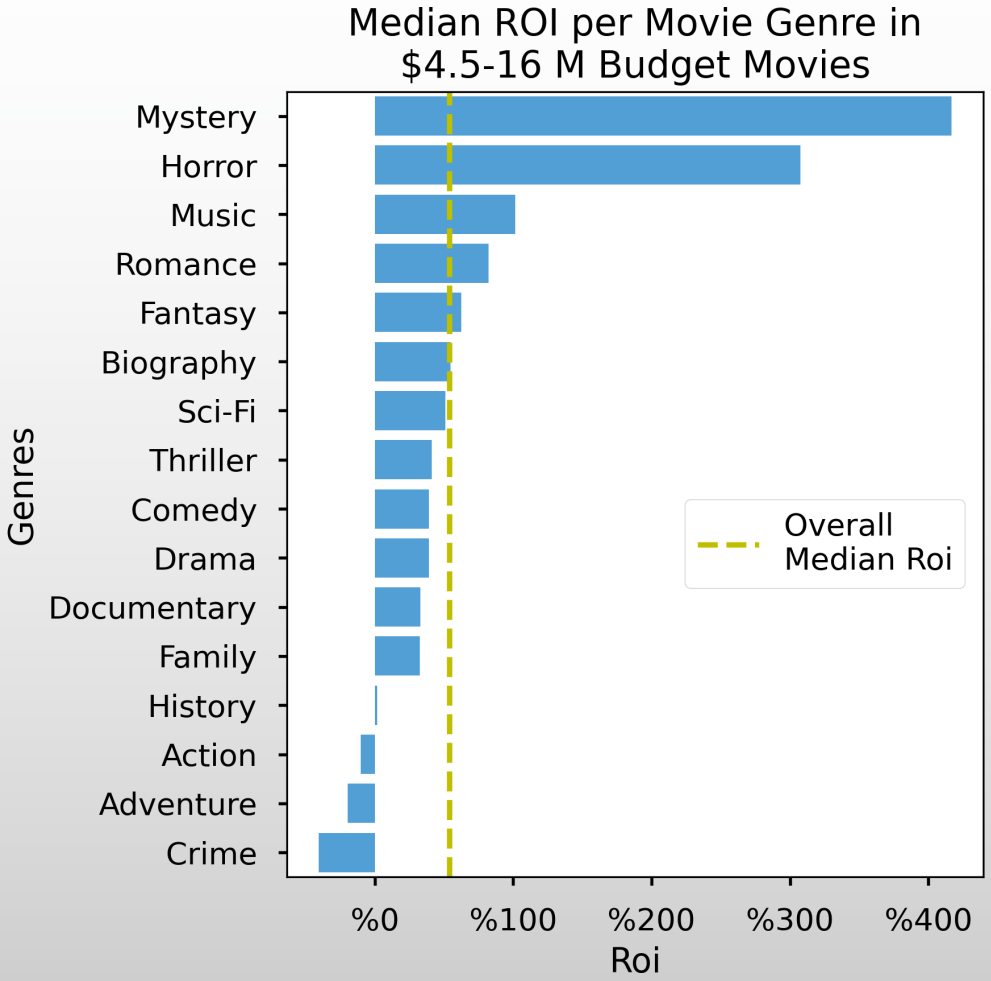
What genres of movies to make ?

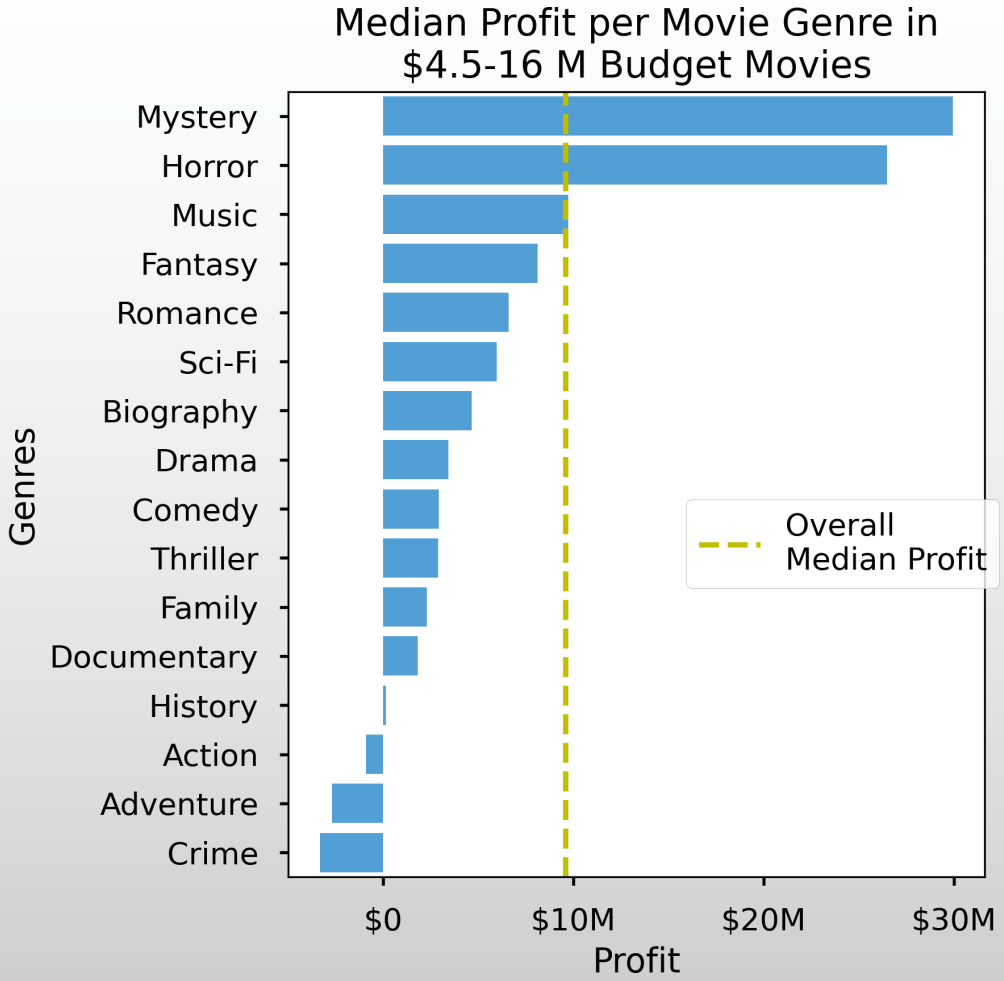




With >$4.5M budget no genre makes it to the median ROI or median profit points.
Avoid this budget range!

