

# Phase 1 Project

---

Flatiron Online Data Science Course

# Summary

This is the first end-of-module project in the Flatiron Online Data Science Course

The project aims at testing and reinforcing the data analysis and engineering concepts taught in phase 1 of the Flatiron School curriculum

Applying the tools of exploratory data analysis (EDA) enables comprehensive analyses resulting in actionable insights

# Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions & Actionable Insights
- Points of Notice

# Business Problem

Microsoft decides to enter the movie industry

In order for this venture to be successful, the head of Microsoft's new movie studio requires data-driven, actionable insights on what type of movies to create



# Data



<https://www.imdb.com/>



<https://www.boxofficemojo.com/>



<https://www.themoviedb.org/>



<https://www.the-numbers.com/>



<https://www.rottentomatoes.com>

# Data



<https://www.imdb.com/>



<https://www.the-numbers.com/>

## Basics

- Title \*
- Runtime
- Genres
- tconst \*

## Ratings

- Average Rating
- Number of Votes
- tconst \*

## Budgets

- Release date
- Title \*
- Production Budget
- Domestic/Worldwide Gross

\* Columns to merge on

# Methods

1. Uploading the data into the Jupyter Notebook in order to take a look which data appears most useful for the task at hand
2. Cleaning the chosen data (NaN, formatting, column titles, etc.)
3. Creating custom columns for prospective analyses (ROI, Budget categories, etc.)
4. Merging DataFrames to enable further analysis (almost certainly requires repetition of step 2)
5. Creating visuals to communicate findings more easily (using Seaborn instead of Matplotlib for enhanced aesthetics)
6. All throughout the project: Create DataFrame snippets or visuals to ensure the desired result

# Results

For the task at hand I have decided upon the following areas to investigate in order to provide guidance and actionable insights for the Microsoft stakeholders:

1. Production Budget
2. Genre
3. Runtime
4. Release Timing
5. Rating



# 1. Production Budget

- Larger production Budget = More successful movie?
- Is a domestic/worldwide gross of 100 million USD better than 10 million USD?

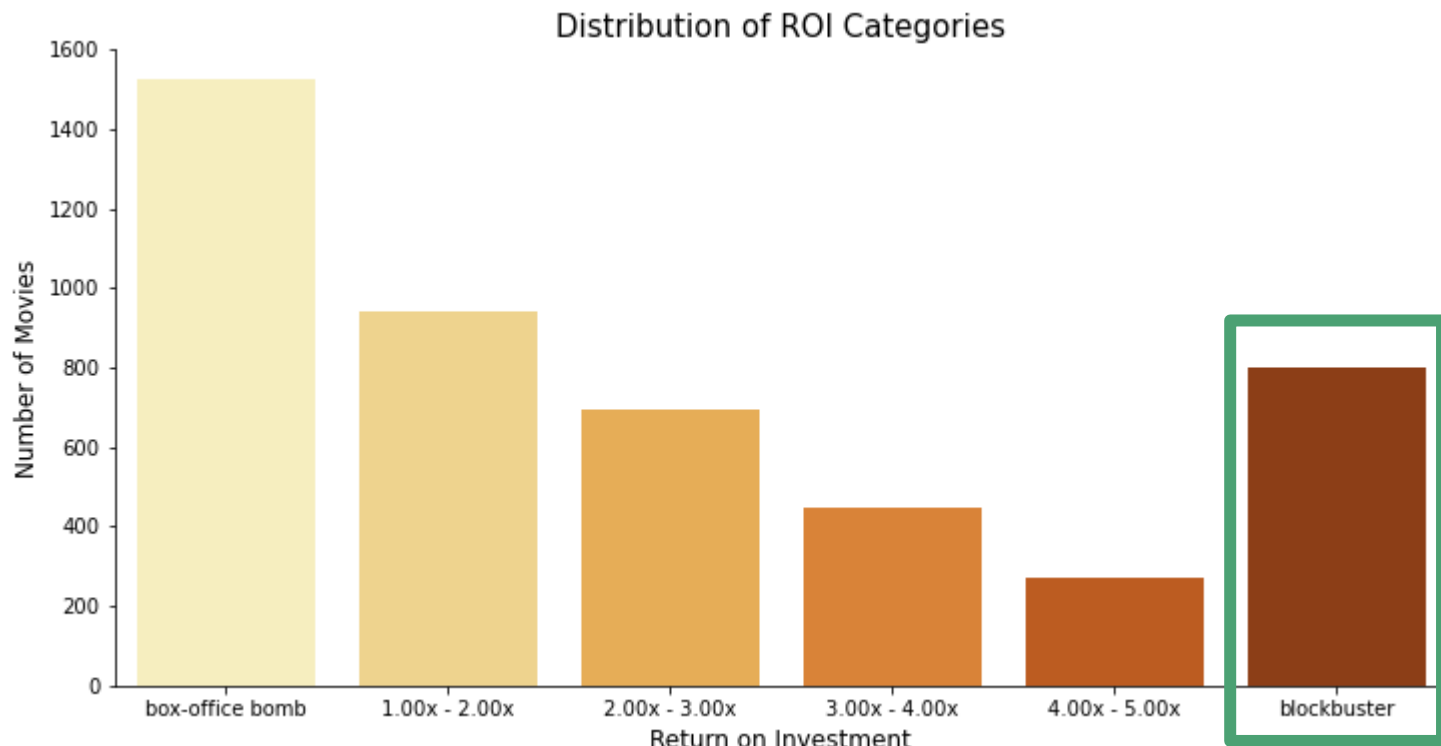
→ „Return on Investment“ (ROI)

