

TOM CRUISE CANNOT LOSE

Flatiron Data Science FLEX Phase 1 Project
Victor Kang

Summary

- Analyzed Financial Performance movie data from publicly-available sources to identify the most successful movies out of the most recent 5000 released movies to-date.
- Provided data-driven **budget, genre, cast & crew recommendations** giving the best chance at making a financially successful debut movie.



Outline

Business Problem

Data

Methods

Results

Conclusions



Business Problem



Business Problem



Data

THE NUMBERS

kaggle

Box Office Mojo
by IMDbPro

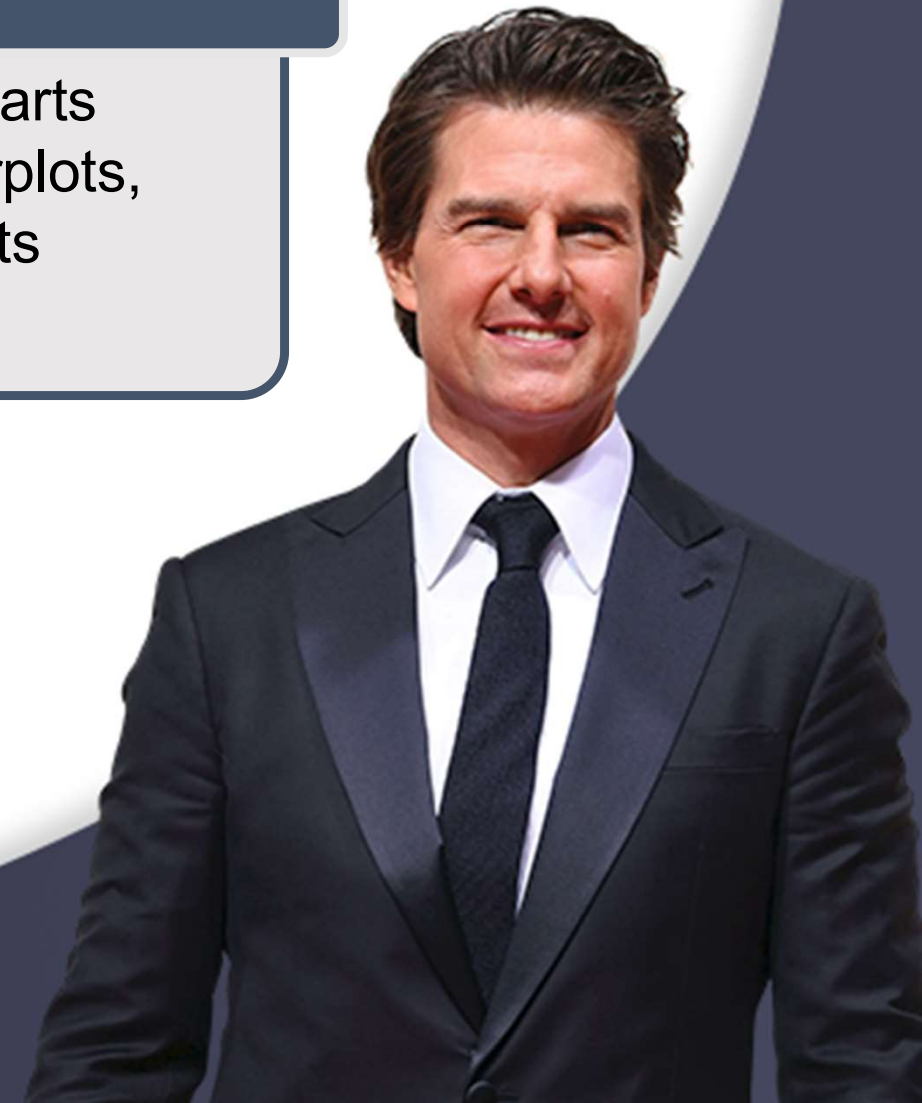
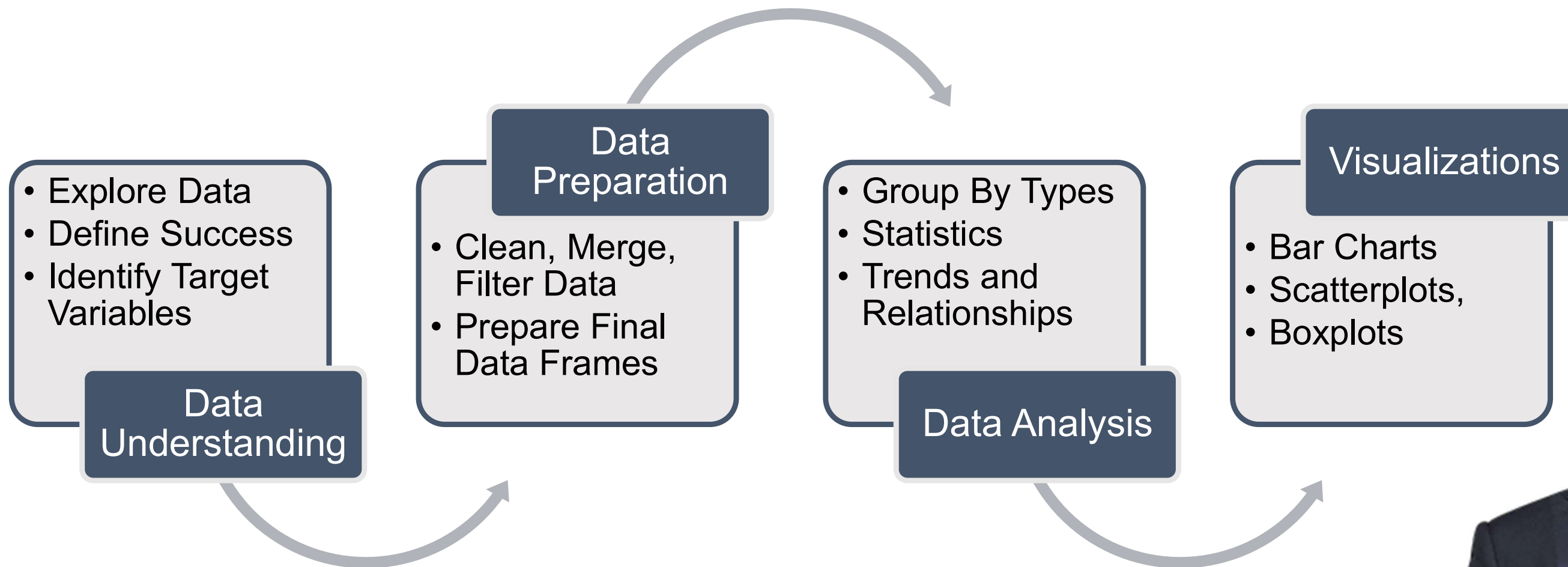
IMDb



Methods



Methods



Methods

“...what **types** of films are **currently** doing the **best at the box office**?”

