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 enables comprehensive analyses resulting in actionable insights
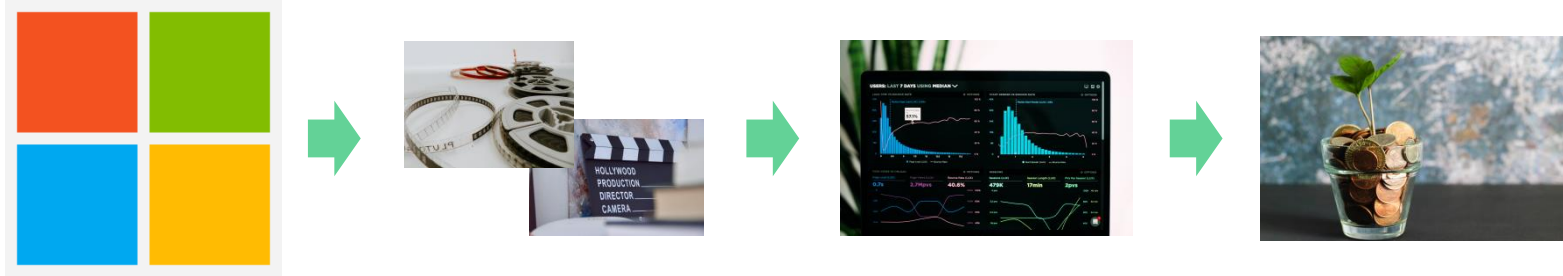
Outline

- **Business Problem**
- **Data**
- **Methods**
- **Results**
- **Conclusions**

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

IMDb

https://www.imdb.com/

Box Office Mojo by IMDbPro

https://www.boxofficemojo.com/

TMDB

https://www.themoviedb.org/

THE NUMBERS ®
Where Data and the Movie Business Meet

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

Rotten Tomatoes ®

https://www.rottentomatoes.com

Data

https://www.imdb.com/

Basics
- Title *
- Runtime
- Genres
- tconst *

Ratings
- Average Rating
- Number of Votes
- tconst *

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

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. Release Timing

3. Genre

4. Runtime

5. Rating

1. Production Budget

- Larger production Budget = More successful movie?

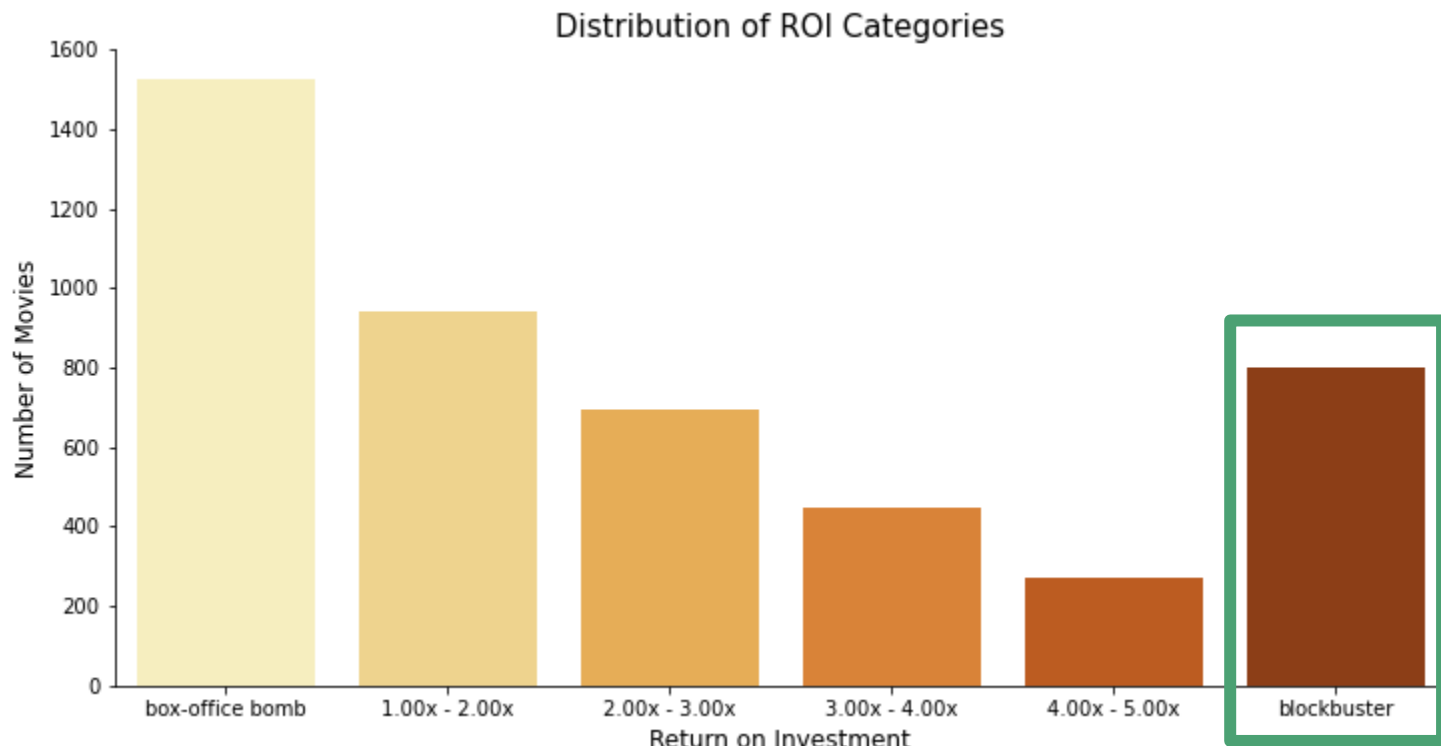- Is domestic/worldwide gross of 10 million USD better than 100 million USD?

