

Analysis of Movie Data

By Genre, ROI, and Review Data

Initial Considerations

Specific challenges in the Movie Investment

- Which features contribute the most to a high ROI movie
 - How do we define a high ROI movie
- Given the volume of this dataset, how do we choose which features to investigate
- Is there enough correlation within the data to make statistically meaningful conclusions

Initial Considerations

Assumptions & Methodology

- Calculate ROI using budget and gross or revenue
 - Create ROI groups (flop, low, medium, high) defined as
 - Flop = $(-1, 0]$
 - Low = $(0, 0.4)$
 - Medium = $[0.4, 0.7)$
 - High = $[0.8, 1)$

Methodology

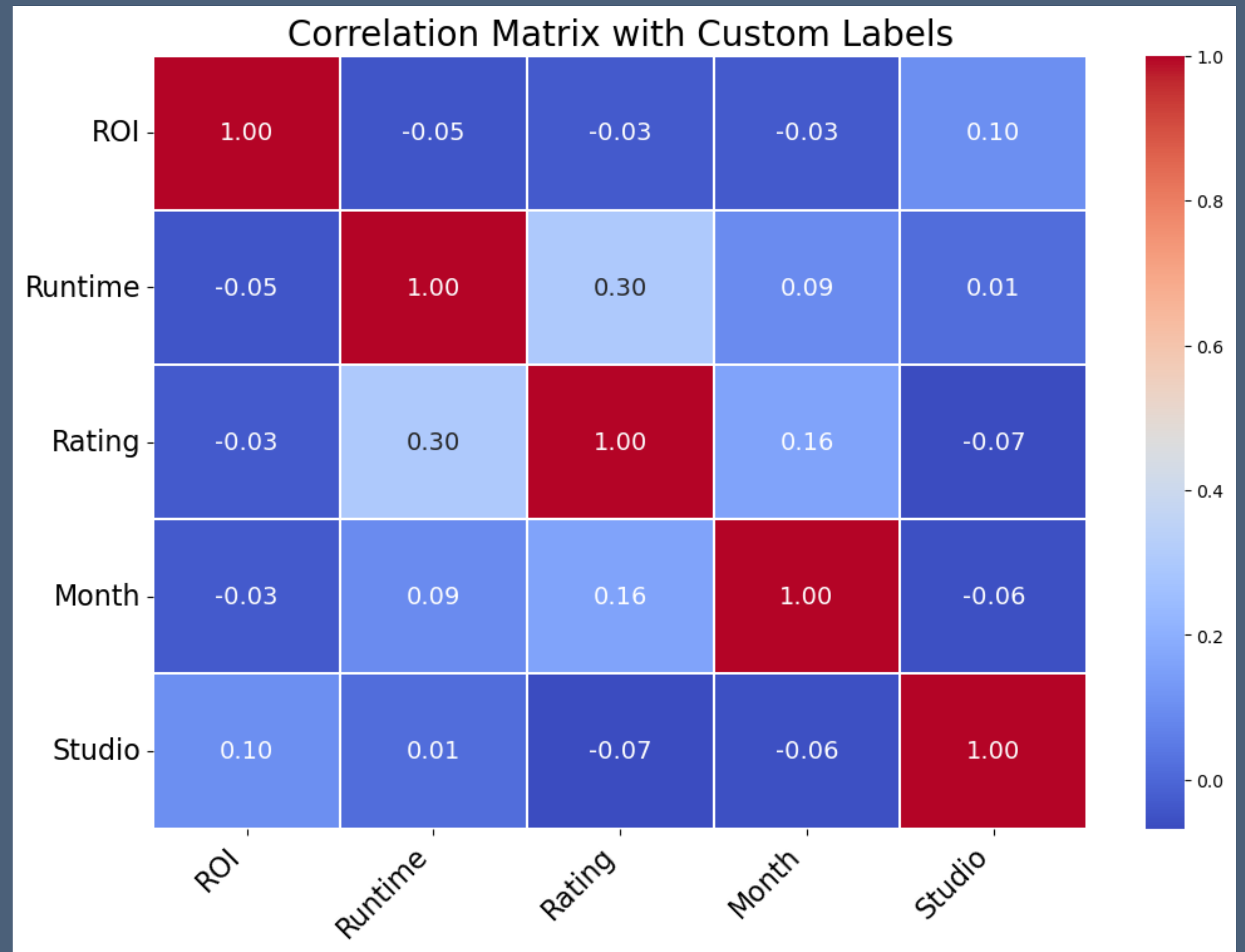
The Approach

- Assemble disparate elements into a master data frame
- Augment data with data provider APIs
- Trim extraneous elements
- Visualize data to detect trends
- Establish correlation matrix to detect trends