- Worldwide Gross / Production Budget
- i.e. „Bang-for-buck“, „How many times have I earned my invest?“

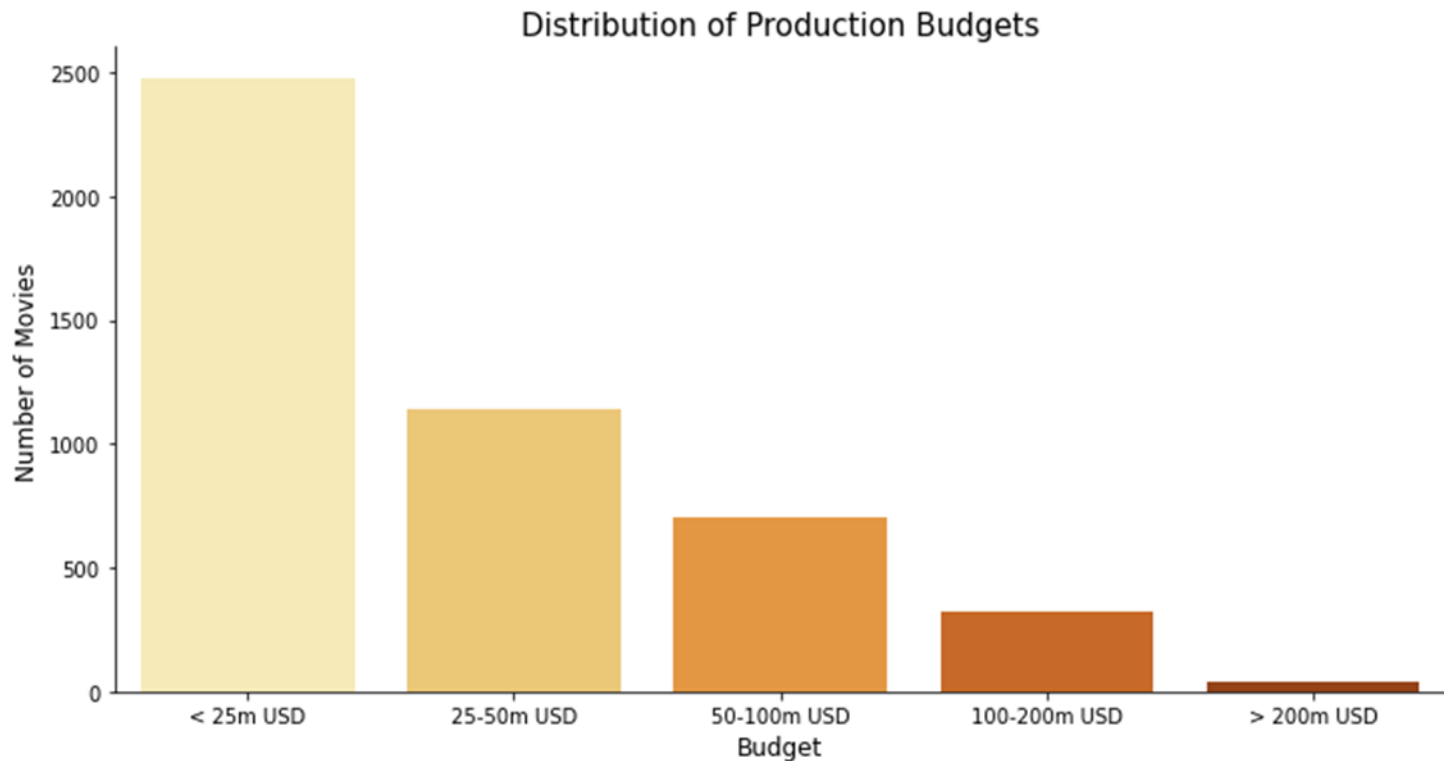
Alternative: Profit Margin