Data Understanding

- Explore Data
- **Define Success**
- **Identify Target Variables**

currently

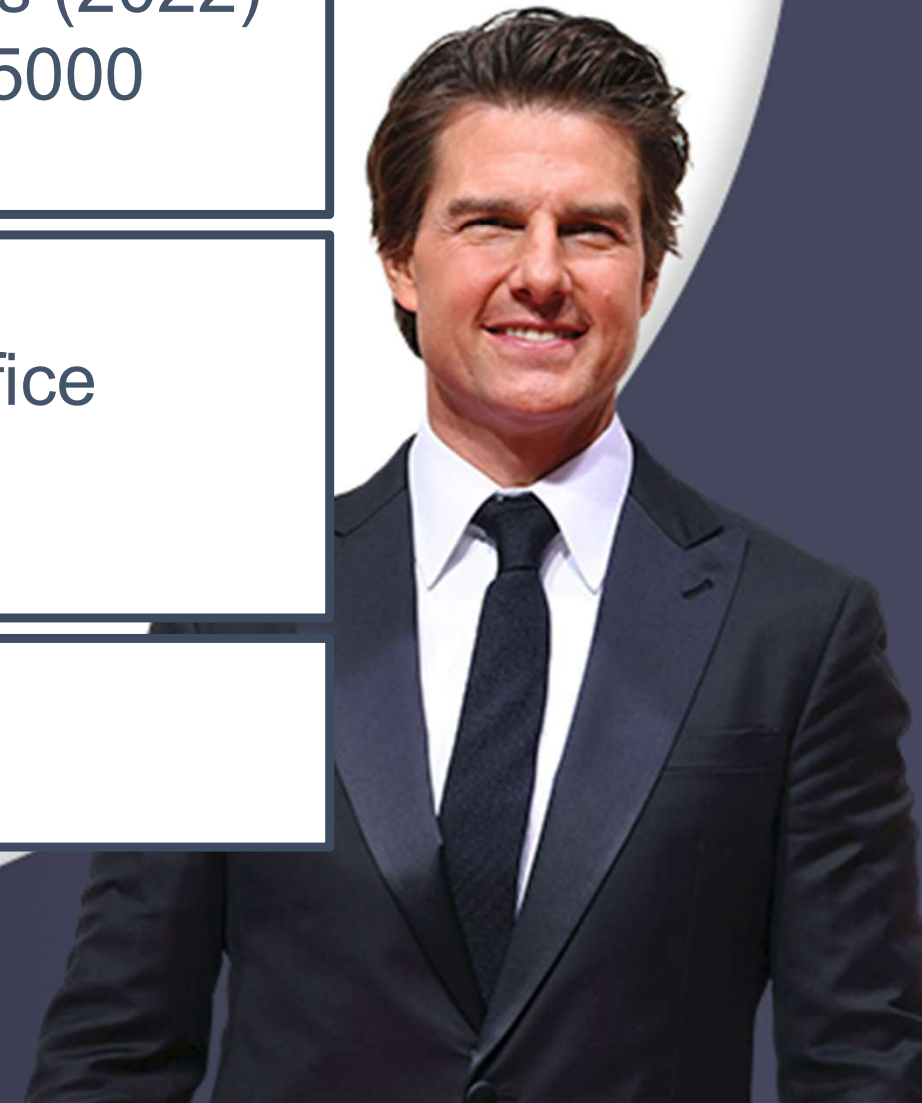
- Modern films only
- Must include recent films (2022)
- Most recently released 5000 films (1999 – 2022)

best at the box office

- Financial Performance
- Worldwide Total Box Office
- Profit / Loss *
- Return-On-Investment *