Solving the Problem

Analyzing Key Attributes

- ROI
- Runtime
- Studio
- Genre
- Movie Rating
- Release Month



Variable Analysis

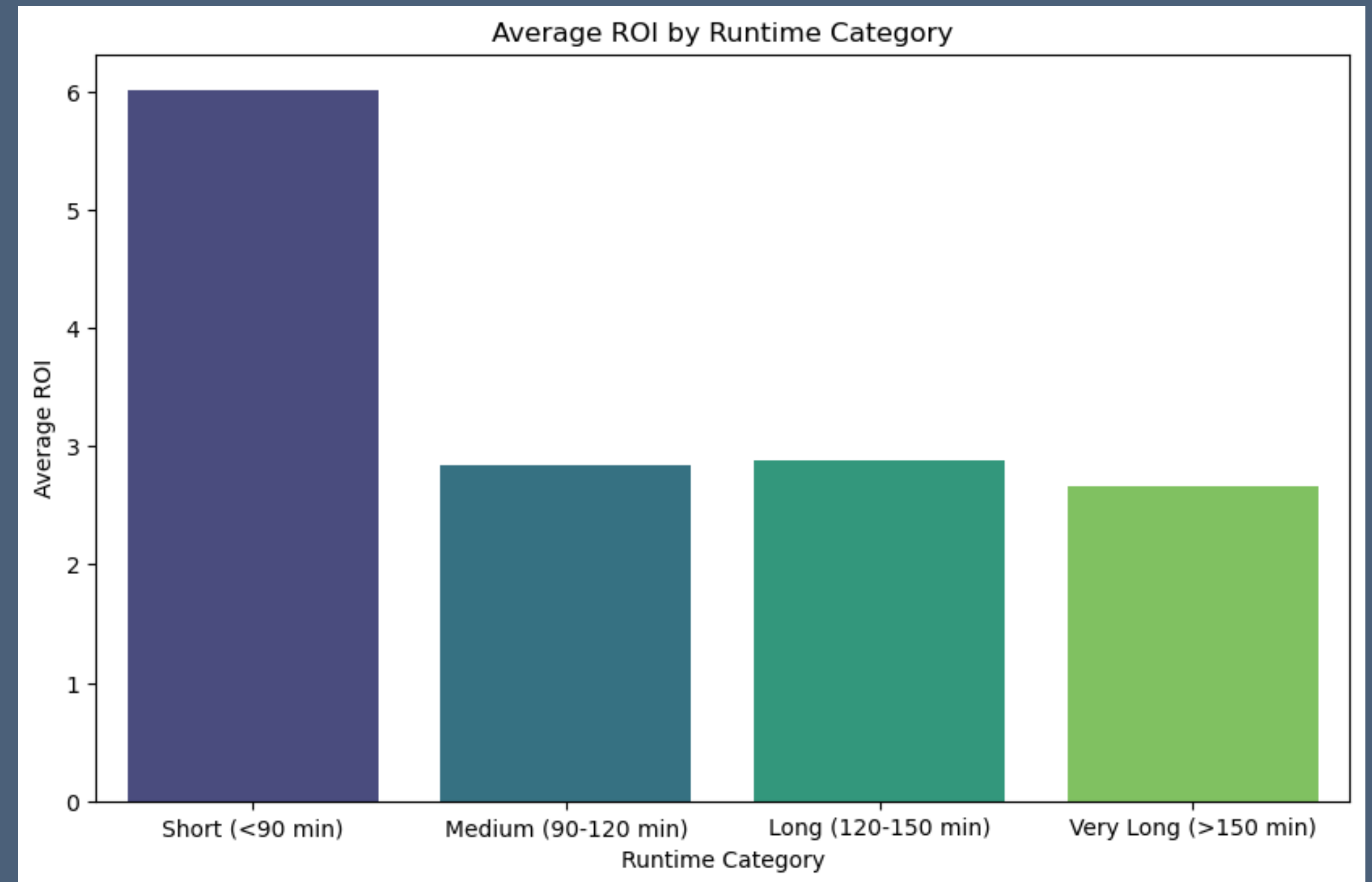
ROI

- Normalized metric (across currencies) that denotes the financial success of a film
- Easily compared to other candidates
- Not overly contrived, or feature engineered
- Basis for organizing films into ROI Groups
 - Establishes guides for understanding most financially successful films

