



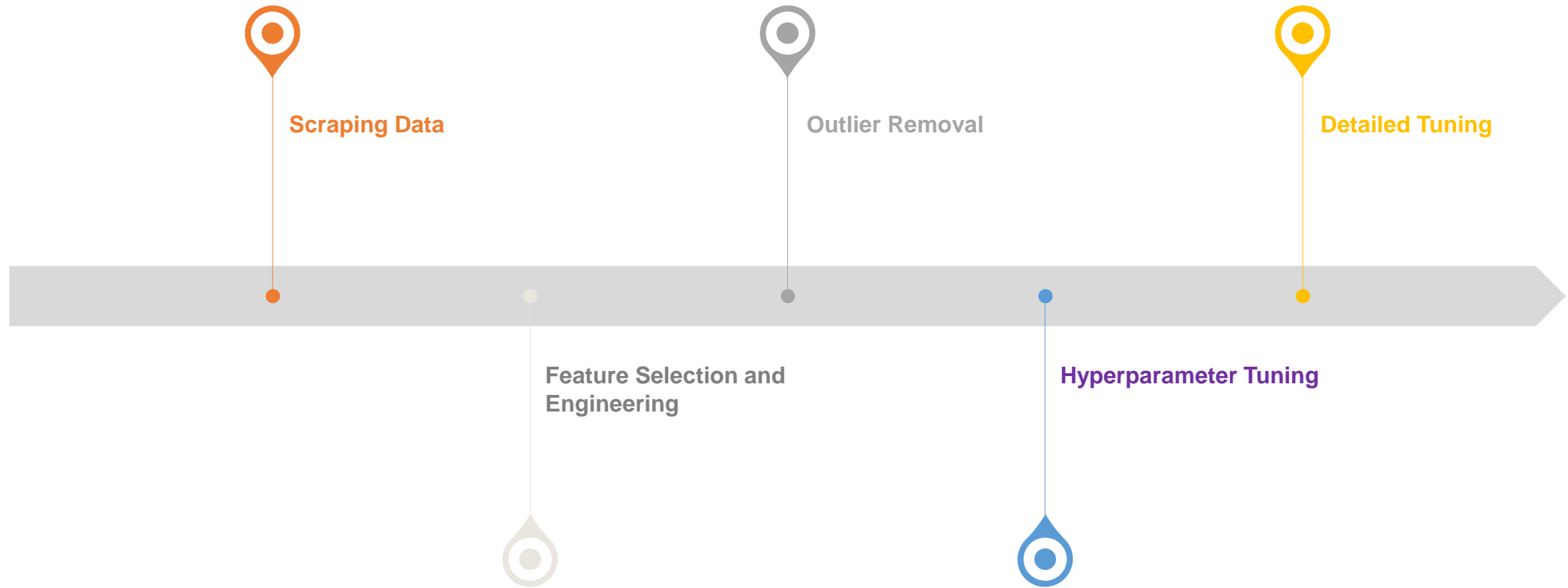
# MOVIE INDUSTRY

Wei Kong

# EXECUTIVE SUMMARY

- Decisions to be impacted
- Business value
- Personal Interest

# MORPHISM





# THE NUMBERS

a website operated by Nash Information Services, LLC.,  
6382 Movies

## SCRAPING DATA

|                                |                                  |
|--------------------------------|----------------------------------|
| moviename                      | Languages                        |
| Budget                         | Leading Cast                     |
| link                           | Supporting Cast                  |
| Domestic Box Office            | Production and Technical Credits |
| International Box Office       | Director                         |
| Worldwide Box Office           | Est. Domestic DVD Sales          |
| Domestic Release               | Est. Domestic Blu-ray Sales      |
| Dom Year                       | Total Est. Domestic Video Sales  |
| International Release          | Lead Ensemble Members            |
| Int Year                       | Cameos                           |
| Rating                         | Uncategorized Crew               |
| Runtime                        | Narrator(s)                      |
| Franchise                      | Extras                           |
| Keywords                       | Documentary Subject(s)           |
| Genre                          | Interviewee(s)                   |
| Production Method              |                                  |
| Creative Type                  |                                  |
| Production/Financing Companies |                                  |
| Production Country             |                                  |

# FEATURE SELECTION AND ENGINEERING

$$\text{ROI} = (\text{Revenue} - \text{Budget}) / \text{Budget} * 100\%$$



Revenue

Release Year

Release Month

Is Franchise?