→ „Return on Investment" (ROI)
  - Worldwide gross/Production Budget
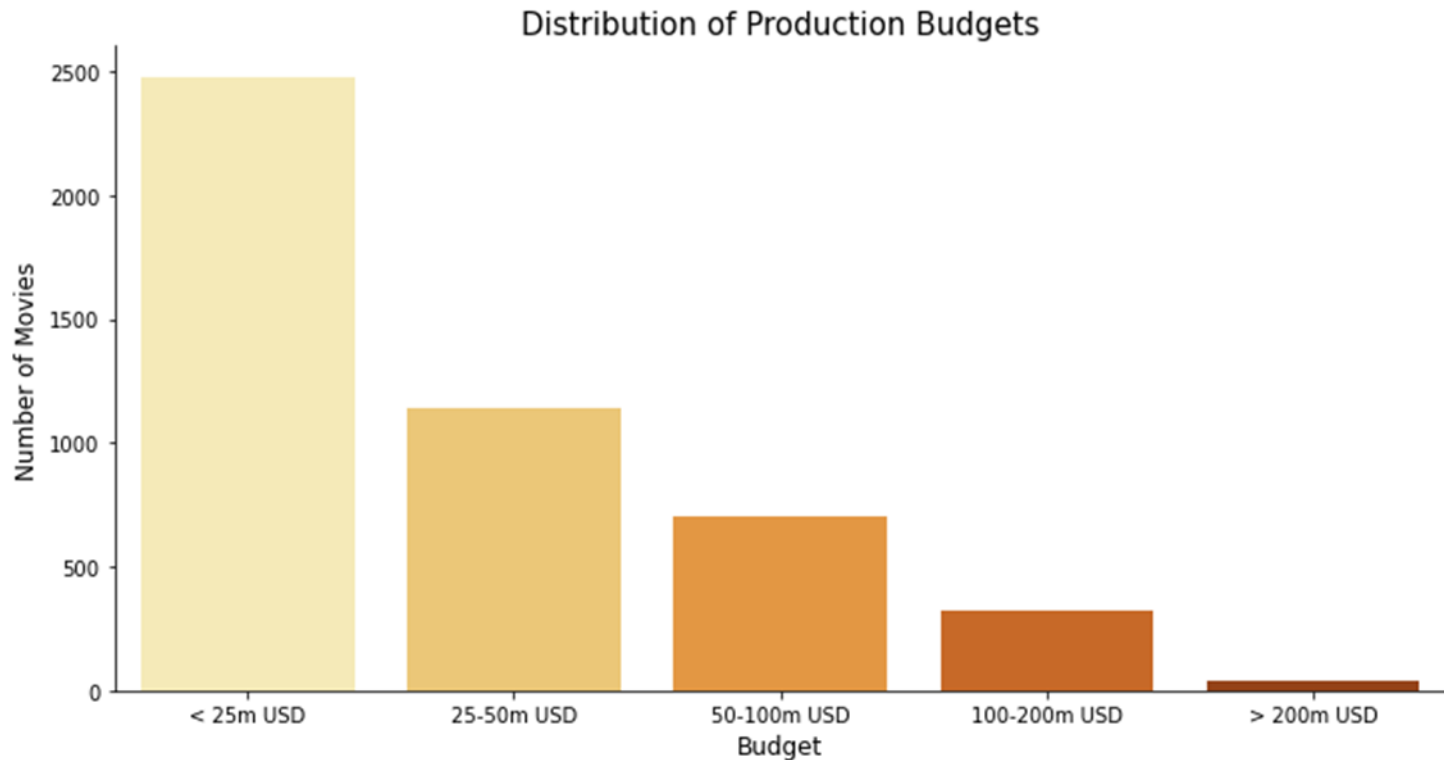 - i.e. „Bang-for-buck", „how many times have I earned my invest?"