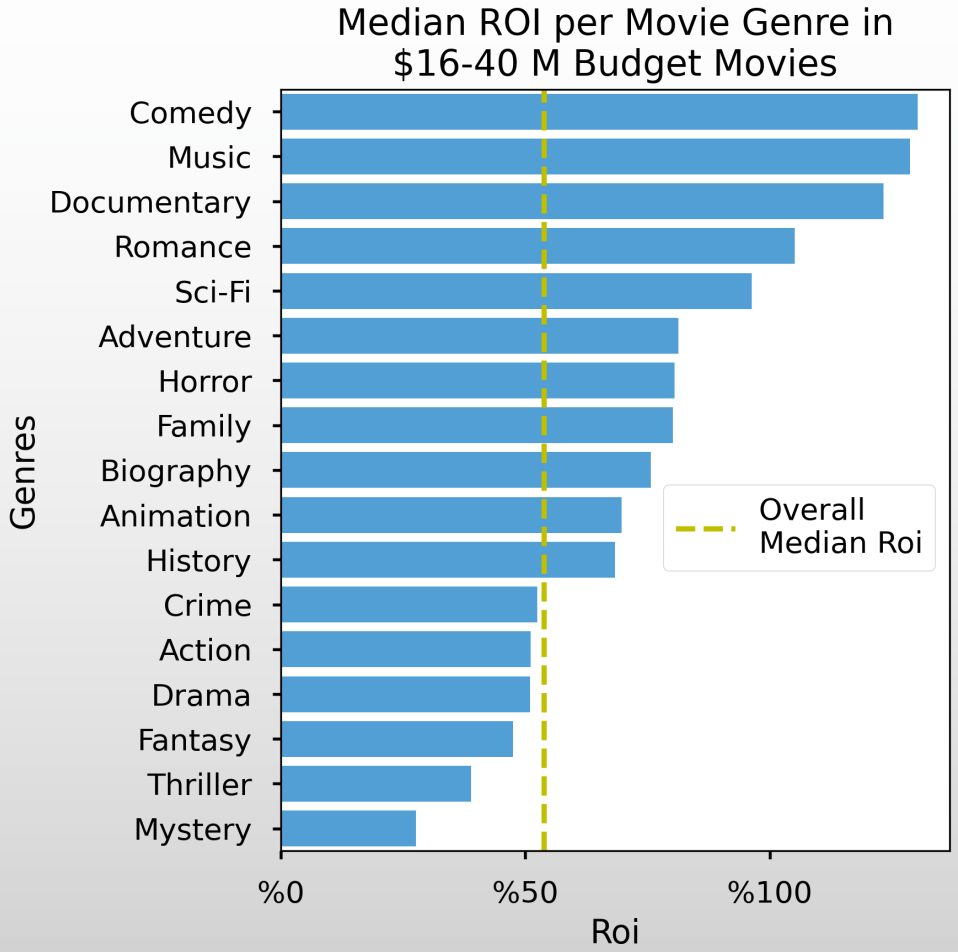
What genres of movies to make?