Leading Casts

~~Languages~~

~~Leading Cast~~

~~Supporting Cast~~

~~Production and Technical Credits~~

~~Director~~

~~Est. Domestic DVD Sales~~

~~Est. Domestic Blu-ray Sales~~

~~Total Est. Domestic Video Sales~~

~~Lead Ensemble Members~~

~~Cameos~~

~~Uncategorized Crew~~

~~Narrator(s)~~

~~Extras~~

~~Documentary Subject(s)~~

~~Interviewee(s)~~

~~moviename~~

~~Budget~~

~~link~~

~~Domestic Box Office~~

~~International Box Office~~

~~Worldwide Box Office~~

~~Domestic Release~~

~~Dom Year~~

~~International Release~~

~~Int Year~~

~~Rating~~

~~Runtime~~

~~Franchise~~

~~Keywords~~

~~Genre~~

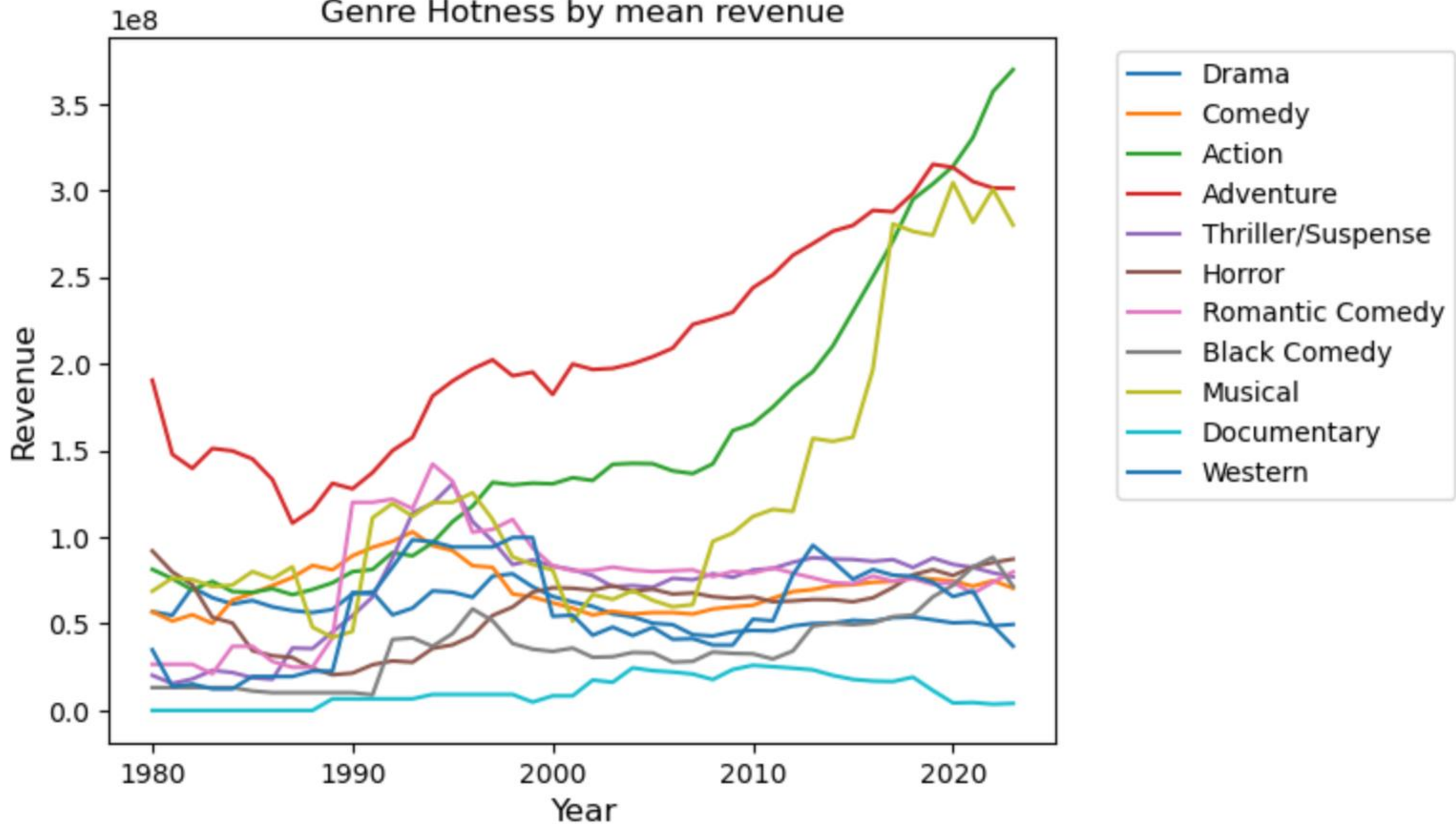
~~Production Method~~

~~Creative Type~~

~~Production/Financing Companies~~

~~Production Country~~

Genre Hotness by mean revenue



# FEATURE SELECTION AND ENGINEERING

|   | moviename                   | Budget    | Genre  | Leading Cast                                       | Director                             | Revenue        | ROI    | Year     | Month |
|---|-----------------------------|-----------|--------|--|--------------------------------------|----------------|--------|----------|-------|