$(\text{Worldwide Gross} - \text{Production Budget}) / \text{Worldwide Gross}$

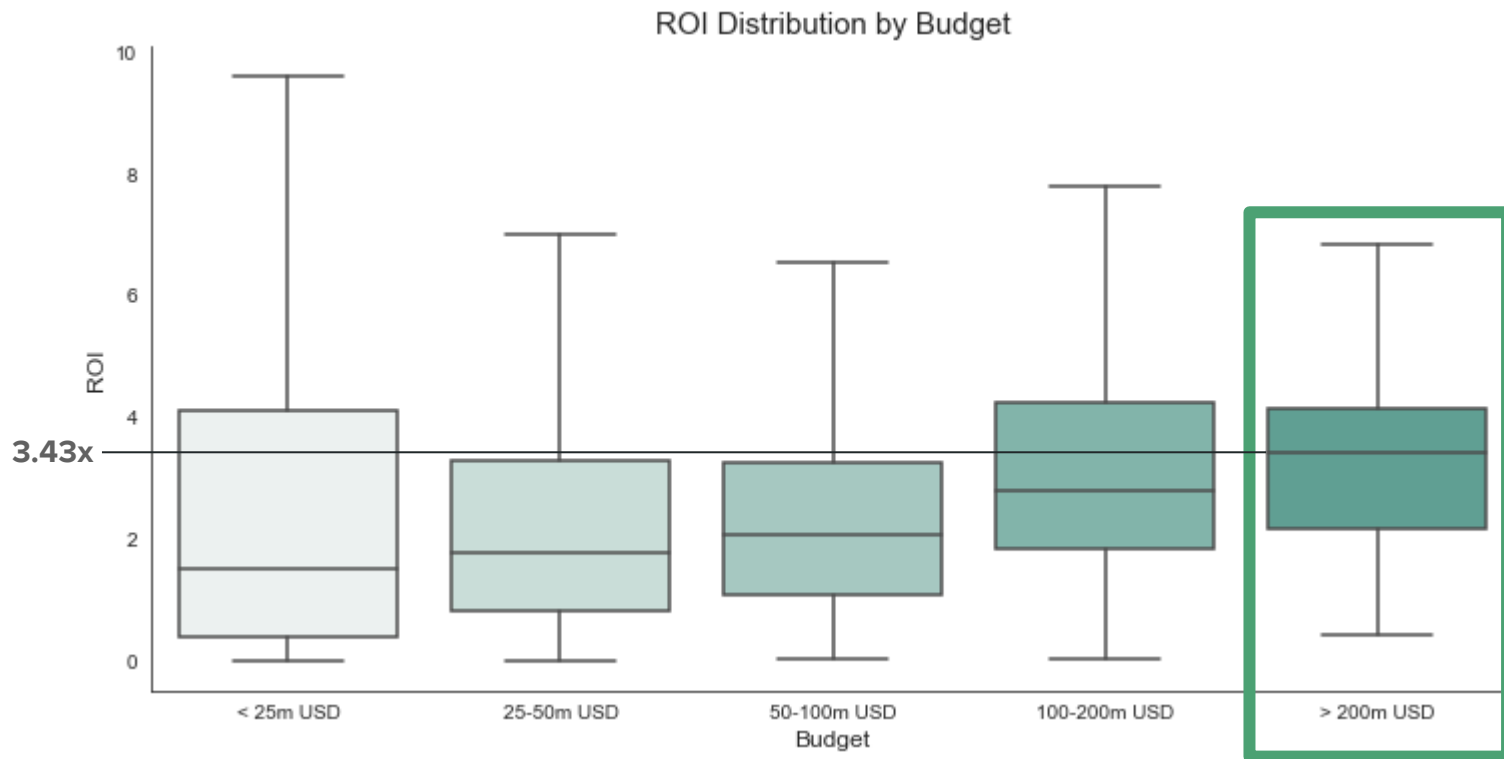
# Return on Investment by Number of Movies