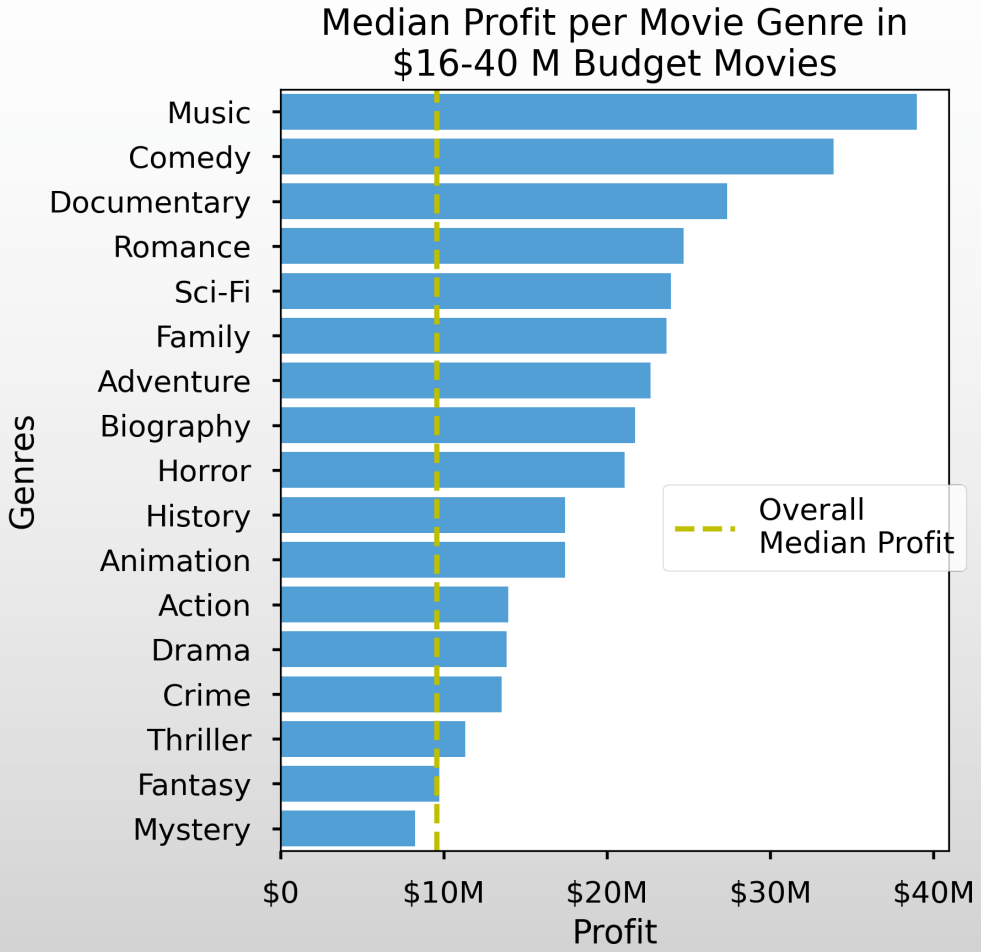




With $4.5-16M budget range make movies of **Mystery** and **Horror:**
They bring more than **300% ROI** and about **$25-30M** in profit.

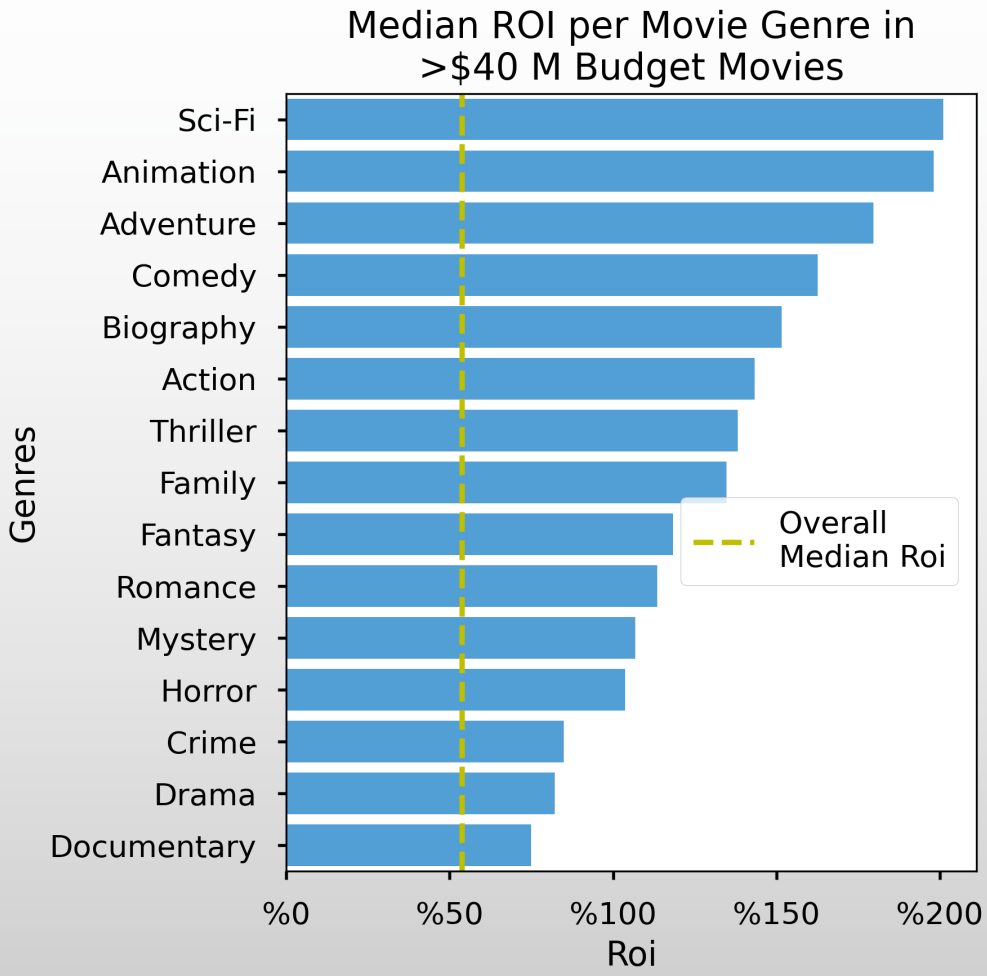
What genres of movies to make?