types

- Production Budget
- Genres



Methods

Data Preparation

- Clean, Merge, Filter Data
- **Prepare Final Data Frames**

Most Recent Movies

- Sample Size: 5000

PnL & RoI

- Profit = WW Gross – Budget
- RoI = Profit / Budget

Top
10% in
Profit



Top
20% in
RoI

Best Recent Movies

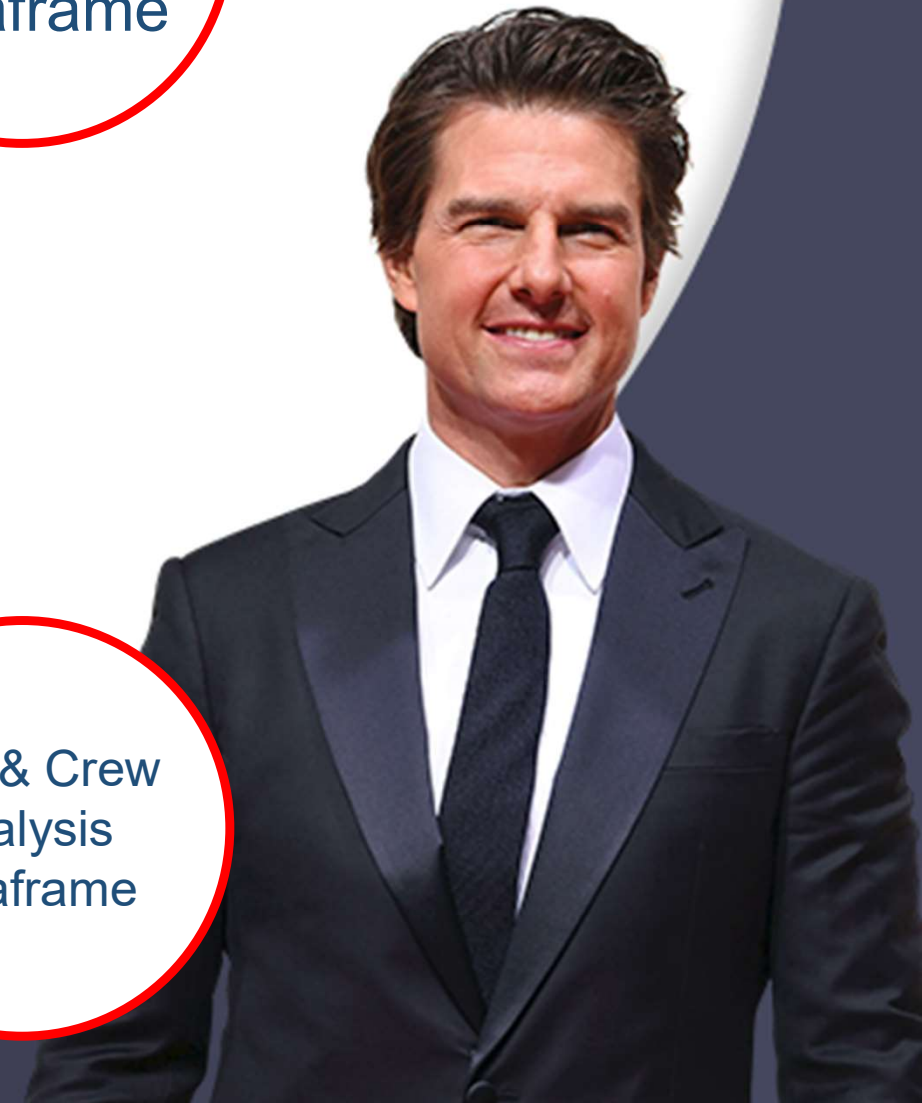
- Sample Size: 308

Merge
Genres

Genre
Analysis
Dataframe

Merge
Cast &
