



# Film Performance Exploratory Data Analysis

AIDAN O'KEEFE

# Overview

- ▶ Microsoft to create a new movie studio.
- ▶ What **types** of films are **currently** doing the **best at the box office**?
  - ▶ **Types** = Genres
  - ▶ **Currently** = last 10 years
  - ▶ **Best** = highest ROI with highest average ratings
- ▶ **Goal:** deliver actionable **insights** to help decide what **type of films to create**.

# Business Understanding

- ▶ Microsoft wants to **spend the least amount of money**
  - ▶ **Test Run**
  - ▶ **Quantity**
- ▶ And wants **high average ratings**
  - ▶ **Quality**



# Data Understanding

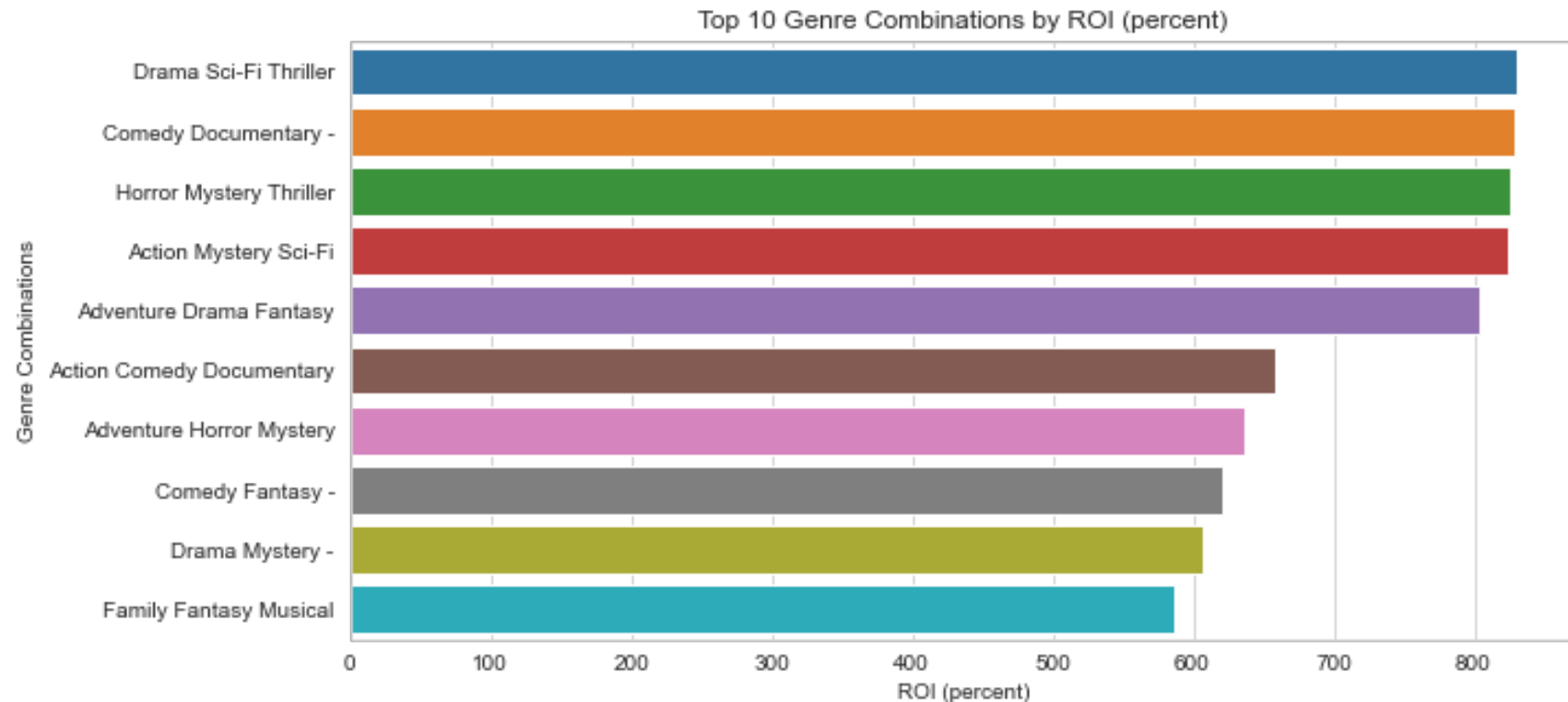
- ▶ **Two Data Sources:**
  - ▶ **IMDB database:** film details and rating information
  - ▶ **The Numbers.com CSV:** budget and sales data
- ▶ Limiting our data to films in the last 10 years where we have financial data.
- ▶ Each record in the data represents a film along with variables such as genres, release year, average rating, vote counts, production budgets, gross sales, director, and more.