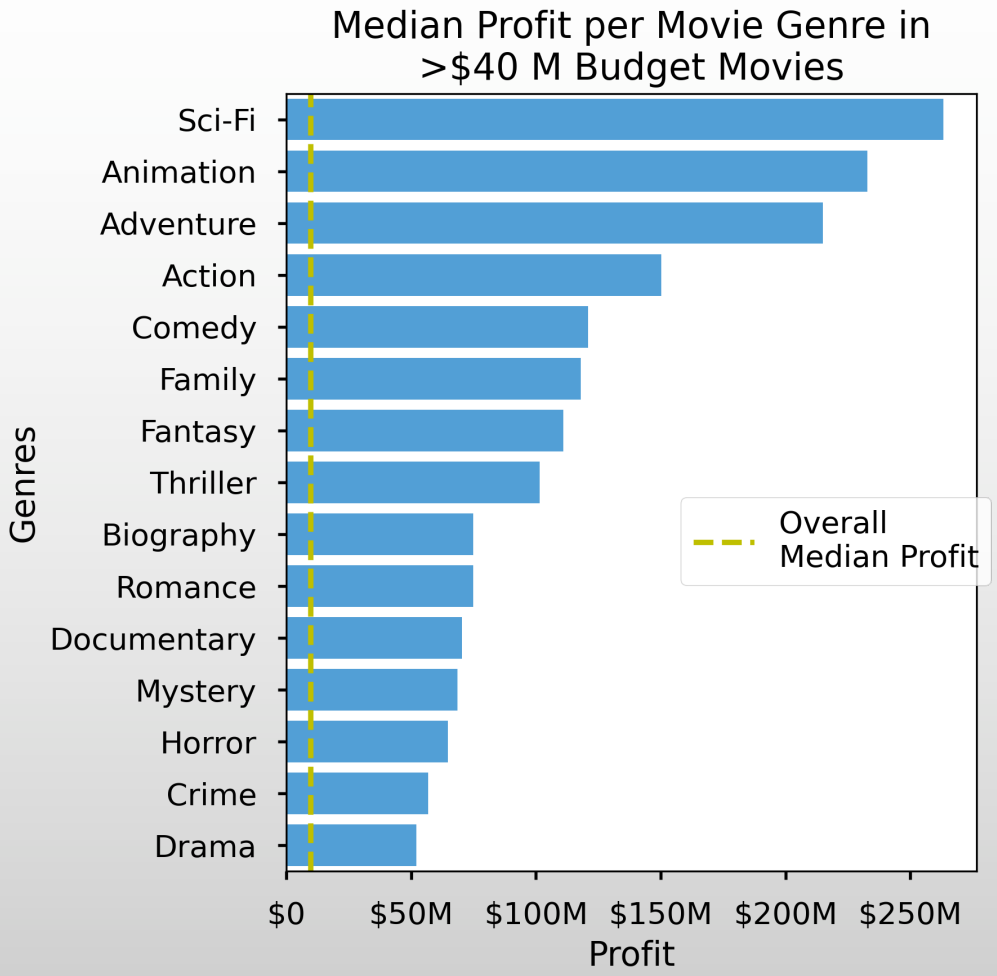




With $16-40M budget range you can make movies of **Comedy** and **Documentary:** However, although they bring the same profit as Horror and Mystery (**$25-30M**), they bring less in return on investment: **100% ROI**.

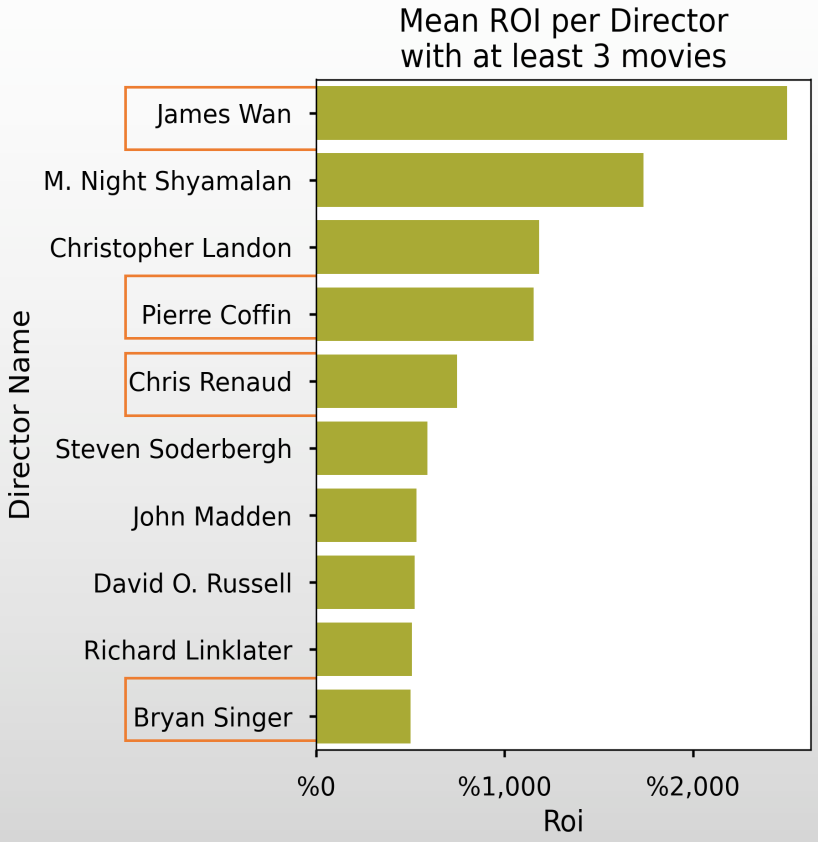
What genres of movies to make?

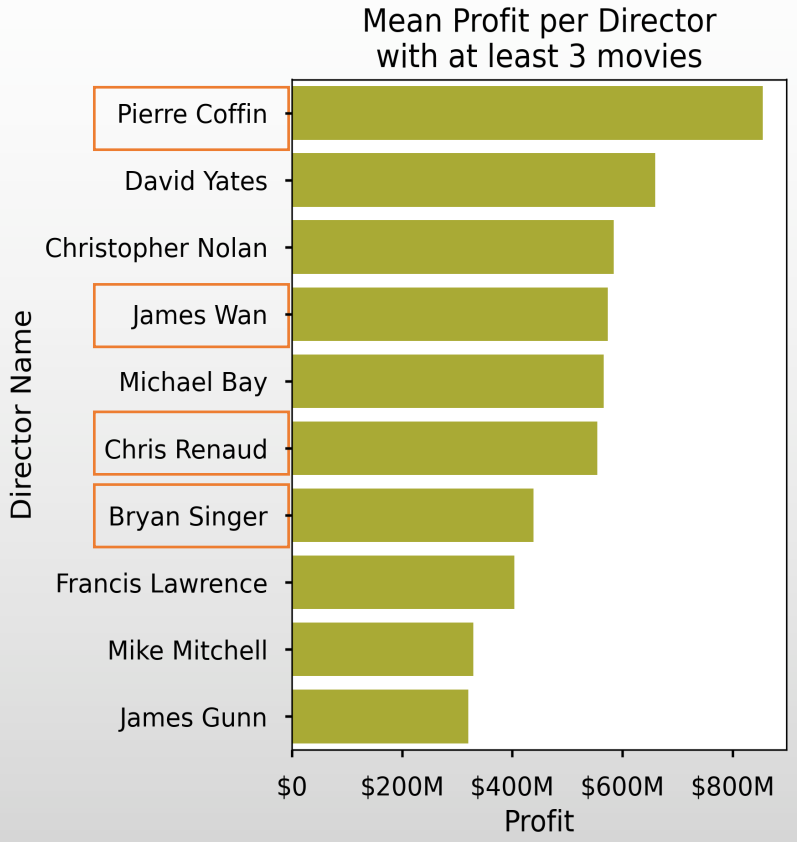




With high budget (>$40M) make movies of **Animation, Sci-Fi** and **Adventure.**
They bring about **200% ROI** and huge $**200-250 M** in profit.