Crew

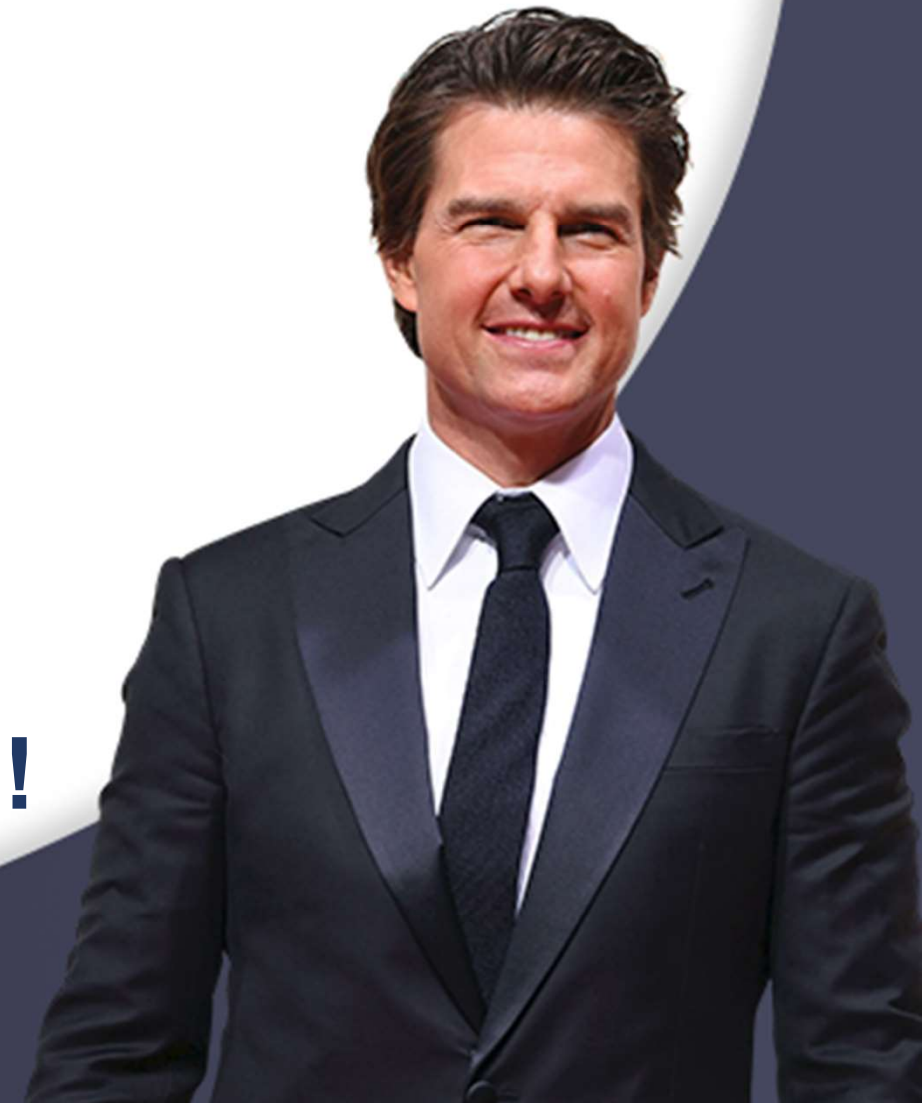
Cast & Crew
Analysis
Dataframe

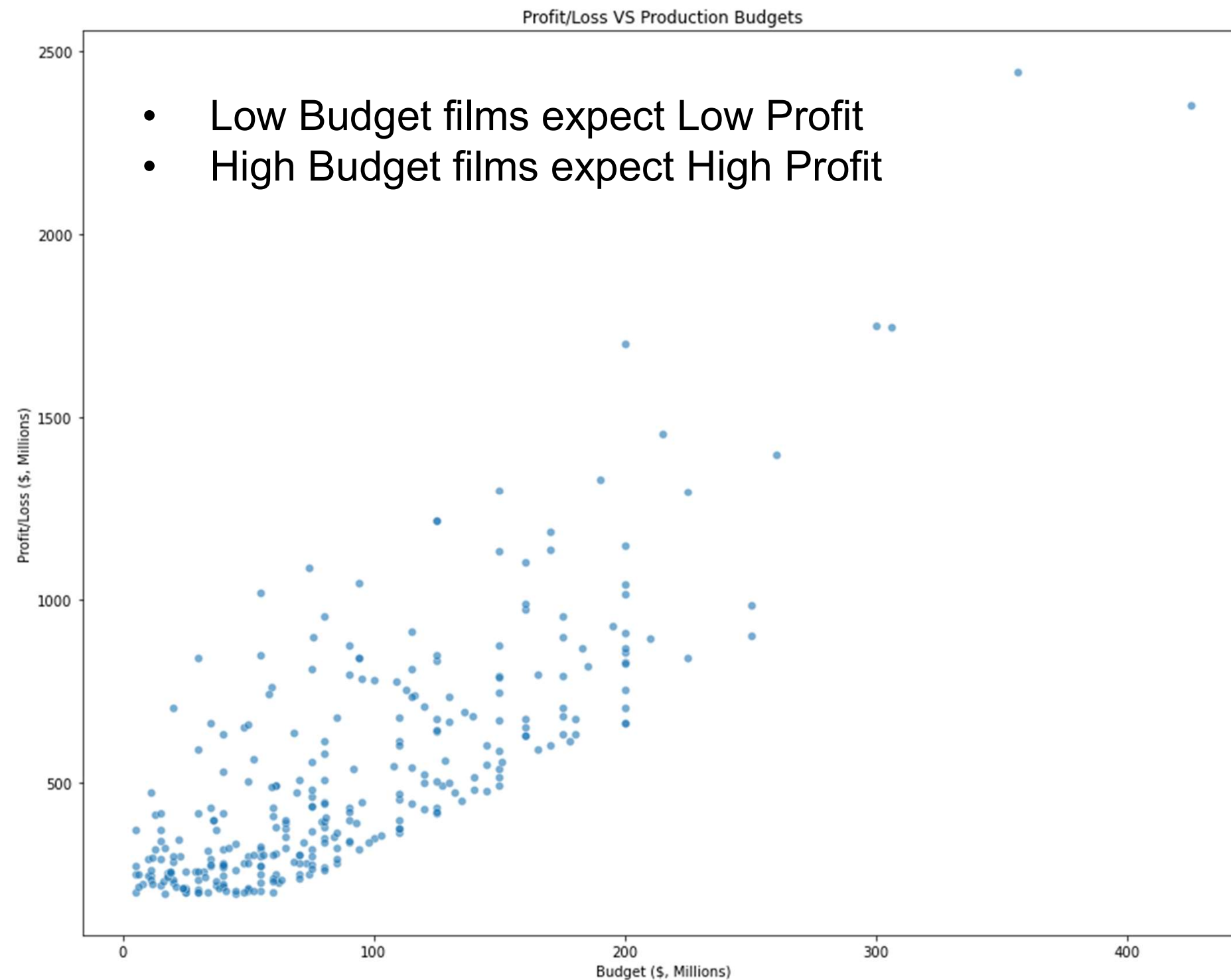


Summary Results

- Production Budget : **~125M**
- Genre Type : **Action & Adventure**
- Cast & Crew
 - Lead Actor : **Tom Cruise**
 - Lead Actress : **Emma Watson**
 - Director : **Anthony Russo**
 - Writer : **Christopher Markus**

See Visualizations for Data Driven analysis!