| 0 | The Super Mario Bros. Movie | 100000000 | Action | ['Chris Pratt', 'Anya Taylor-Joy', 'Charlie Da...] | ['Aaron Horvath', 'Michael Jelenic'] | 509,221,843.00 | 409.22 | 2,023.00 | April |

Time window of  $y = 3$  years

Production Period

Release



How much each Actor made in this time window? → Top  $n_1$  actors? → Has star actor?

Directors? → Top  $n_2$  directors? → Has star director?

How much does actors in this movie made in the past time window? → Sum of actors' past revenue

Directors? → Sum of directors' past revenue

Movie of this Genre? → Sum of genre performance

# OUTLIER DETECTION

Fact Check based on ROI  
-95 to 50000

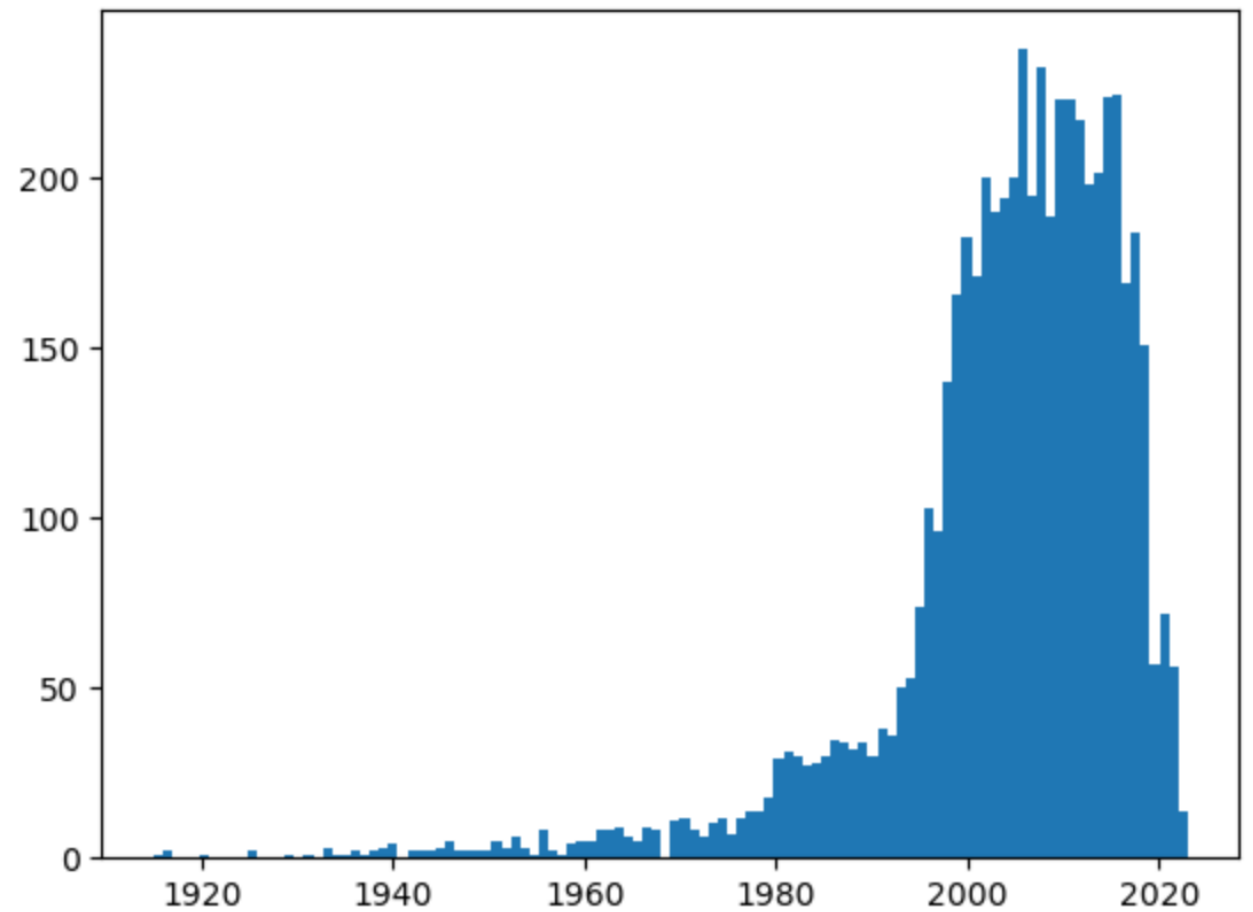
| moviename                           | Budget   | Revenue  | ROI     |
|-------------------------------------|----------|----------|---------|
| Kirikou Et Les Hommes Et Les Femmes | 7500000  | 17.00    | -100.00 |
| Capricorn One                       | 5000000  | 401.00   | -99.99  |
| Pandaemonium                        | 15000000 | 1,438.00 | -99.99  |
| Perrier's Bounty                    | 6600000  | 828.00   | -99.99  |