Which directors to work with?





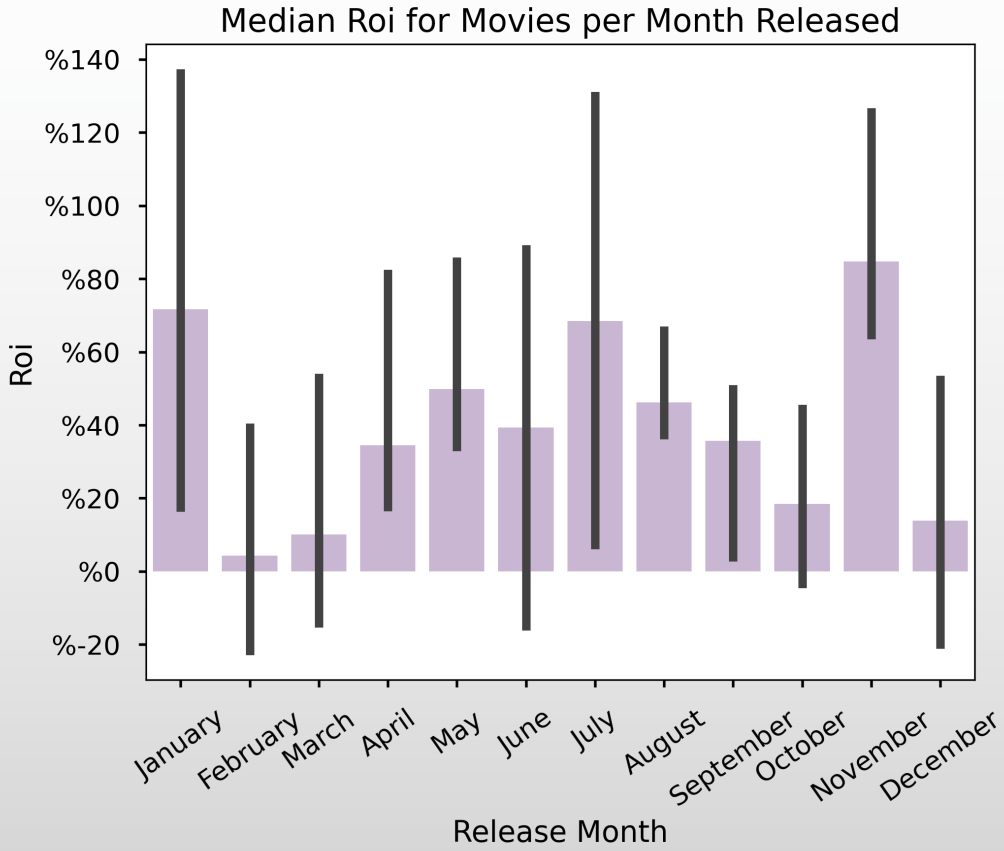
For at least 1000% ROI invest on:
James Wan
M. Night Shyamalan
Christopher Landon
Pierre Coffin

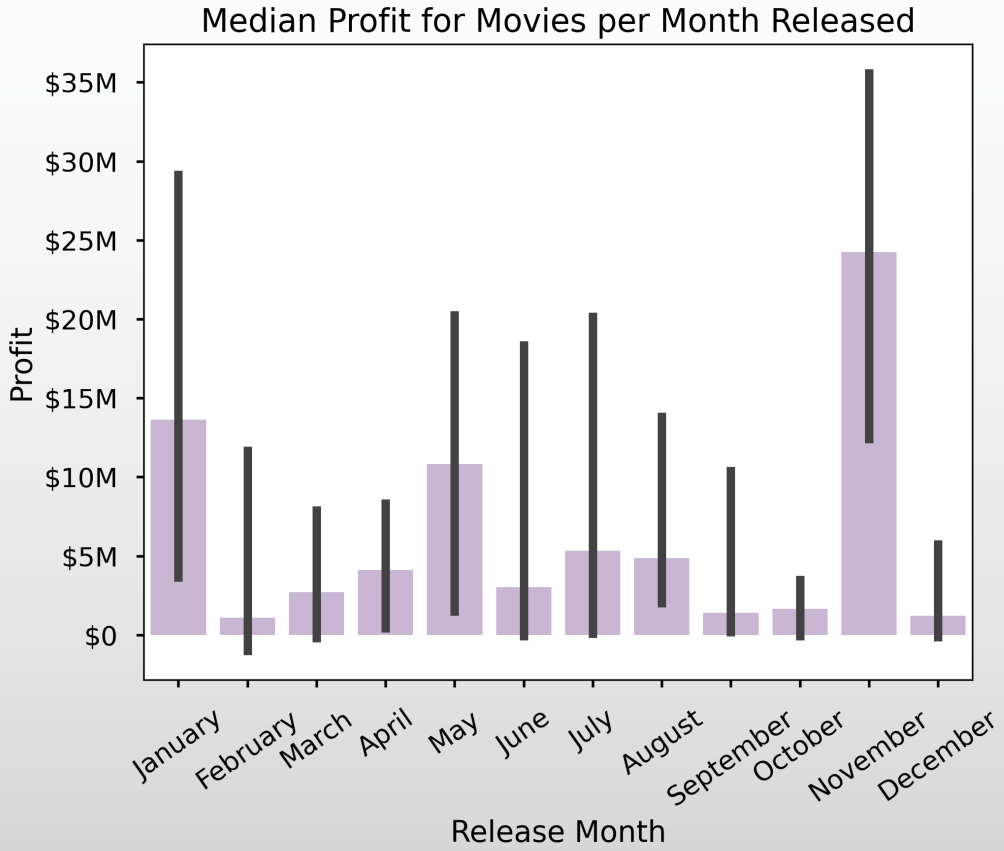
For at least $550M profit invest on:
Pierre Coffin
David Yates
Christopher Nolan
James Wan