Alternative: Profit Margin
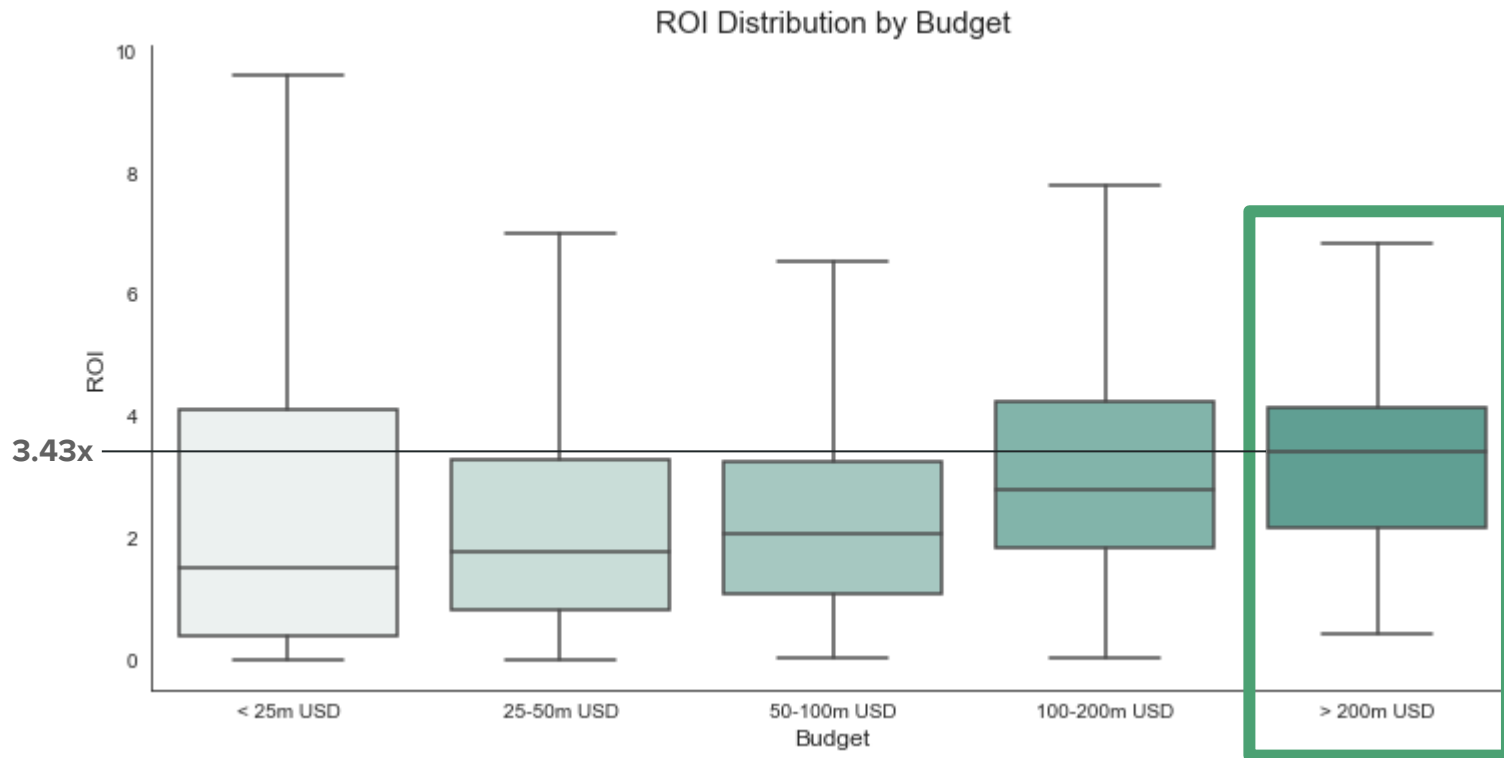(Worldwide Gross – Production Budget) / Worldwide Gross