Variable Analysis

Runtime

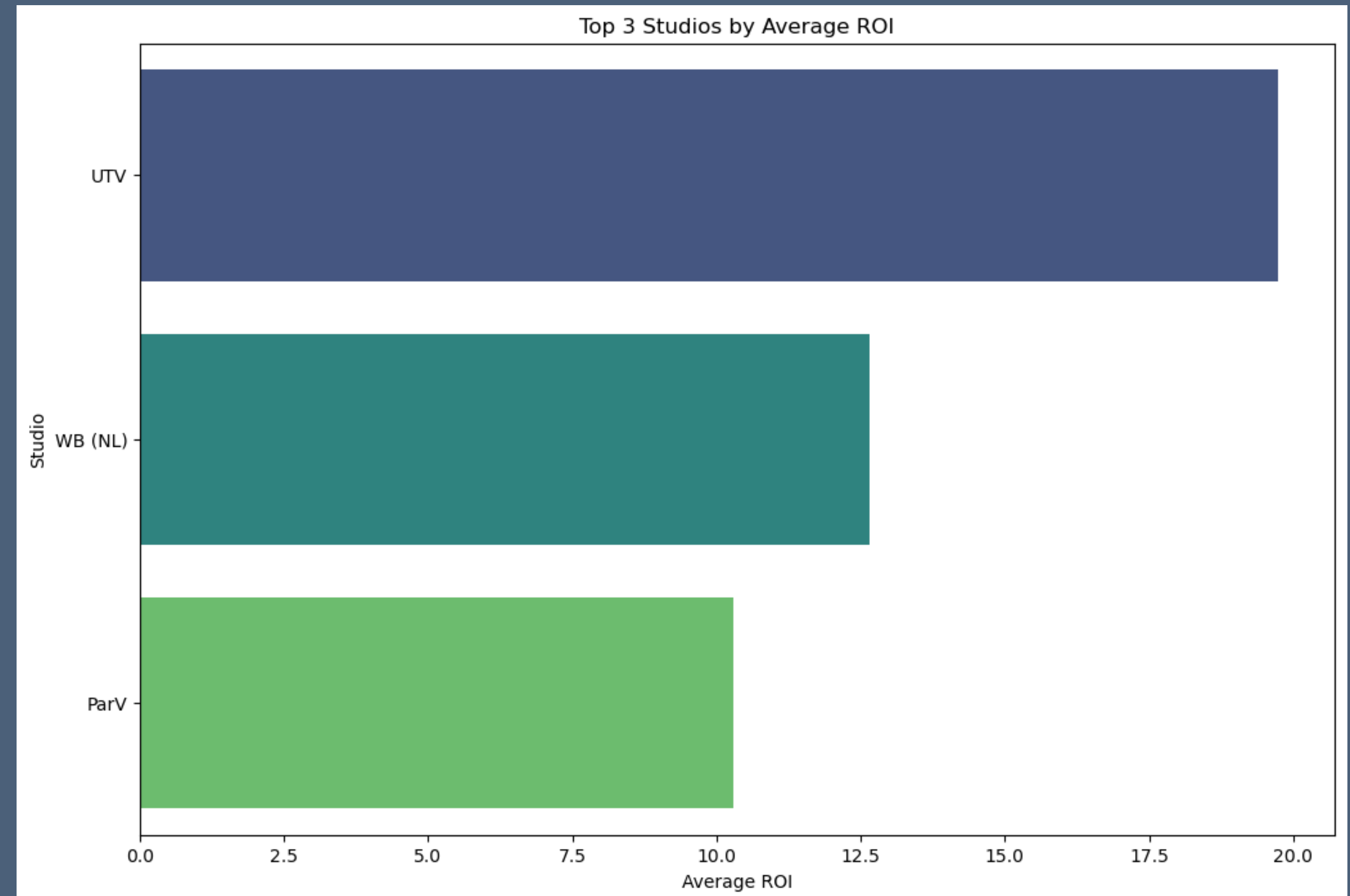
- Movies are expensive so longer movies incur greater cost
- Runtime is positively correlated to budget & revenue
- Runtime is correlated with high rated ROI Groups