# Production Budgets by Number of Movies



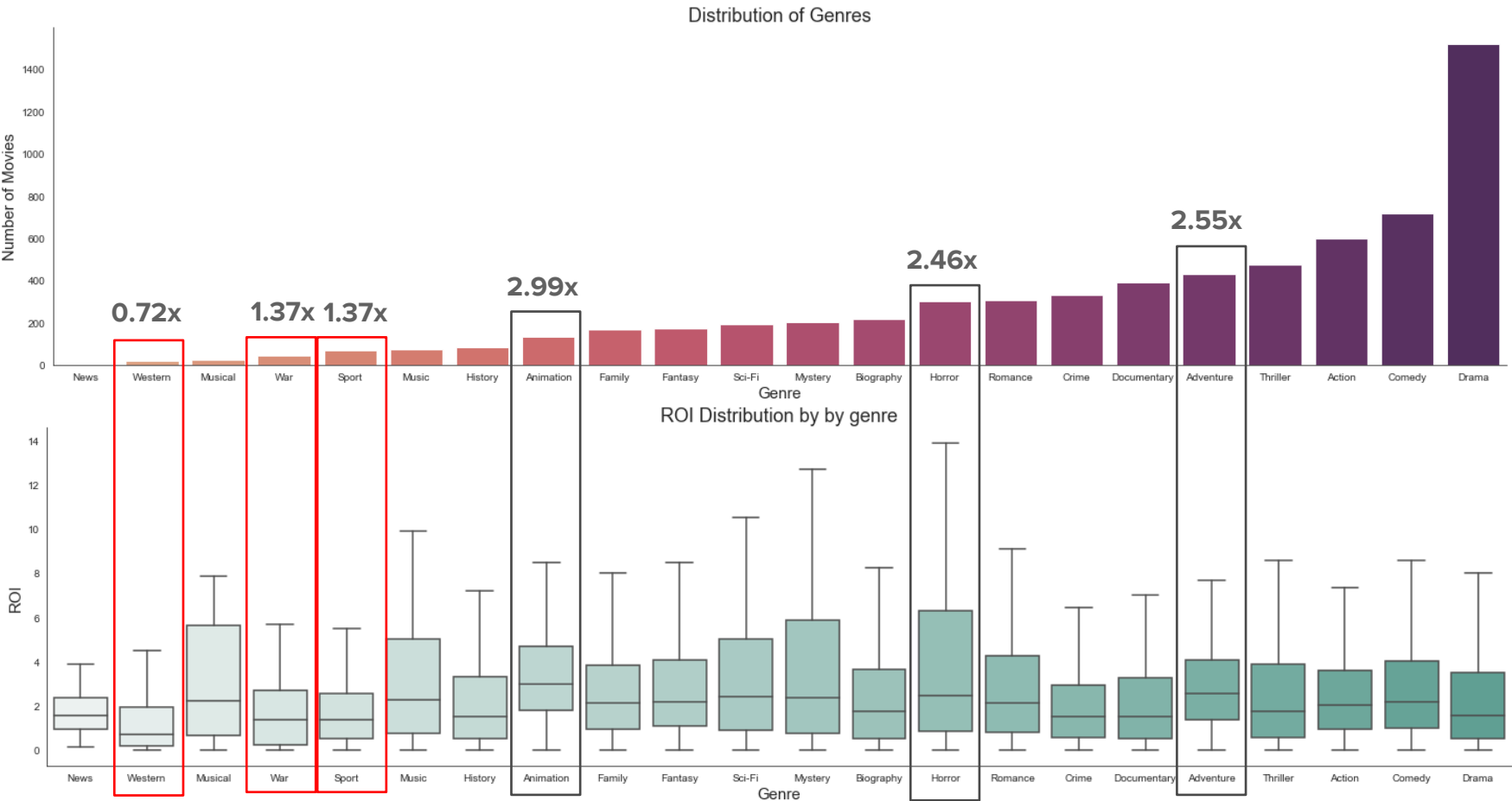
# Profitability by Production Budget



## 2. Genre

- Which genres turn out to be the most successful in terms of profitability?
- An understanding of the most profitable genres will serve as a benchmark for scoping movies and themes