Return on Investment by Number of Movies

Production Budgets by Number of Movies

Distribution of Production Budgets

Profitability by Production Budget

ROI Distribution by 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

Distribution of Genres

ROI Distribution by by genre
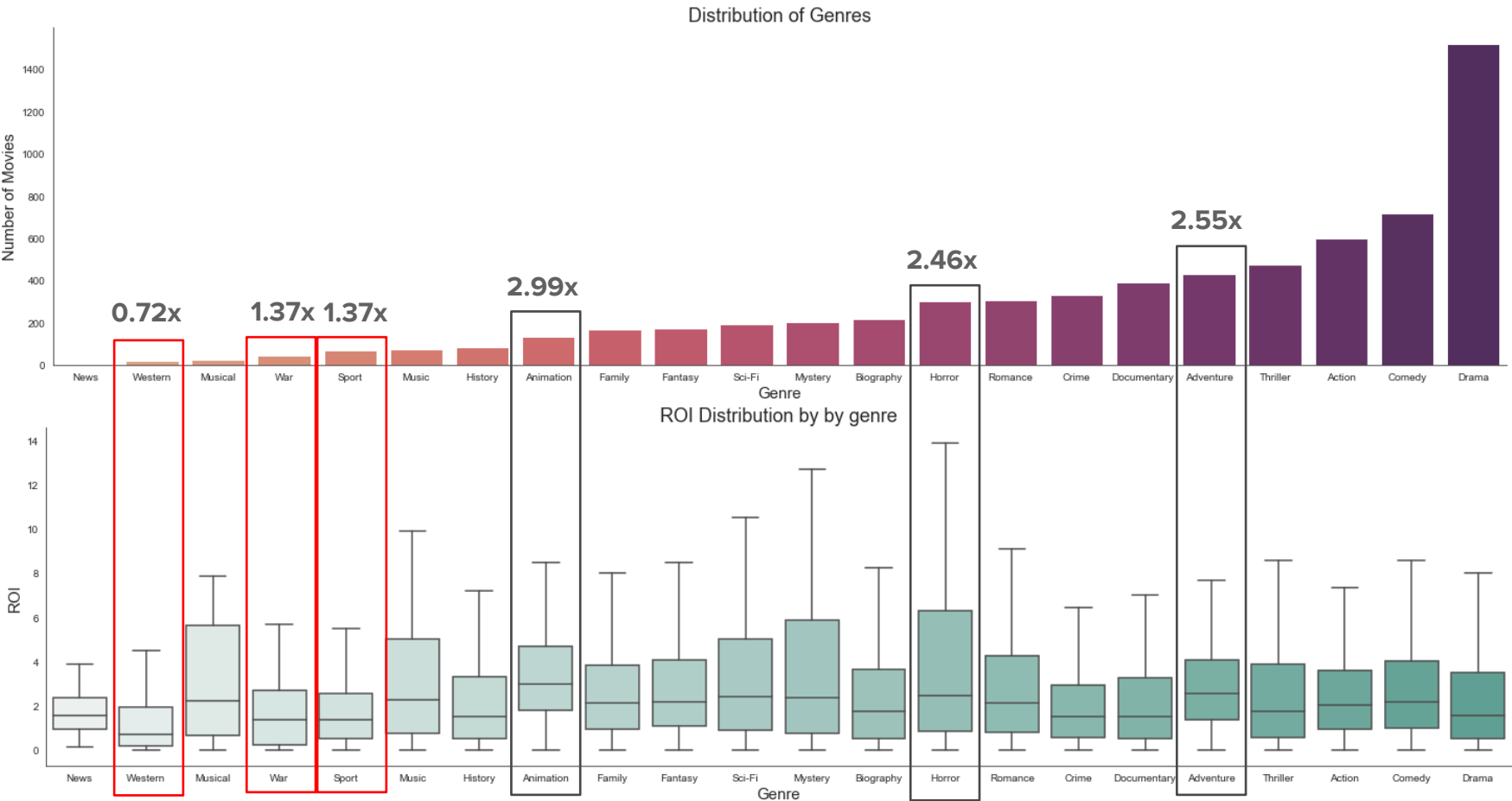
3. Runtime

- 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