Skin Trade    9000000    1,242.00    -99.99

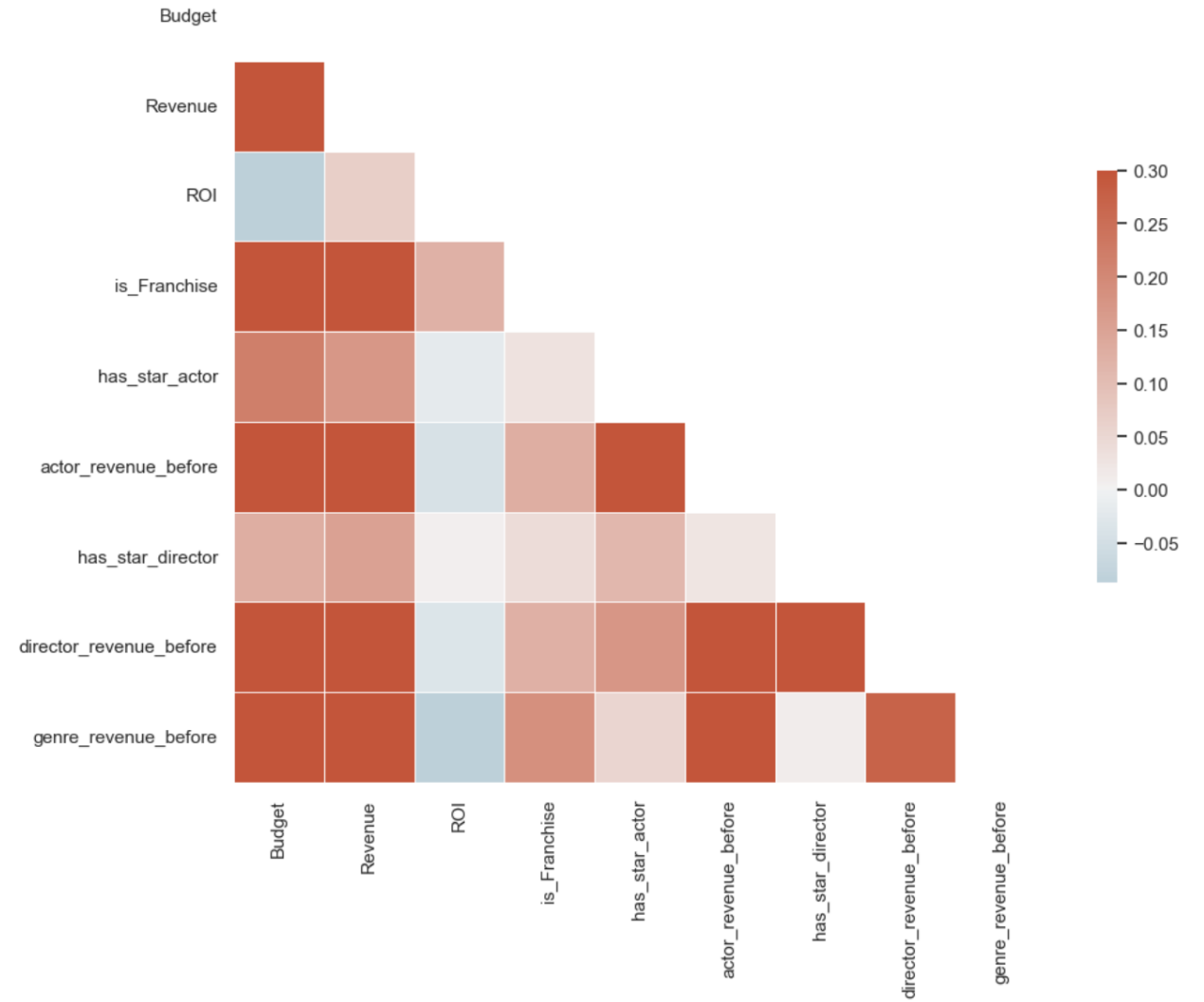
| moviename               | Budget | Revenue        | ROI        |
|-------------------------|--------|----------------|------------|
| The Blair Witch Project | 600000 | 248,300,000.00 | 41,283.33  |
| The Gallows             | 100000 | 41,656,474.00  | 41,556.47  |
| Paranormal Activity     | 450000 | 194,183,034.00 | 43,051.79  |
| Mad Max                 | 200000 | 99,750,000.00  | 49,775.00  |
| Deep Throat             | 25000  | 45,000,000.00  | 179,900.00 |