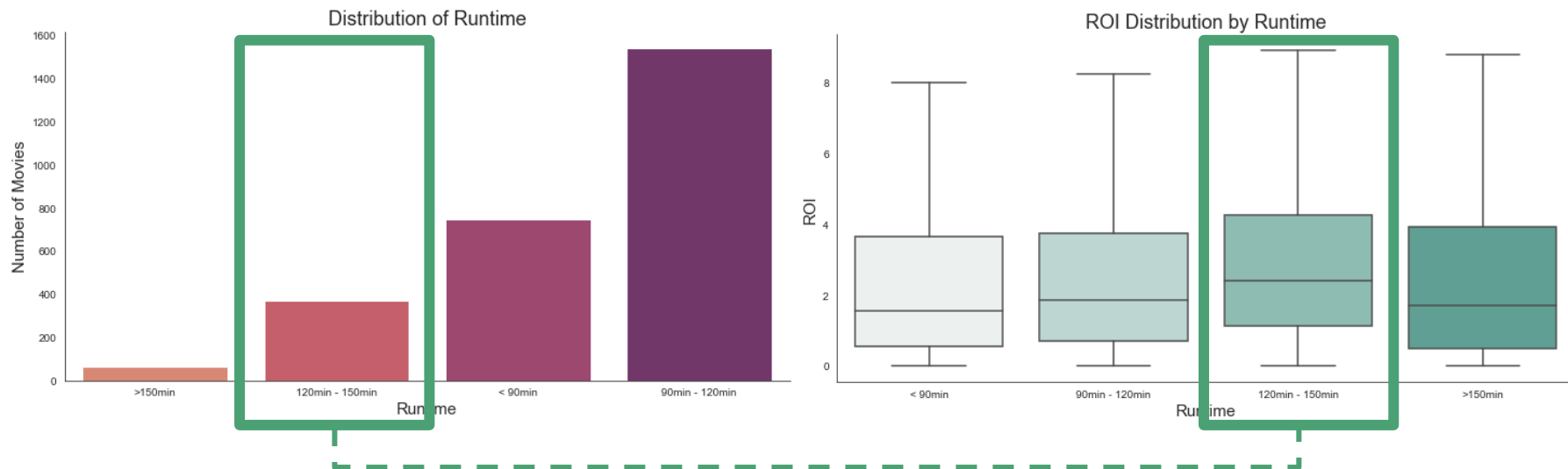
# Distribution and Profitability by Genre



### 3. Runtime

- How long are the most profitable movies?
- Do viewers crave (or financially reward) epic film marathons or rather short, to-the-point narratives?