Distribution of Runtime

ROI Distribution by Runtime
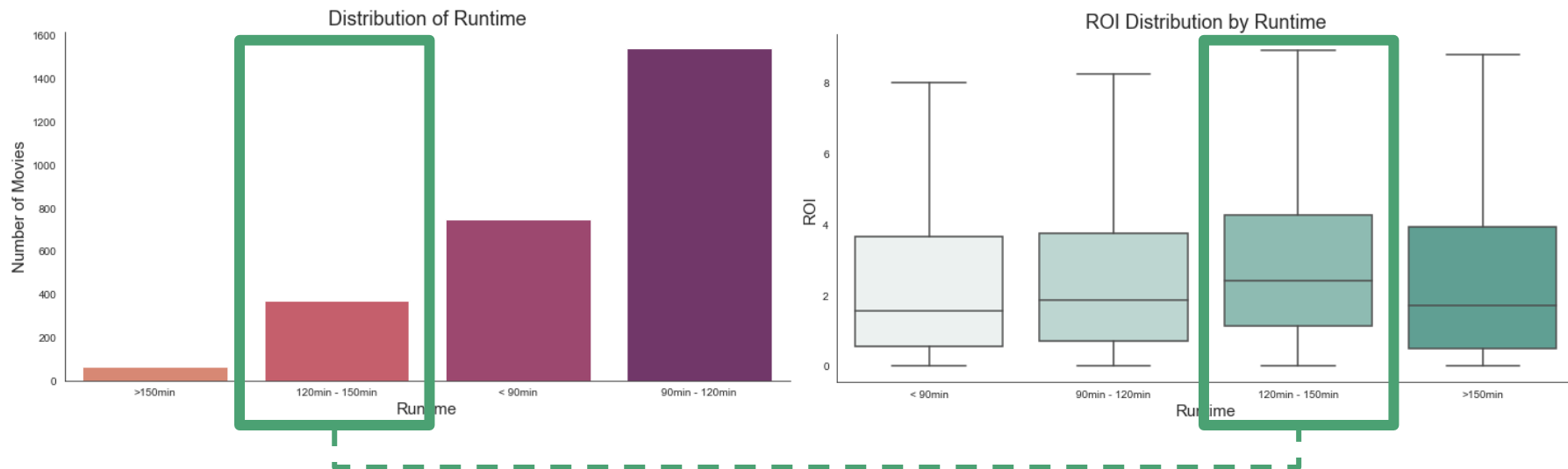
4. Release

- 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
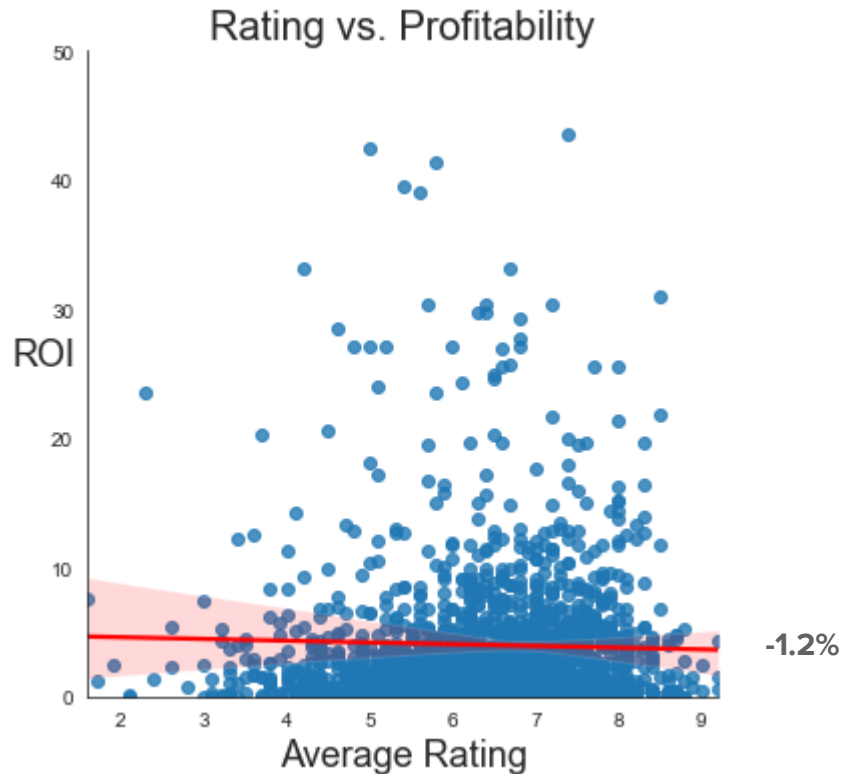
Profitability by Weekday and Month
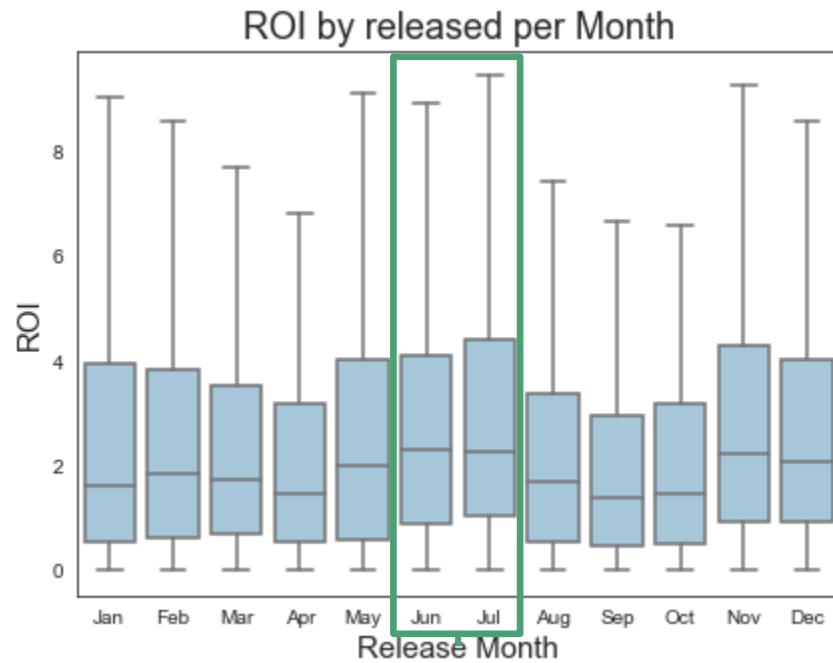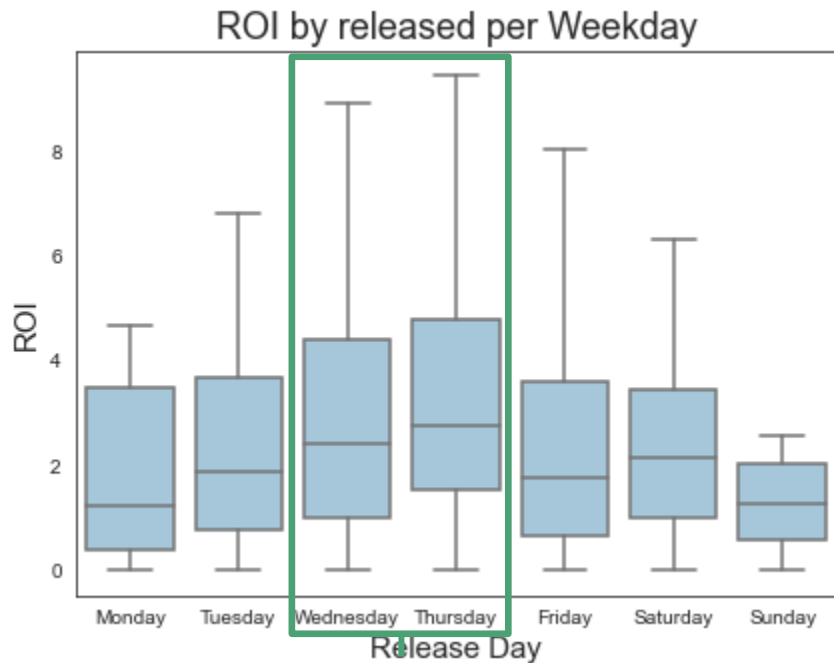
5. Rating

- Does a good rating imply profitability?

- Is rating of moviegoers too subjective?

- Should Microsoft „listen" to their viewers, by favoring types of movies with start rating?

Correlation of Rating and Profitability

Rating vs. Profitability

-1.2%

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

Thank You!

Email: kontakt@oliverzimmer.eu
GitHub: @senseize
LinkedIn: linkedin.com/in/username/oliver-zimmer-cfa-8824881ab/