## NUMBER OF MOVIE BY YEAR



# CORRELATION PLOT





# MODEL SELECTION

Random Forest and XGBoost

# HYPER- PARAMETER TUNING

## Grid Search

Time window length for  
actor, directors, and genre  
[3, 5, 10]

Number of star actor and  
director to be considered in  
the time window [3, 5, 10]

Criteria: RMSE

## Grid Search Cross Validation

5 Folds

Max Depth

Max Features

Criteria: RMSE

## Best Hyperparameters

# HYPER- PARAMETER TUNING

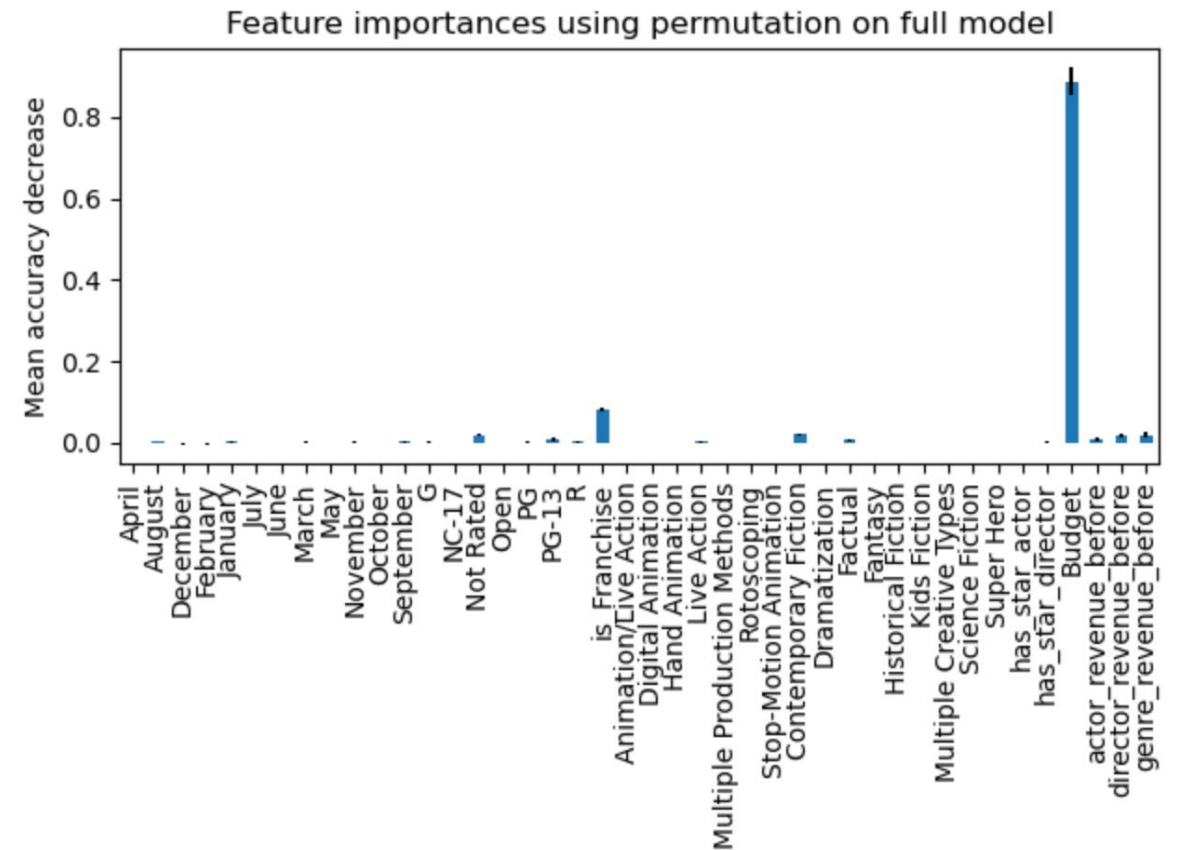