Variable Analysis

Studio

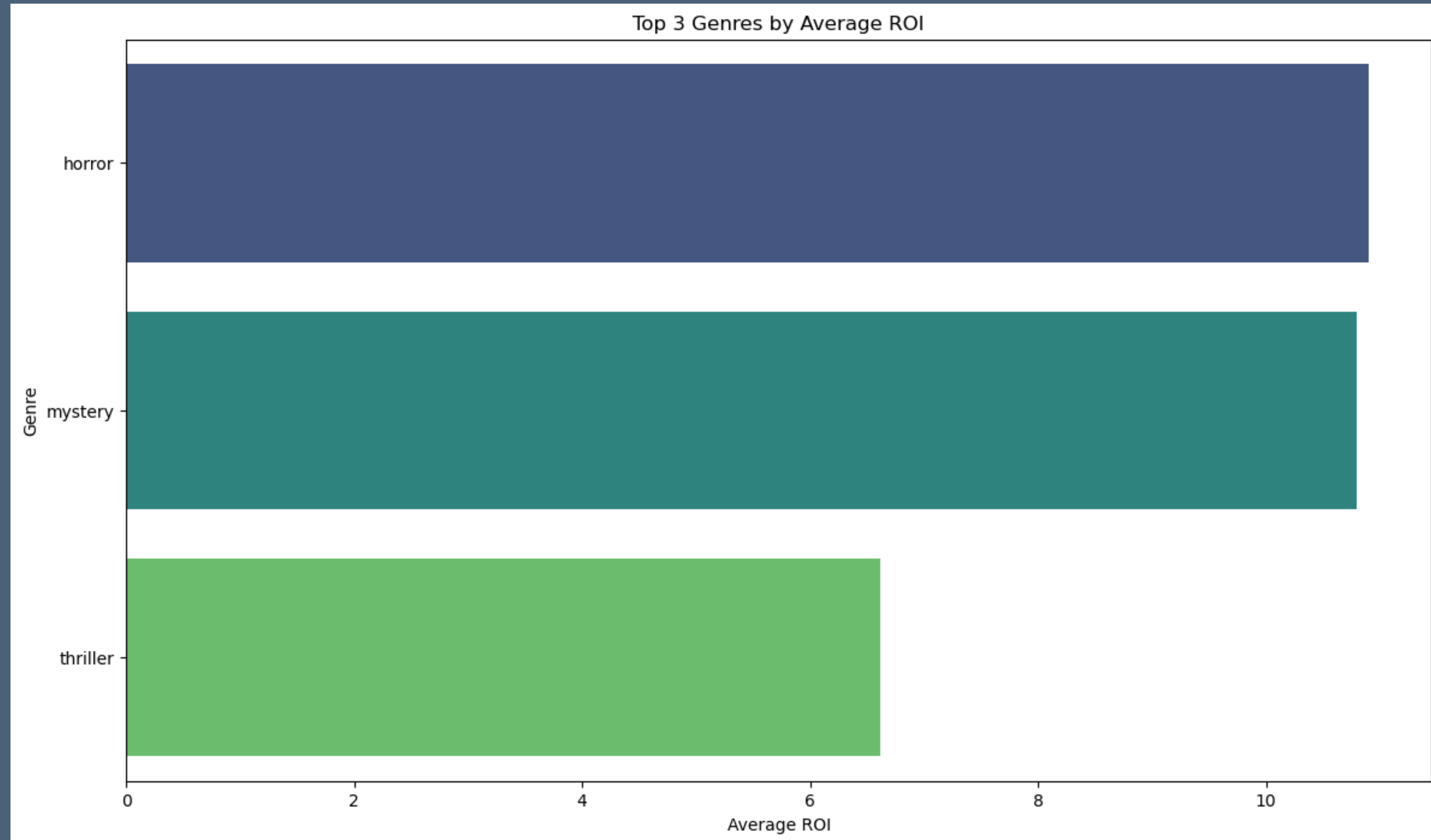
- Studios have a positive correlation with ROI
- Analysis is simple