# Data Analysis – Genres by ROI

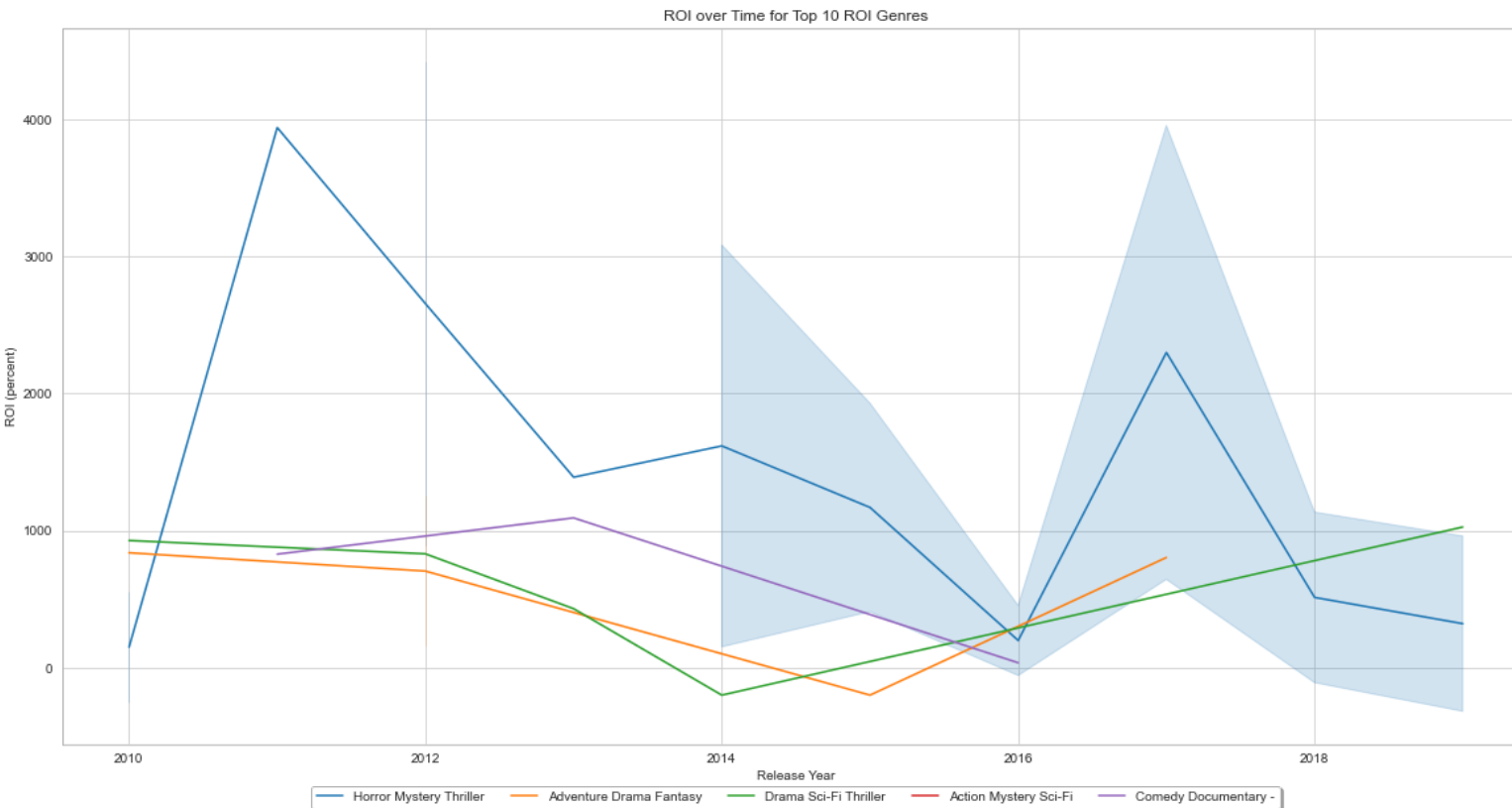
Practically tied among the Top 5 Genres at over 800% ROI.

Interesting to see which genres appear multiple times in Top 5:

- Drama 2x
- Mystery 2x
- Thriller 2x
- Sci-Fi 2x



# Data Analysis – ROI over Time by Genre



Keeping in mind the distribution of films in each genre (competition).