|    | timewindow_actor | timewindow_director | timewindow_genre | star_actor_num | star_director_num | params           | score |
|----|------------------|---------------------|------------------|----------------|-------------------|------------------|-------|
| 54 | 3                | 10                  | 3                | 3              | 3                 | {'max_depth': 6} | -1.38 |

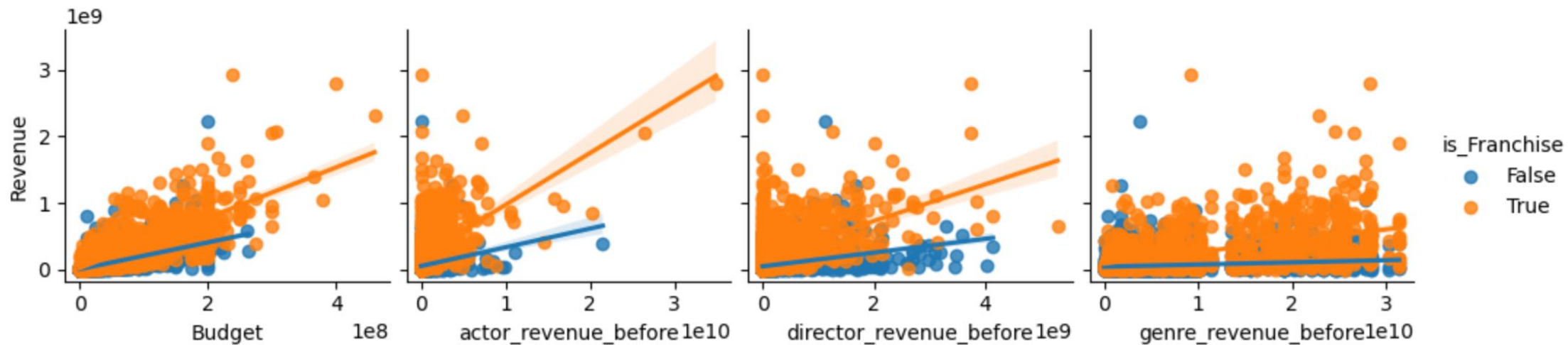
Do Grid Search Cross Validation Again  
Tune more details of XGBoost Model