4 common names between the top 10 directors - you can invest on with trust:

James Wan, Pierre Coffin, Chris Renaud, Bryan Singer

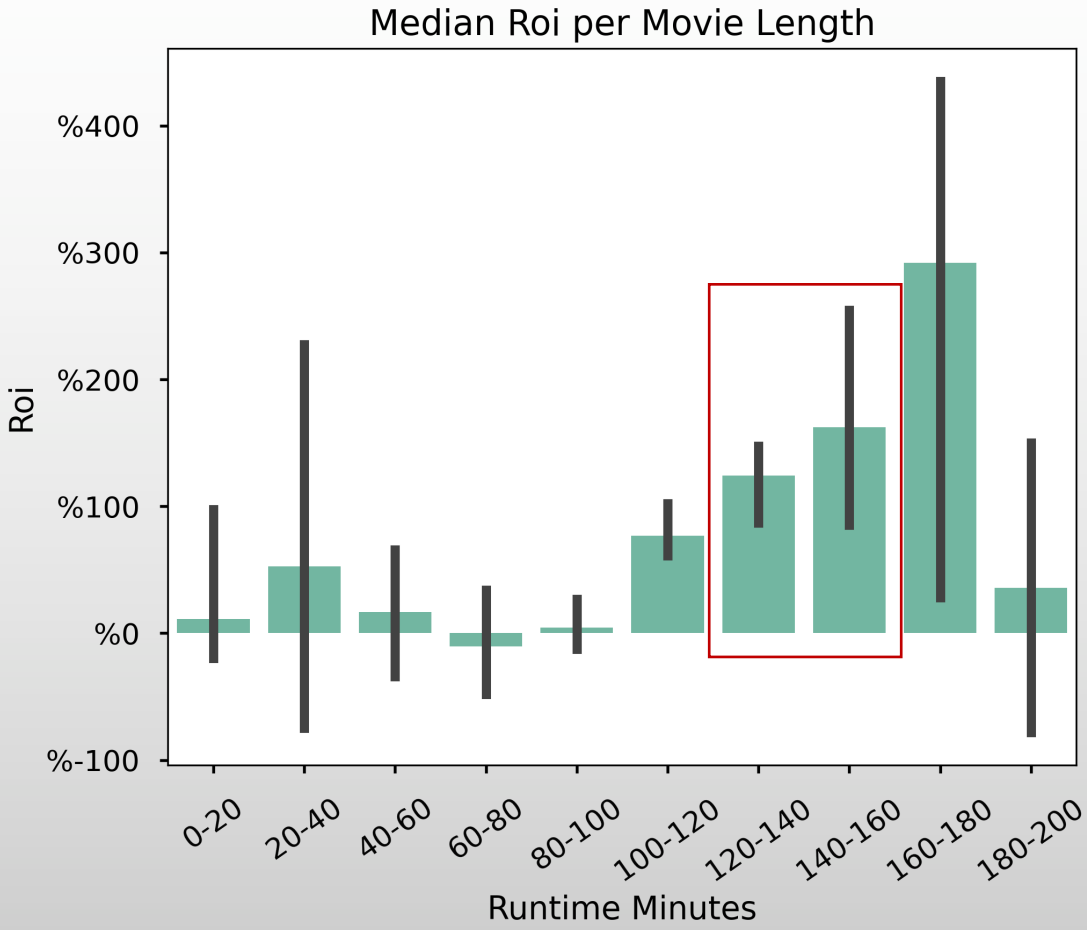
When to release the movie?