Variable Analysis

Genre

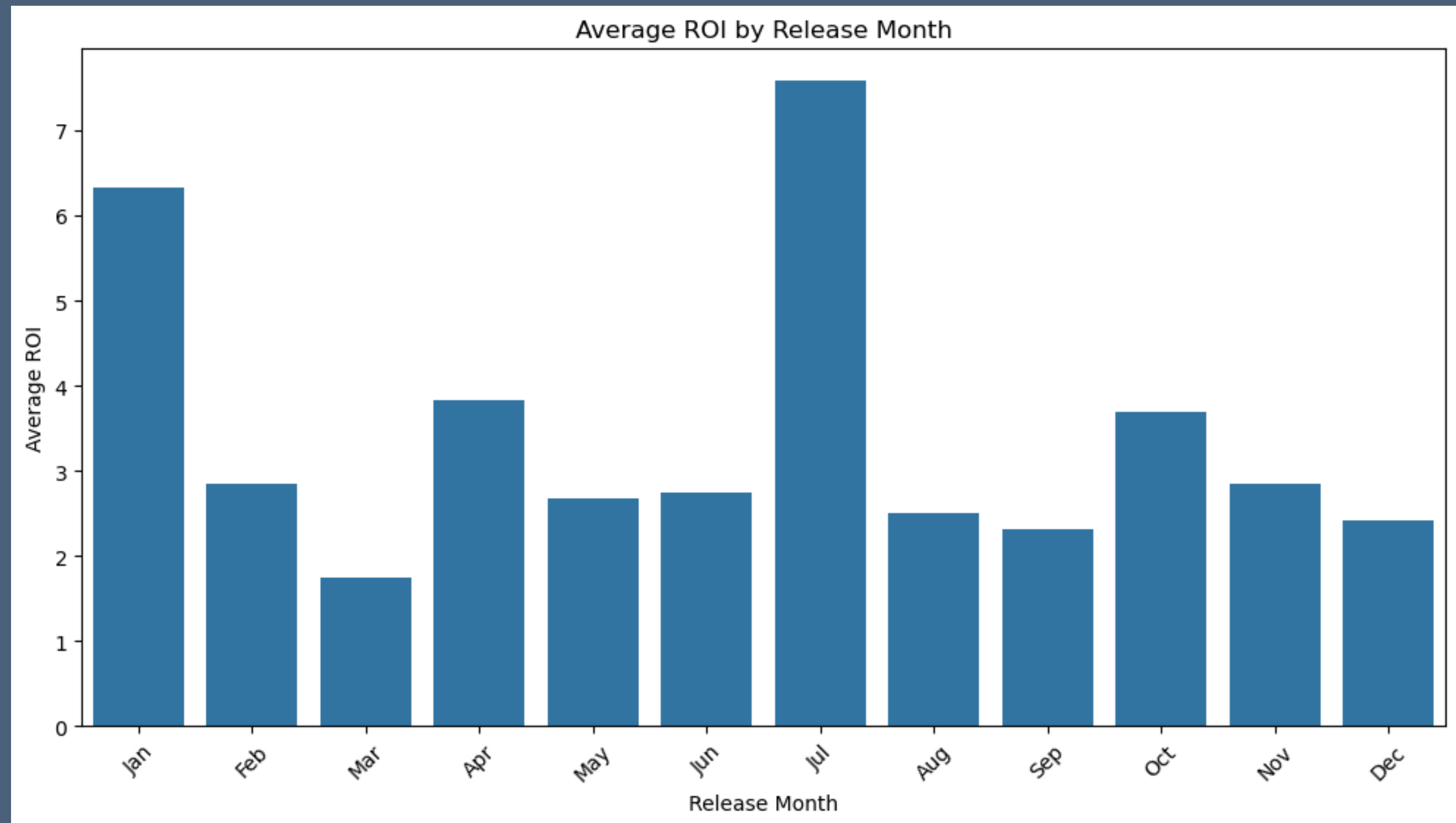
- Genres are correlated to movie ratings
- Horror and Thriller genres tend to produce the highest ROI movies
- These genres tend to be rated lower (meaning higher ROI)