# Profitability by Runtime

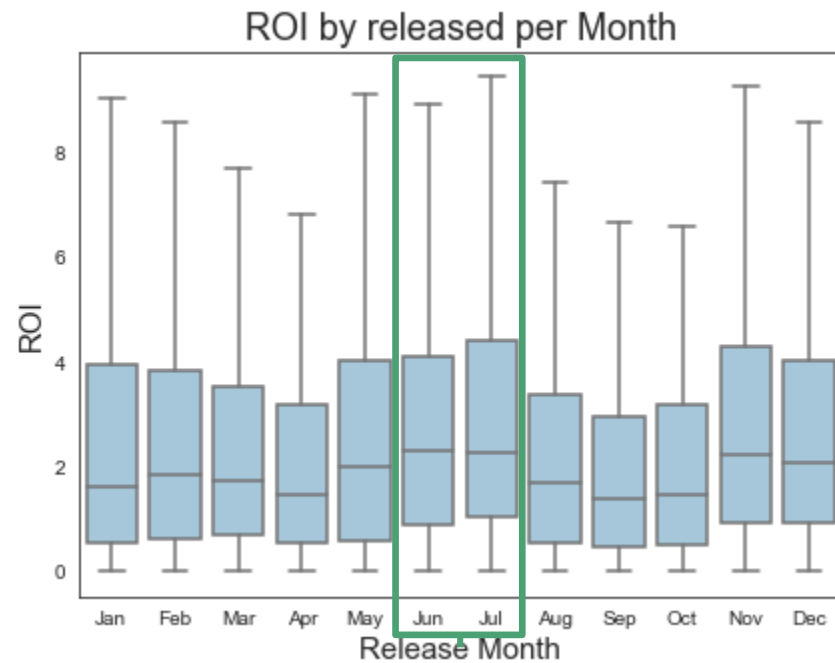
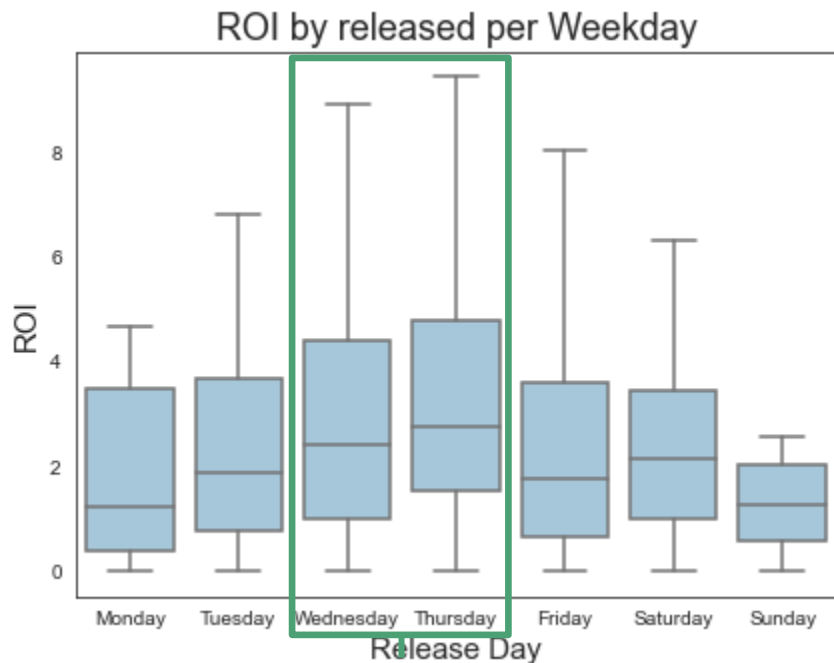