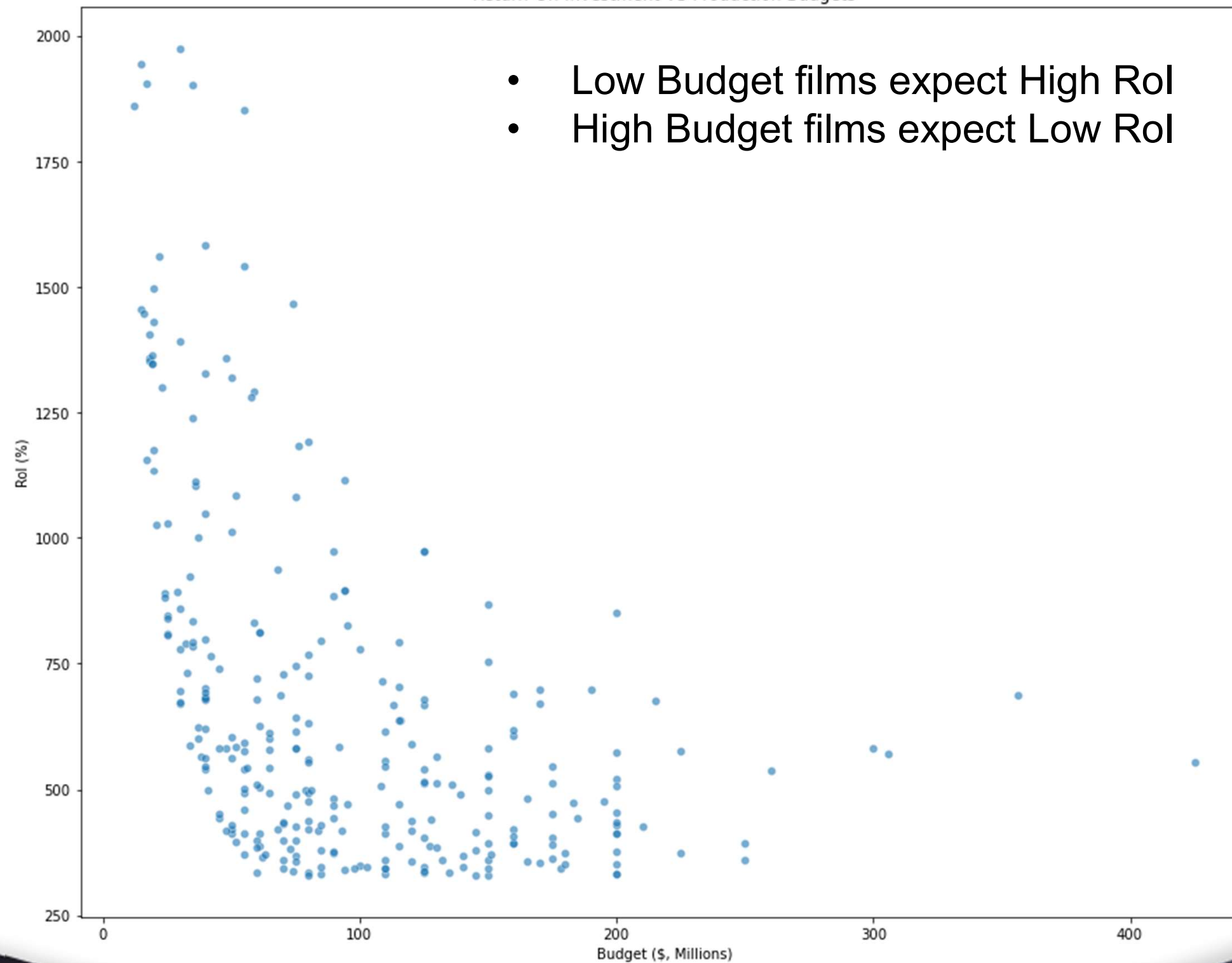




- Strongly Correlated Positive Linear Relationship
- Pearson Coeff : **0.8**



Return-On-Investment VS Production Budgets

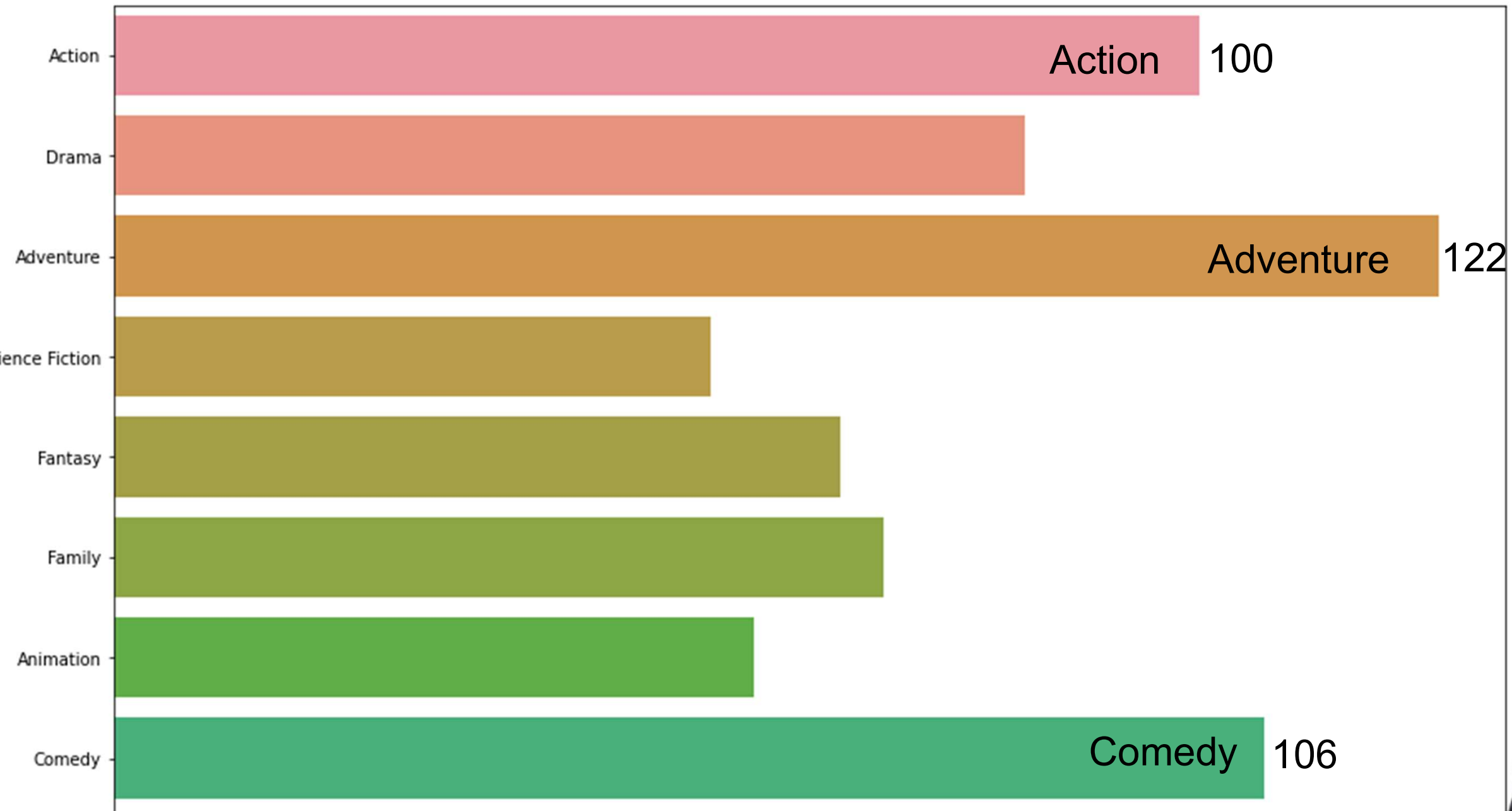


- Low Budget films expect High RoI
- High Budget films expect Low RoI

- Inverse Non-Linear Relationship
- Above **125M** budget, RoI is expected to drop below the mean RoI of 845%.