**Apparent choice:** Horror Mystery Thriller

- Downward trend
- High competition

**Best choices:** Drama Sci-Fi Thriller or Adventure Drama Fantasy

- Upwards trend
- Low competition

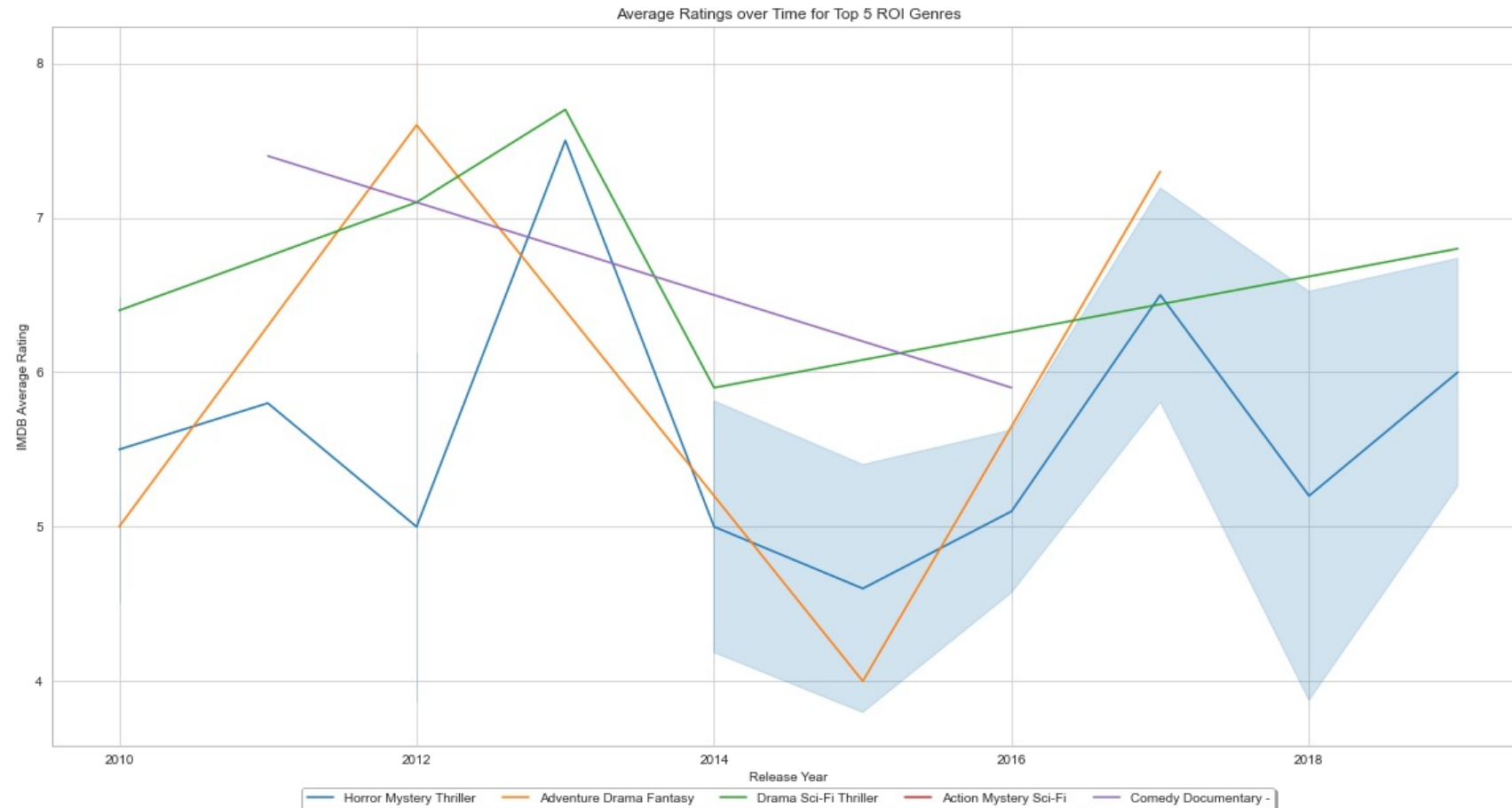
# Data Analysis – Average Ratings of Top Five ROI Genres

## Adventure Drama Fantasy

-often "hit or miss" with ratings oscillating from 7.5 to 4.