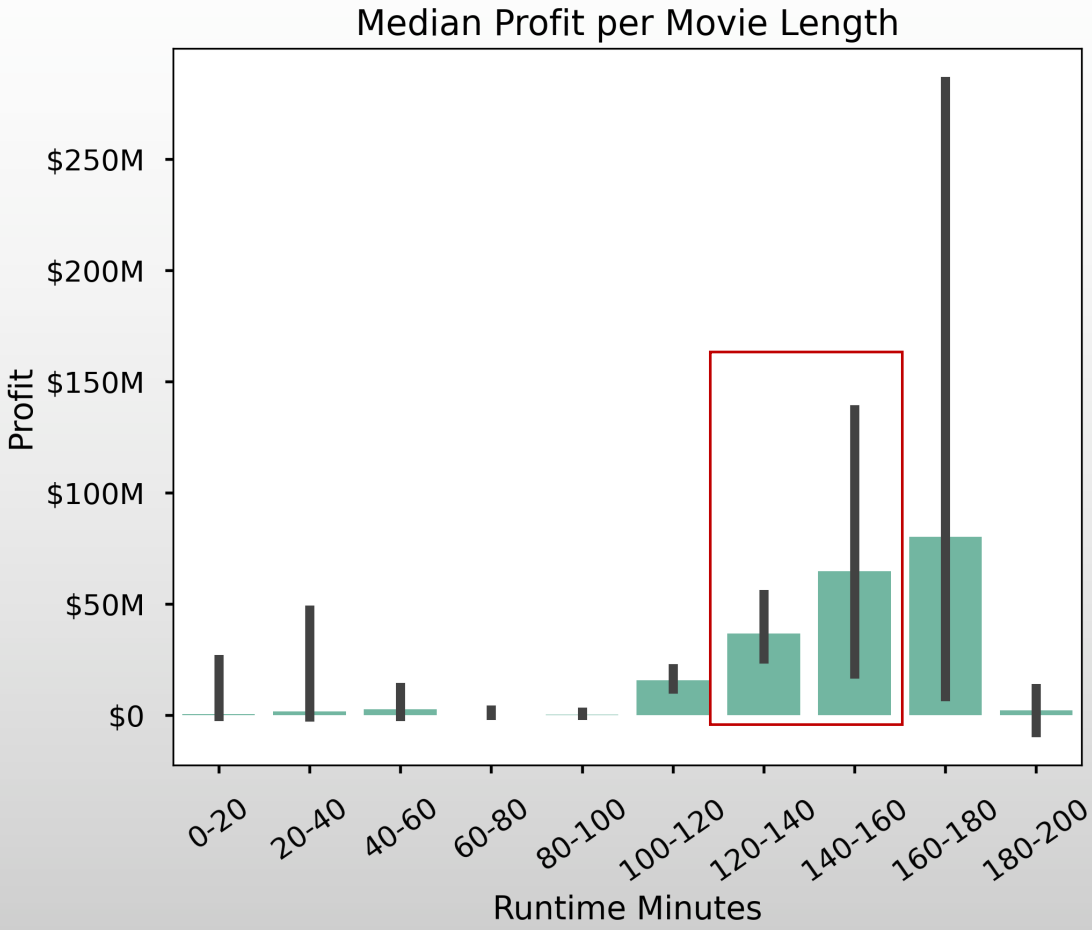




- For highest ROI and Profit release the movie in **November**.
- If you miss November, do NOT release in December, wait for **January**.
- **Summer months** are the next best options.

Which movie length to focus on?





For the highest Roi and Profit target **120-160** min length for the least risk. This is a bit longer than 2 hours.

Conclusions

- ❑ Focus on **Animation, Sci-Fi** and **Adventure** for high budget movies and **Thriller** and **Mystery** for lower budget movies.
- ❑ Hire **James Wan, Pierre Coffin, Chris Renaud,** or **Bryan Singer** as directors.
- ❑ Release the movie in **November**, in **January** or in the **summer**.
- ❑ Make movies **slightly longer than 2 hours**.

Limitations and Improvements

- ❑ Small sample size - due to lack of budget and gross information. API calls or wed scraping?
- ❑ Movie names not coded the same way in different datasets - perform a more rigorous cleaning.
- ❑ Lots of outlier movies making the statistical analyses more challenging.
- ❑ Need more information about Microsoft's allocated budget?