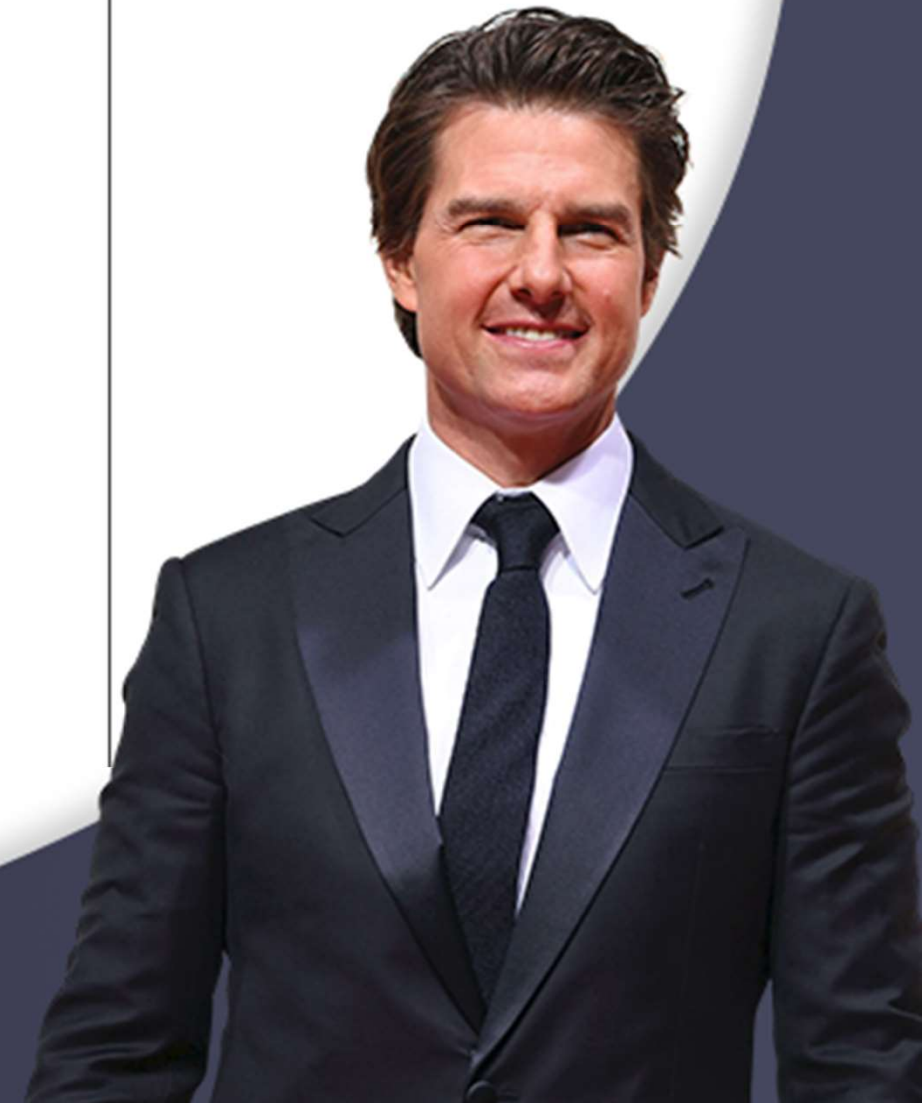


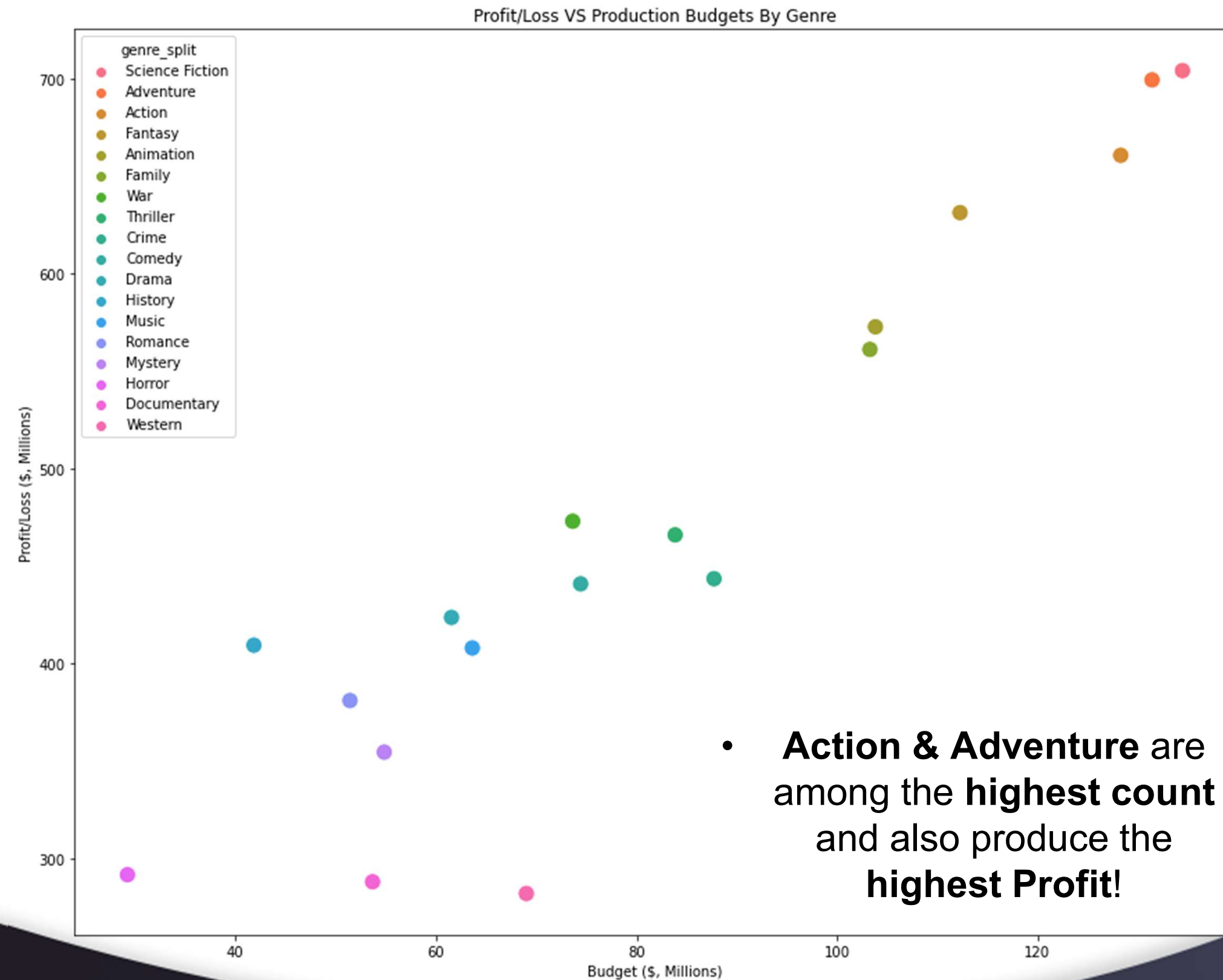
Genre Counts of Successful Movie Pool



Sample Size:
883 Exploded
Genre-Split Movies

18 Genre types

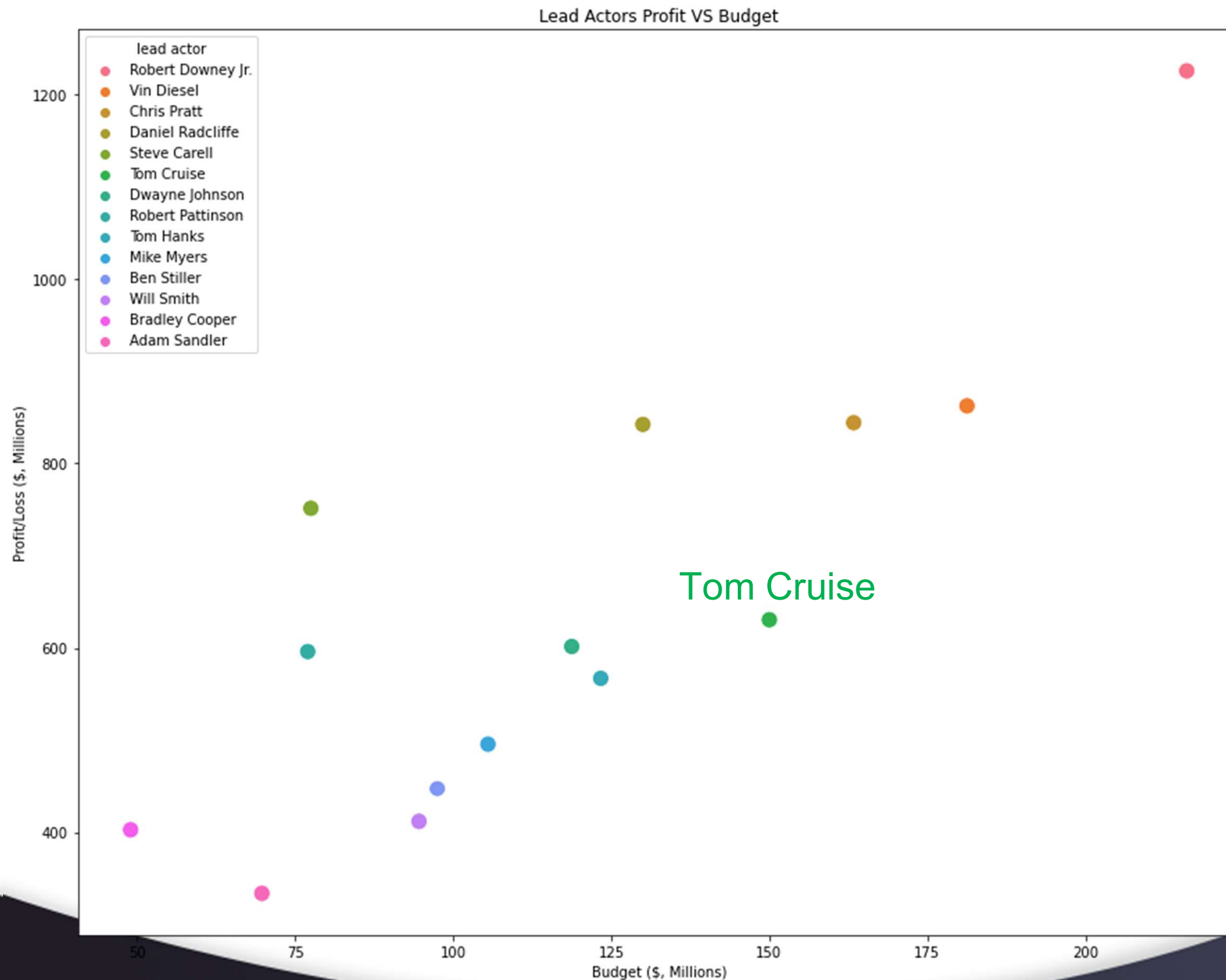




- **Action & Adventure** are among the **highest count** and also produce the **highest Profit!**