# RESULTS

r2 score: 0.49796386164252404  
MSE: 1.6210515813363064

r2 score: 0.5738890076632135  
MSE: 1.3877082380000316





### OLS Model

r2 score: 0.1260607791334808  
MSE: 2.023095142953555

r2 score: 0.37567420925092965  
MSE: 0.844808891064326

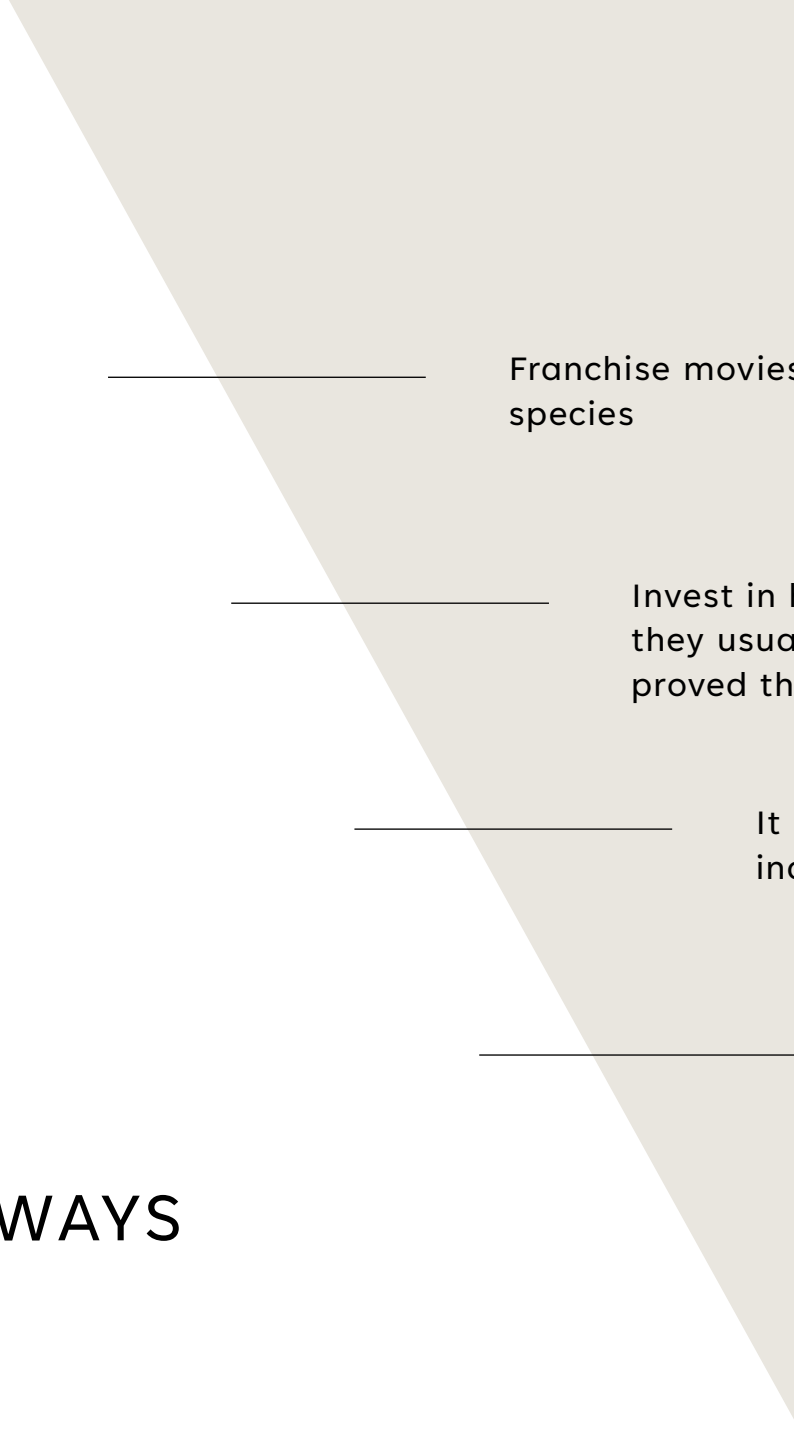
OLS Model for Franchise only

r2 score: -0.28570693372472533  
MSE: 2.367045898228619

OLS Model for non-Franchise only

r2 score: 0.35777696155216854  
MSE: 1.6029046969447214

XGBoost model for non-Franchise



---

Franchise movies and non-Franchise movies are two different species

---

Invest in higher budget movies are more likely to succeed, as they usually have more successful directors and actors that has proved their ability before

---

It is much harder to predict the success of non-franchise individual movies and they exhibits more randomness

---

TAKEAWAYS