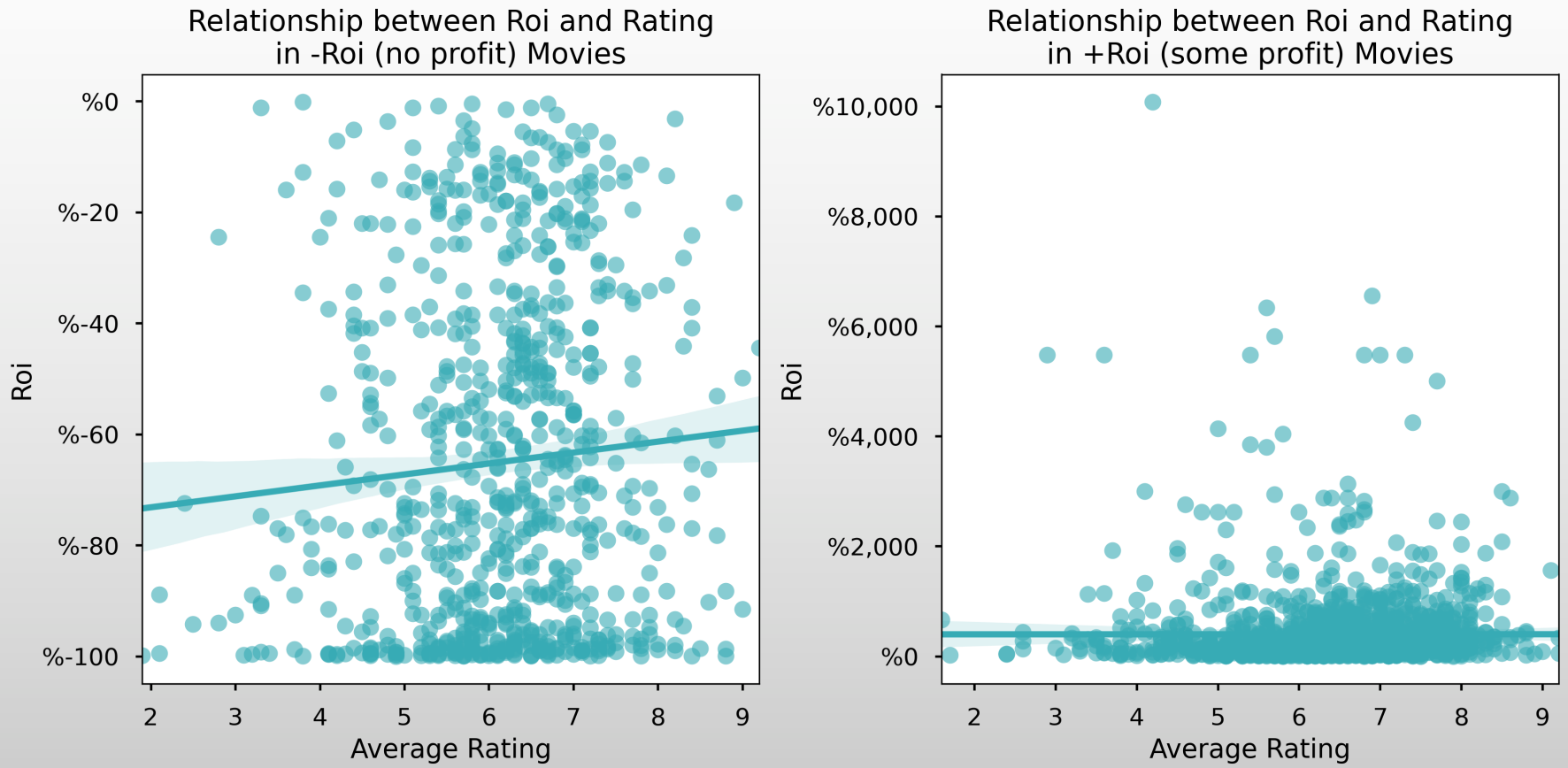
Email: erdemiraysu@gmail.com

GitHub: @erdemiraysu

LinkedIn:
linkedin.com/in/aysuerdemir/

Appendix I

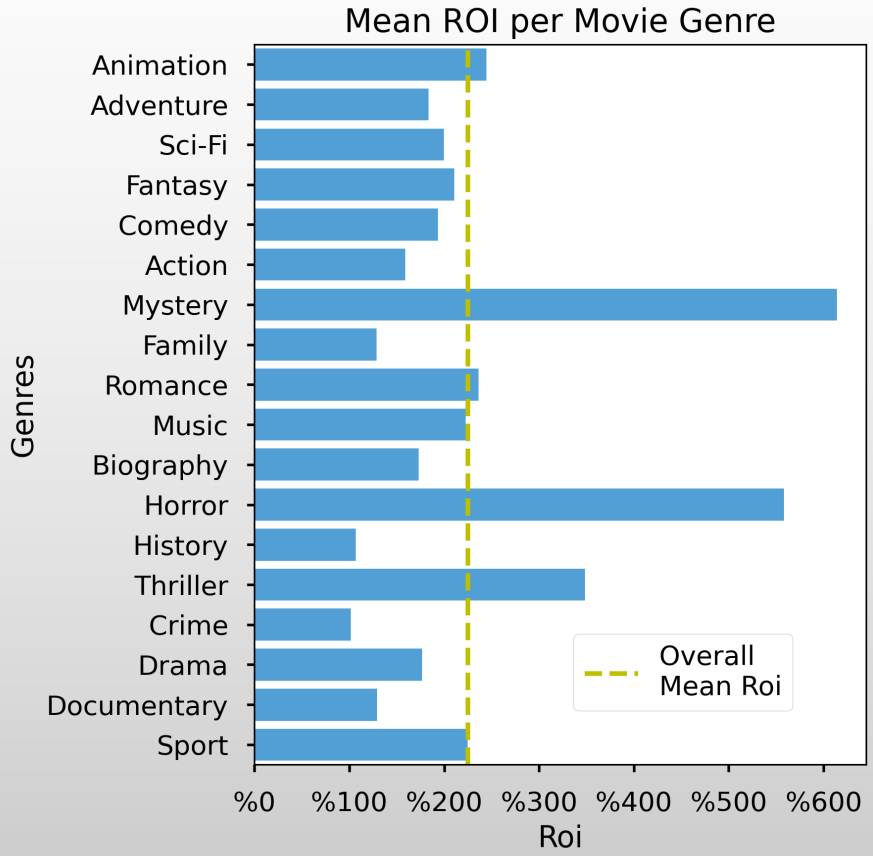
Ratings could be misleading for assessing a movie's success:

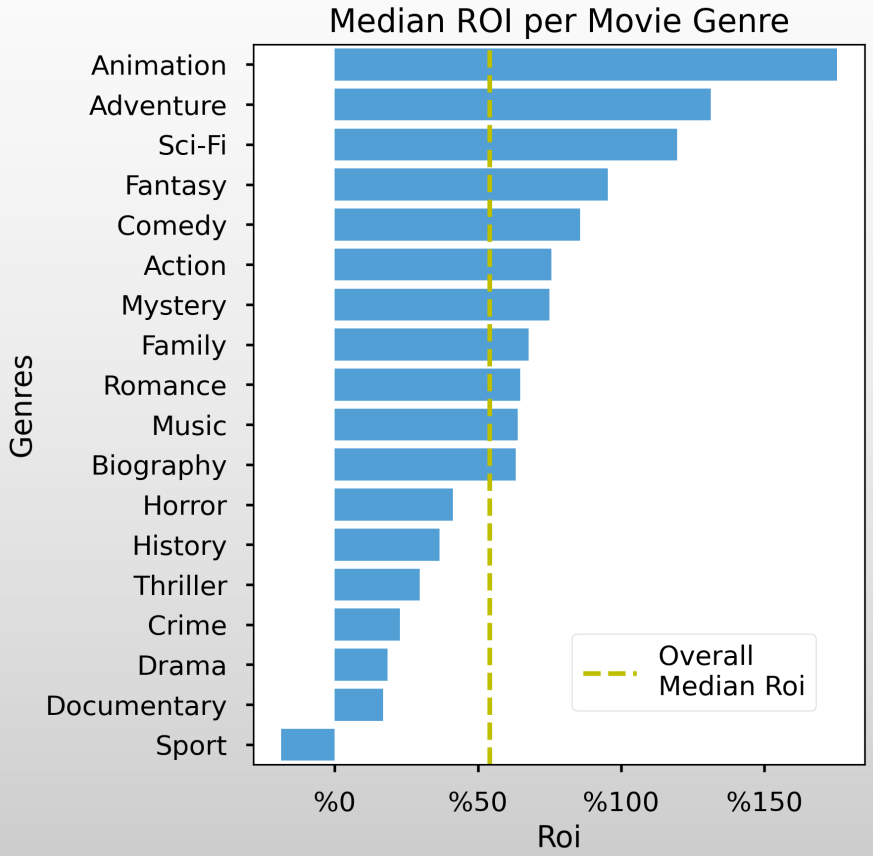


Do NOT rely on ratings when assessing a movie's success especially if the movie makes a profit.

Appendix II

Using MEAN can lead to misleading results like this:





Use **MEDIAN** as a measure of central tendency due to:

- The skewness of the distributions.
- The presence of outlier movies with extremely high success.