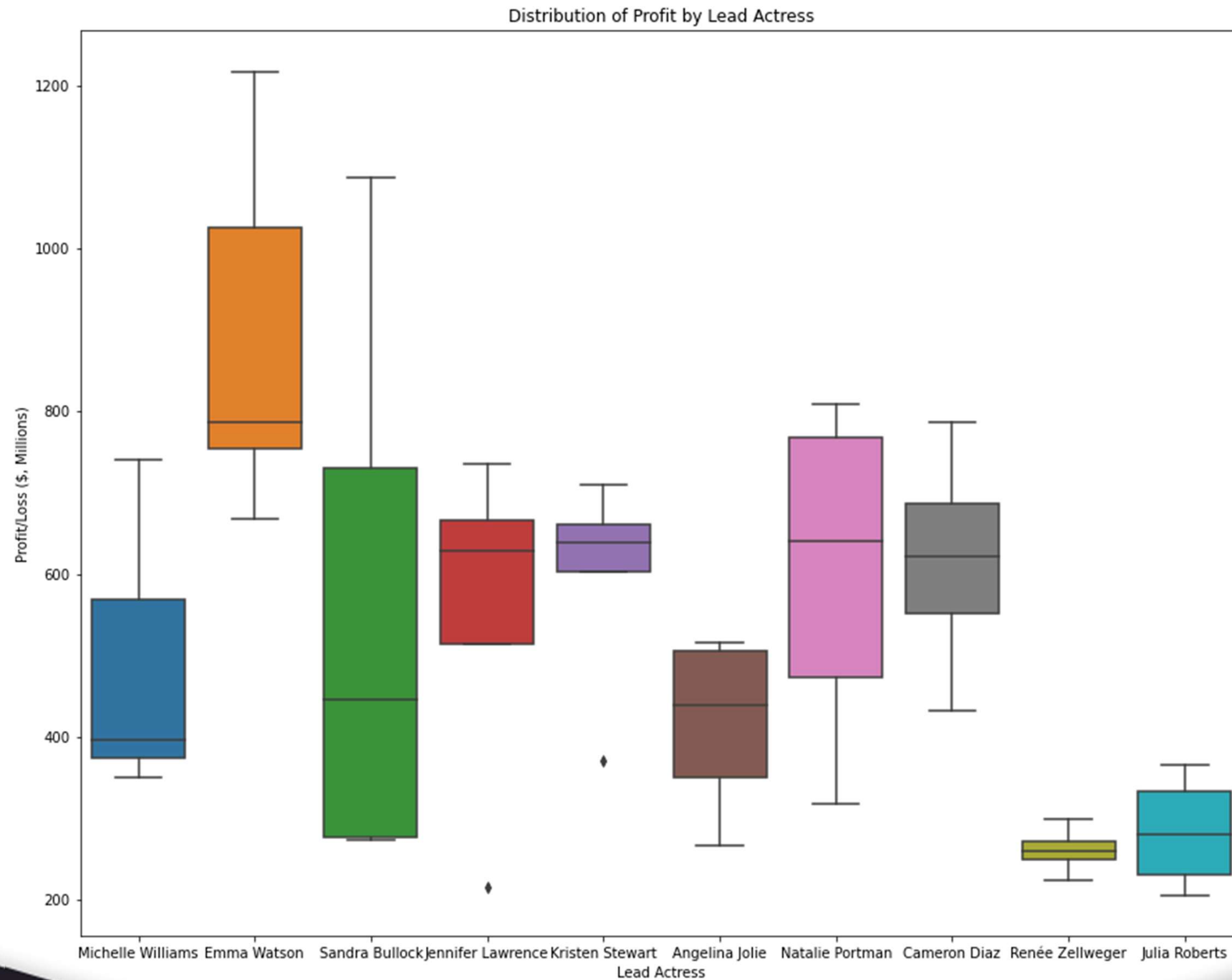
- **Sci-Fi, Adventure, and Action** movies average the most profit. They also typically have larger budgets





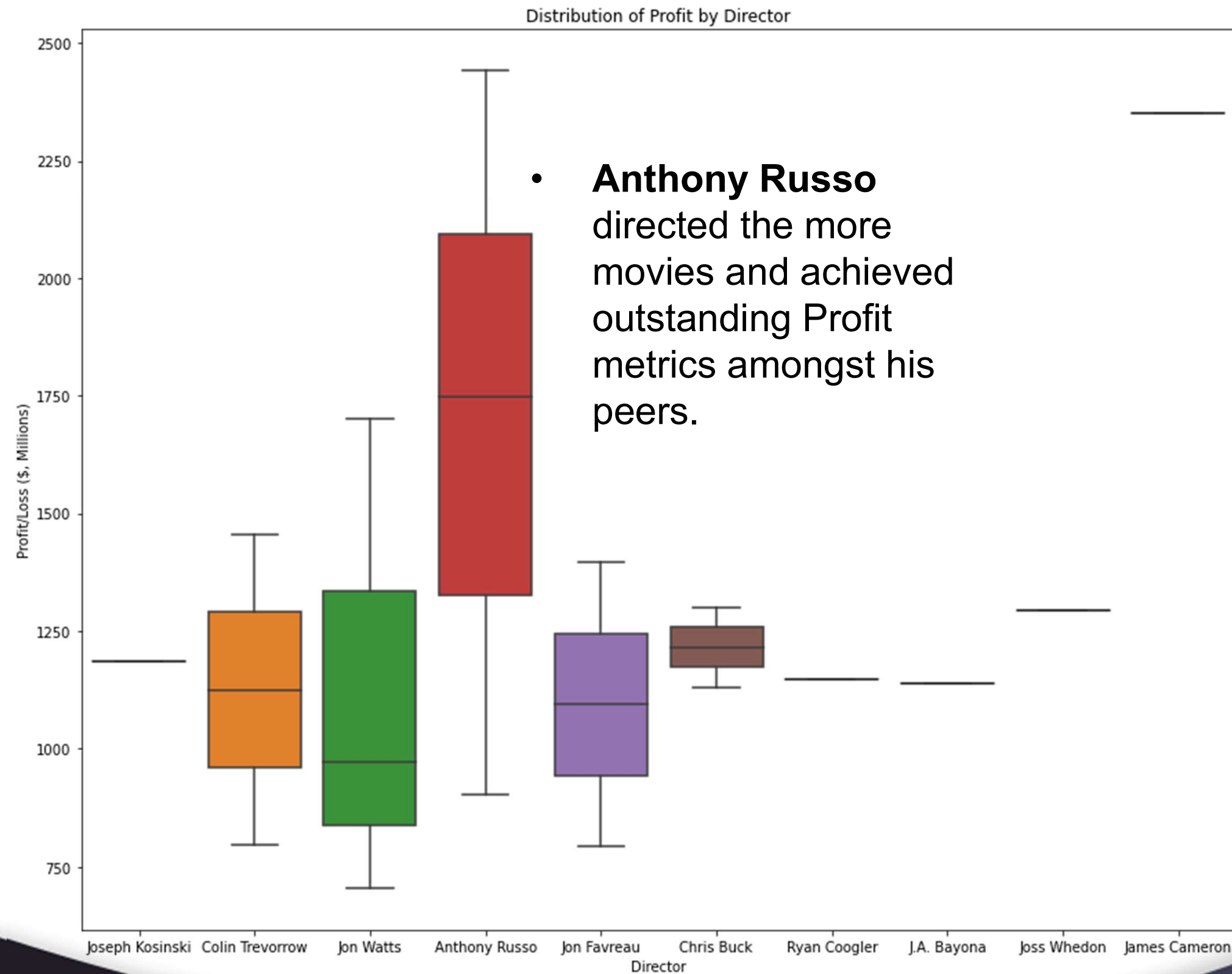
- Of the 14 highest count actors, **Tom Cruise** has the best proven track record of producing the most profitable **Action & Adventure** films within our budget range of **\$125M**.





- **Emma Watson** outperforms her peers on both median and mean Profits ~\$800M!