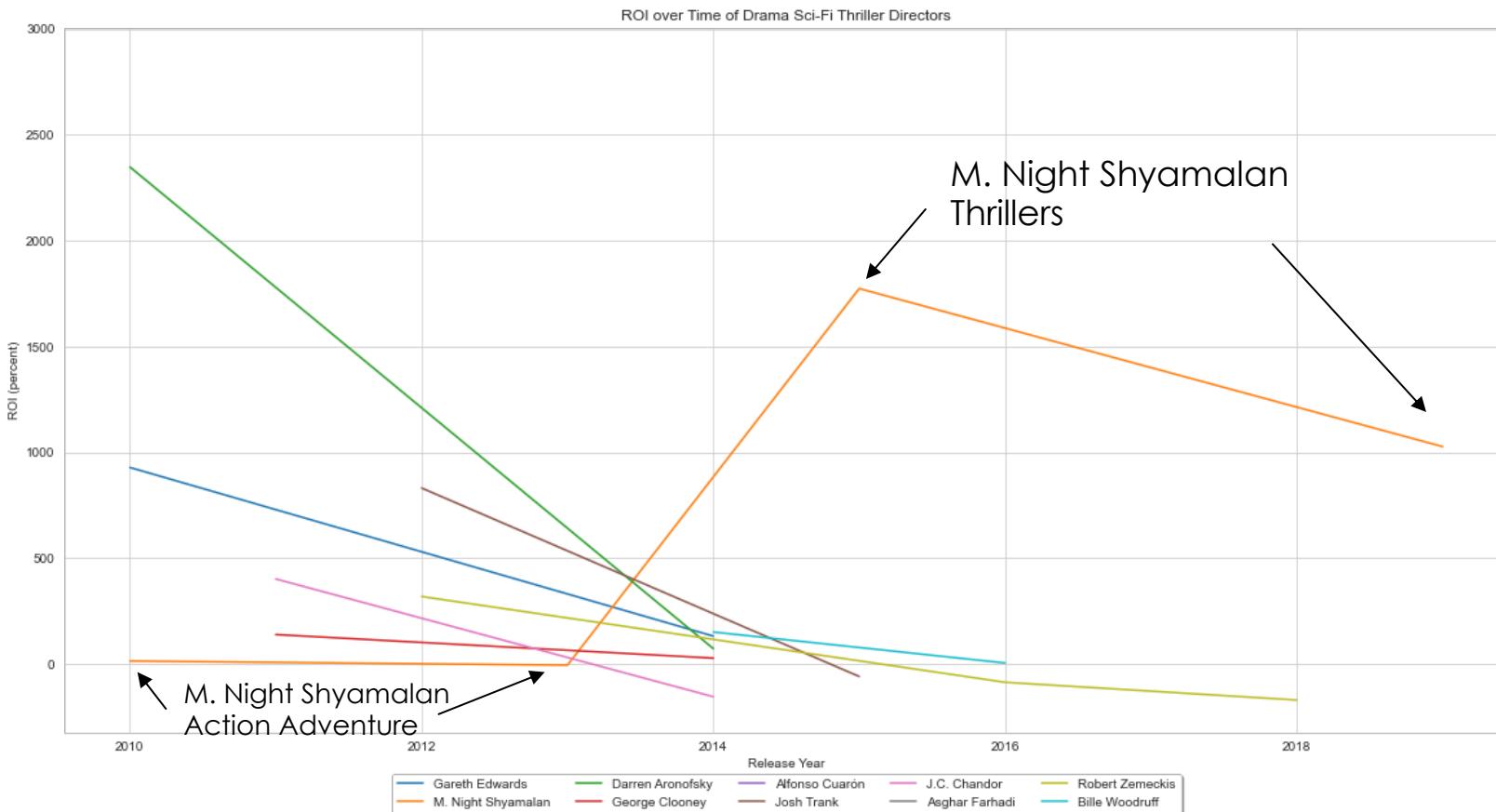
## Drama Sci-Fi Thriller

-the most consistent and highest average rating.





# Data Analysis – Directors of Drama Sci-Fi Thrillers



Based on ROI over time of all movies by DST Directors, I recommend:

1. M. Night Shyamalan (Glass)
2. Gareth Edwards (Monsters)
3. Darren Aronofsky (Black Swan)
4. Asghar Farhadi (A Separation)
5. Alfonso Cuarón (Gravity)