## 4. Release

- What's the most profitable day of the week and month of the year for movies to be released?
- While it can be expected that Microsoft will release movies throughout the year, for expensive or risky movies the timing of a release might make a strategic difference (e.g. considering sequels)

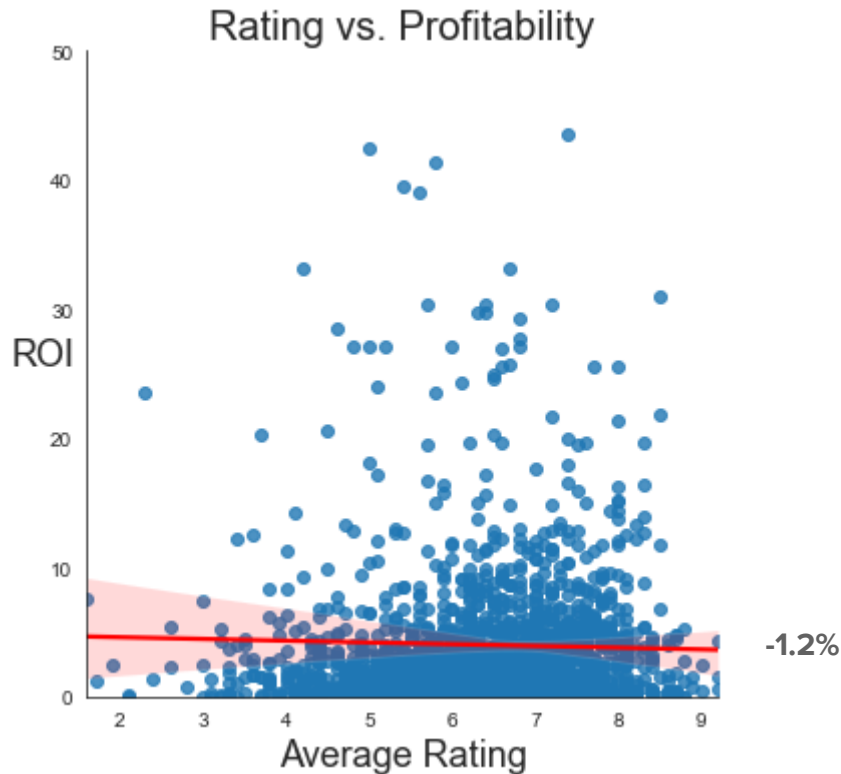
# Profitability by Weekday and Month



## 5. Rating

- Does a good rating imply profitability?
- Is rating of moviegoers too subjective?
- Should Microsoft „listen“ to their viewers, by favoring types of movies with 5-star rating?

# Correlation of Rating and Profitability



# Conclusions & Actionable Insights

## 1. Production Budget

- Higher production budget = higher profitability
- Largest ROI's begin at > 100m USD

## 2. Release Timing

- June and July most profitable release months
- Wednesday and Thursday most profitable release days

## 3. Genre

- Most common genres are not necessarily the most profitable
- Focus on Animation, Adventure and Horror movies while avoiding Western, War and Sports for highest profitability

## 4. Runtime

- 120 – 150 minute movies showcase highest ROI
- Beyond 150 minutes ROI diminishes

## 5. Rating

- Too subjective
- Negatively correlated with ROI

# Points of Notice

- Different Data can lead to different results
- Further analyses to consider:
  - Which producers/directors/actors to engage
  - How to effectively allocated production budget (marketing, staff, merchandise, etc.)
  - A more in-depth look of actual financials
  - Webscraping and use of API potentially lead to further/more recent insights
  - Choice of time frame analyzed (here 1990 – 2021) potentially leads to different results
  - Cross-analysis (e.g. most profitable genre & most profitable budget)

# Thank You!

Email: [kontakt@oliverzimmer.eu](mailto:kontakt@oliverzimmer.eu)

GitHub: @senseize

LinkedIn: [linkedin.com/in/username/oliver-zimmer-cfa-8824881ab/](https://www.linkedin.com/in/username/oliver-zimmer-cfa-8824881ab/)