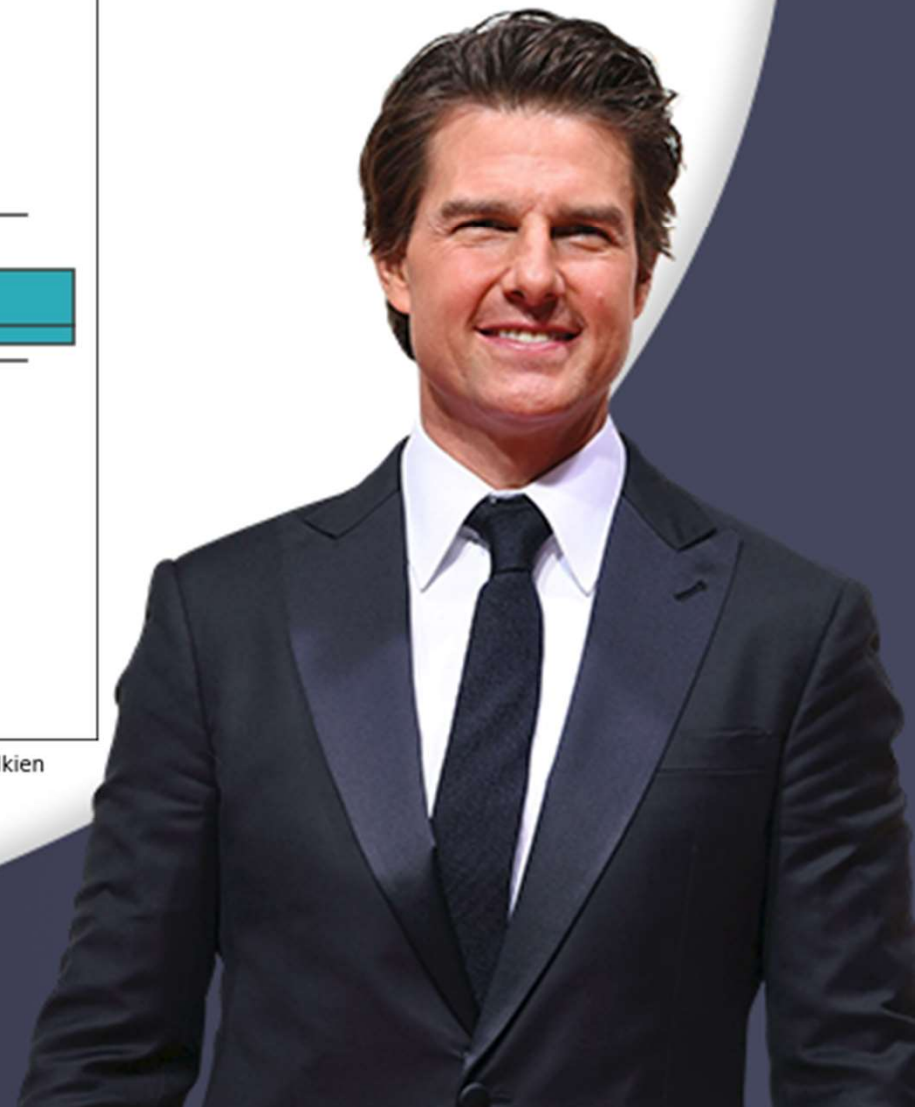
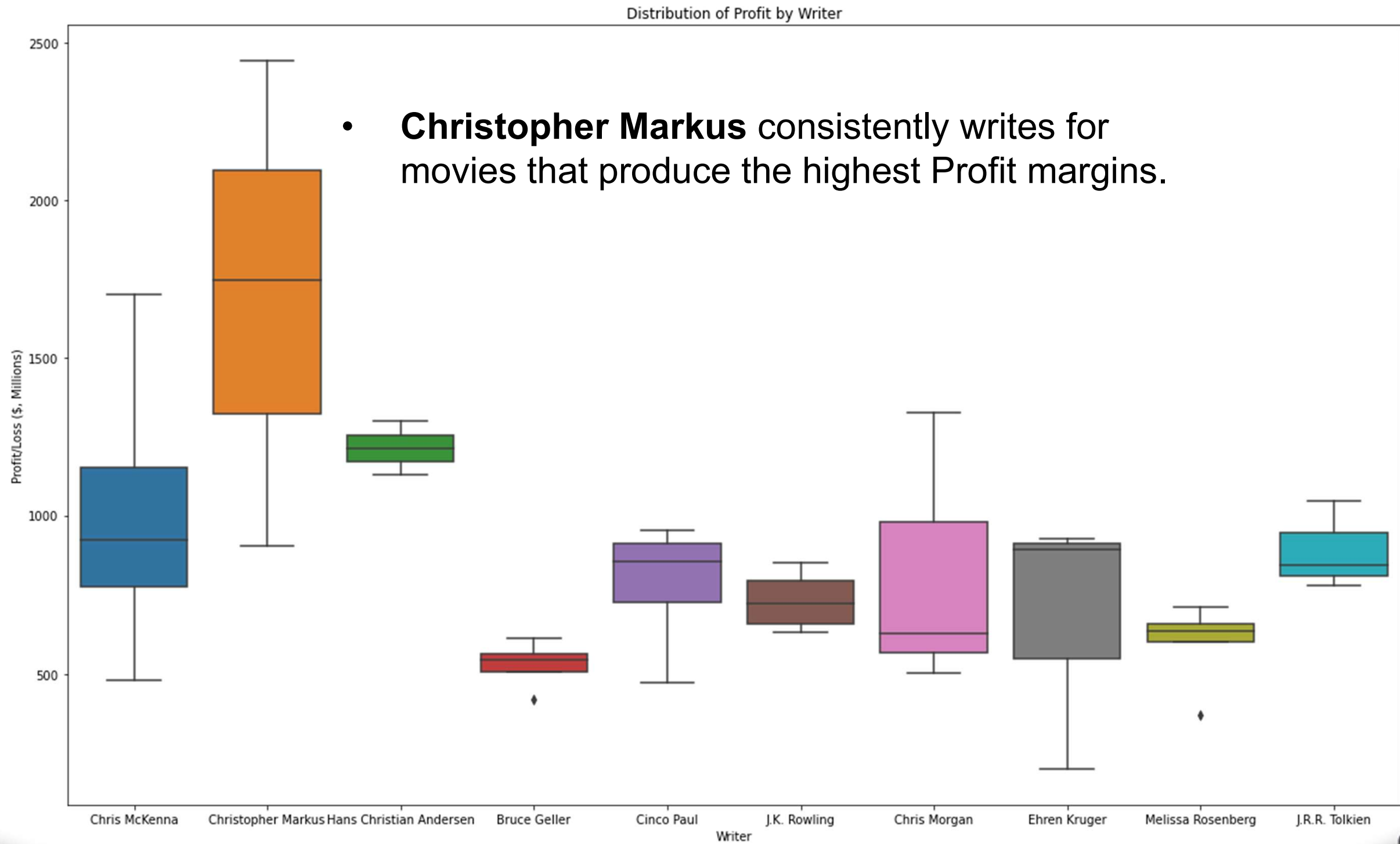




- **Anthony Russo** directed the more movies and achieved outstanding Profit metrics amongst his peers.

- **James Cameron** has the highest average Profit, but his sample size is only 1 movie! That's too small a sample!





Conclusion

Based off the data analysis, I would recommend to Microsoft's new Movie Studio to secure a **production budget of ~\$125M** to make the next **Action & Adventure** movie starring **Tom Cruise** and **Emma Watson**, with **Anthony Russo** directing a story written by **Christopher Markus**.

TOM CRUISE CANNOT LOSE



Thank You!

Email: mr.victorkang@gmail.com

GitHub: @vict0rkang

LinkedIn: [linkedin.com/in/vict0rkang](https://www.linkedin.com/in/vict0rkang)