Variable Analysis

Release Month

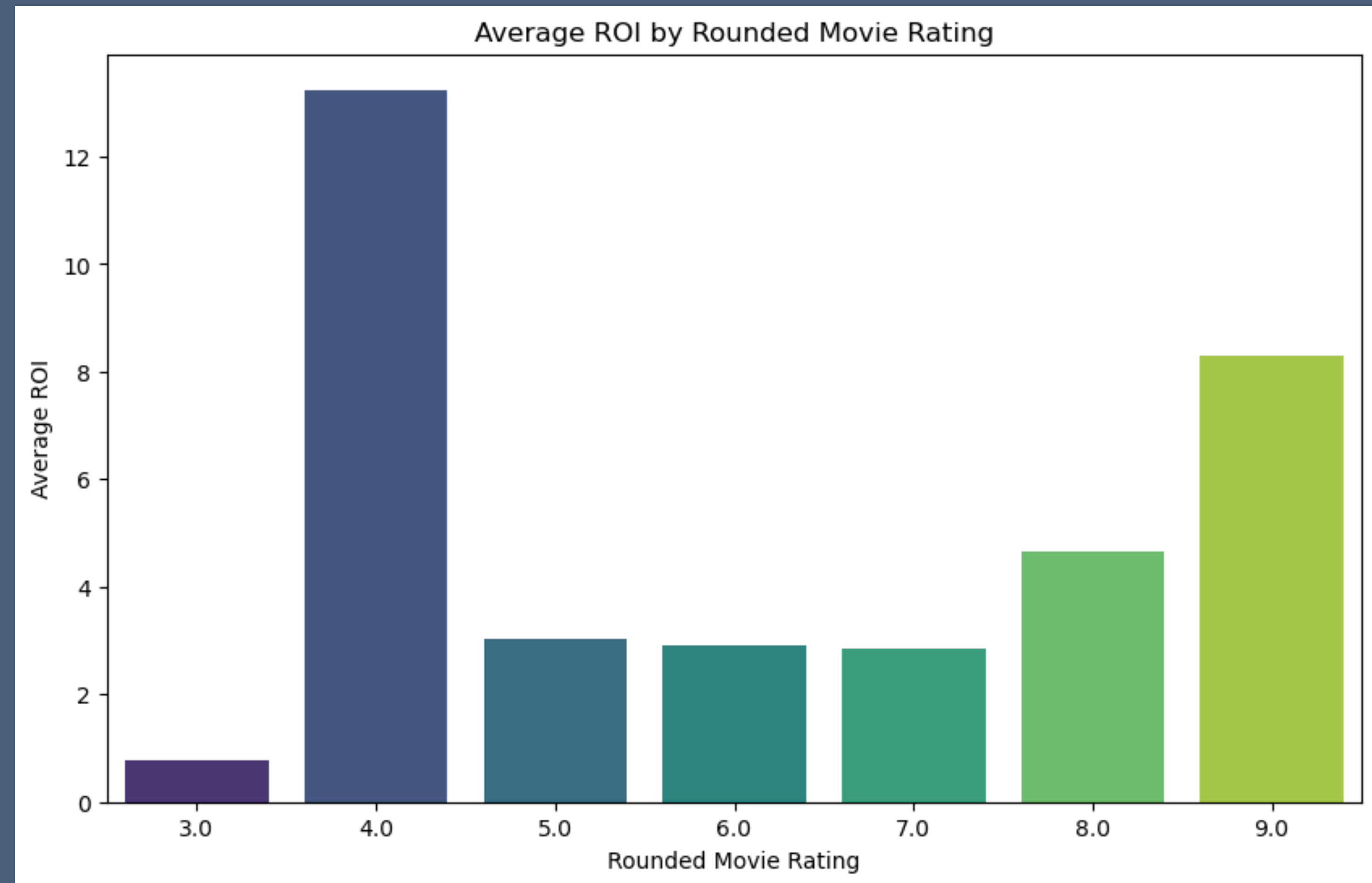
- Analysis is straight forward
- No statistical correlation with ROI
- Data trends imply certain months are preferable however



Variable Analysis

Movie Rating

- Low rated movies (around 4/10) have the highest ROIs
- Very high rated movies (9/10 and greater) have the second highest rated ROI
 - Genres with typically high movie ratings (Western, Music, etc.) have low ROI
- July releases tend to be lower rated (but higher ROI)
- November releases tend to be higher rated



Recommendations & Conclusion

- Invest in Movies that are:
 - Short - 90 minute runtime or less
 - Low rated movies (particularly those released early to mid year)
 - Movies in the Horror/Thriller genre
 - Produced by a top studio (UTV, WB or ParV)
 - Released in July