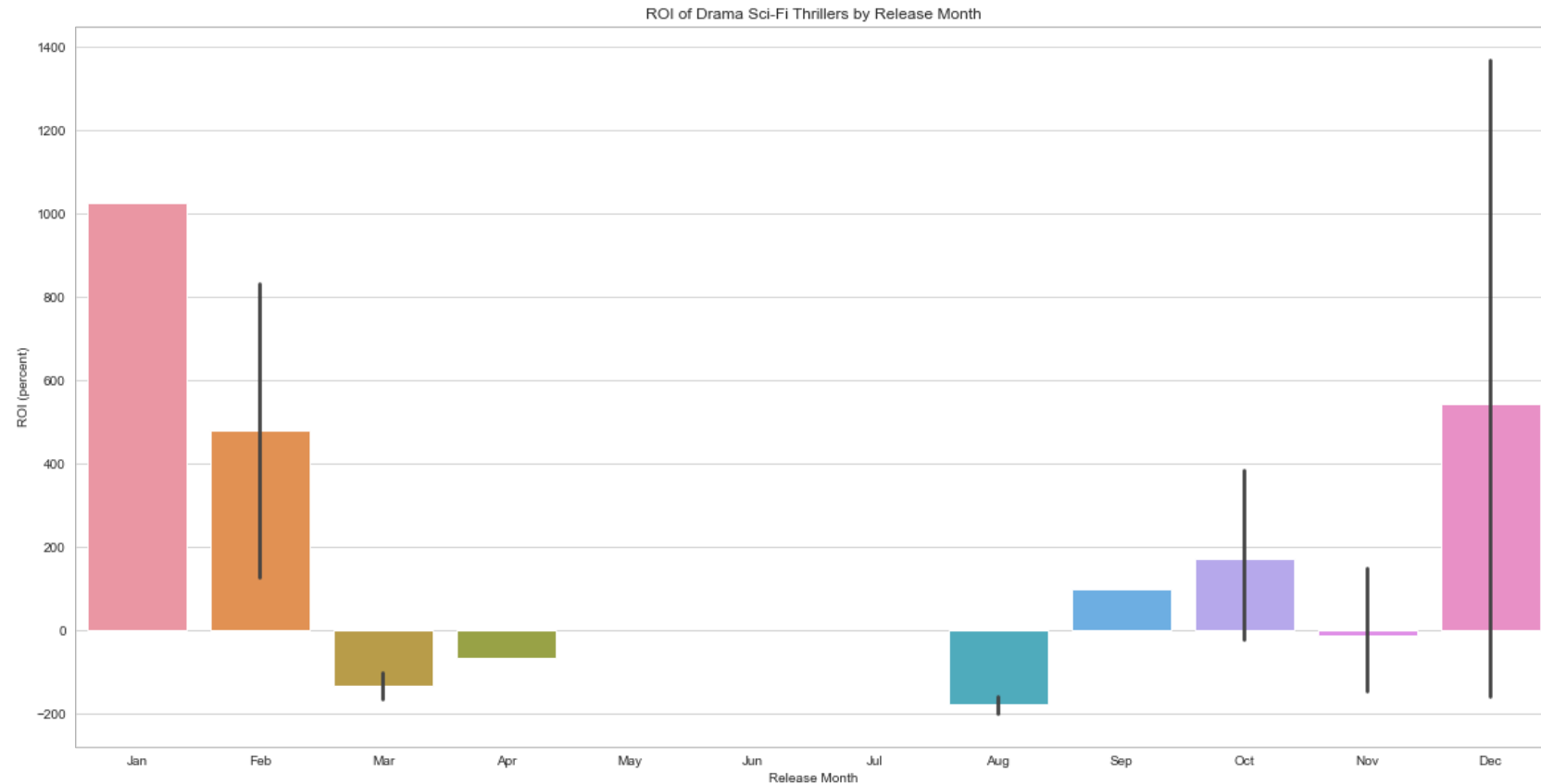


# Data Analysis – Best release time for ROI

## Top 3 Months for ROI for Drama Sci-Fi Thrillers:

- January
- December
- February

- Do not release DST films in March, April, or August



# Recommendations

- ▶ What **types** of films are **currently** doing the **best at the box office**?

Genre: **Drama Sci-Fi Thriller**

e.g. Ex-Machina, Gravity

Director: **M. Night Shyamalan**

Release Time: **Winter** months (Dec – Feb), specifically January

# Next Steps

- ▶ I focused on films for which I had financial data. I would try to **find additional financial data** to import and **increase the size of my data frame** and see if my analysis still held true.
- ▶ Additionally, I would expand my analysis to answer if more genres added onto a movie widens the appeal to audiences. I would look to see if there is a **correlation between** a movie's **number of genres and its profitability and/or ratings**.





# Thank you!

## Questions?

AIDAN O'KEEFE

[HTTPS://WWW.LINKEDIN.COM/IN/AIDAN-O-KEEFE/](https://www.linkedin.com/in/aidan-o-keefe/)