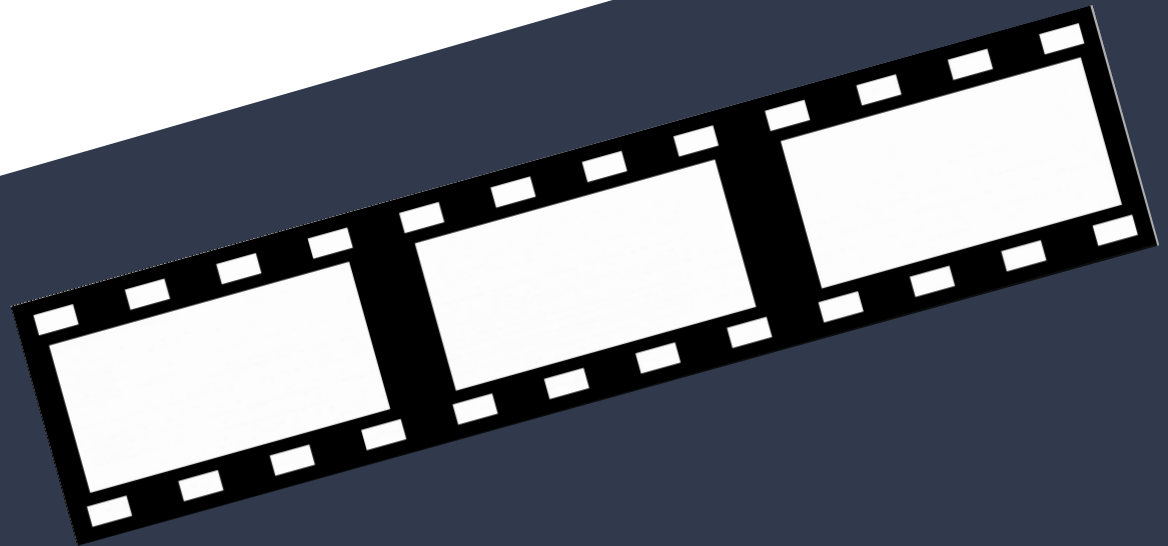
Analysis of Movie Ratings

Predictions and Trends

Adam Ariel



Ratings

- IMDB
- Rotten Tomatoes
  - Tomatometer
    - Top
    - All
  - Average Rating
    - Top
    - All
  - Audience Score

Data

- IMDB Public Data
  - Significant points
- Scrapy:
  - http://www.imdb.com/ ~4000 entries
  - https://www.rottentomatoes.com/ ~2500 entries
- Cleaning
  - Formatting
  - Discrepancies
  - Scaling
- ~1500 Data Points

Features:

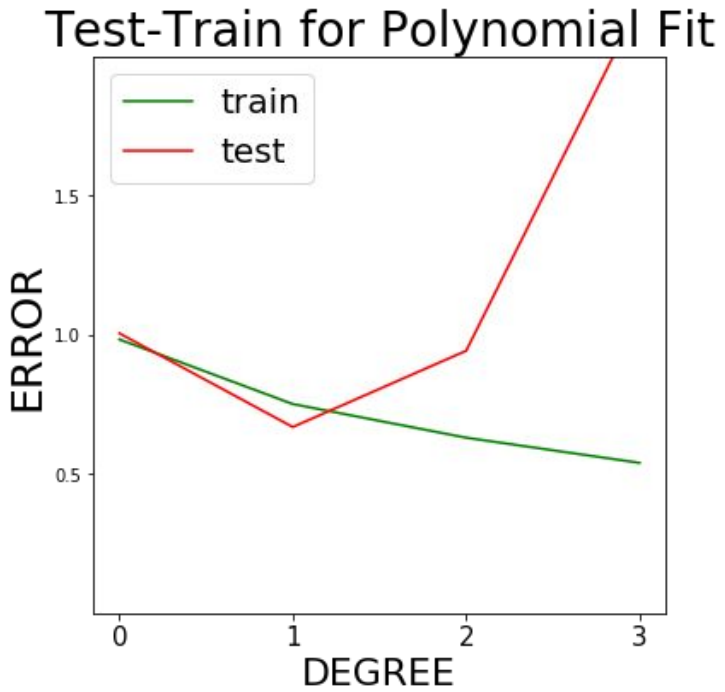
- Runtime
- Domestic Gross
- Initial Weekend Gross
- Budget
- Genre
- MPAA Rating

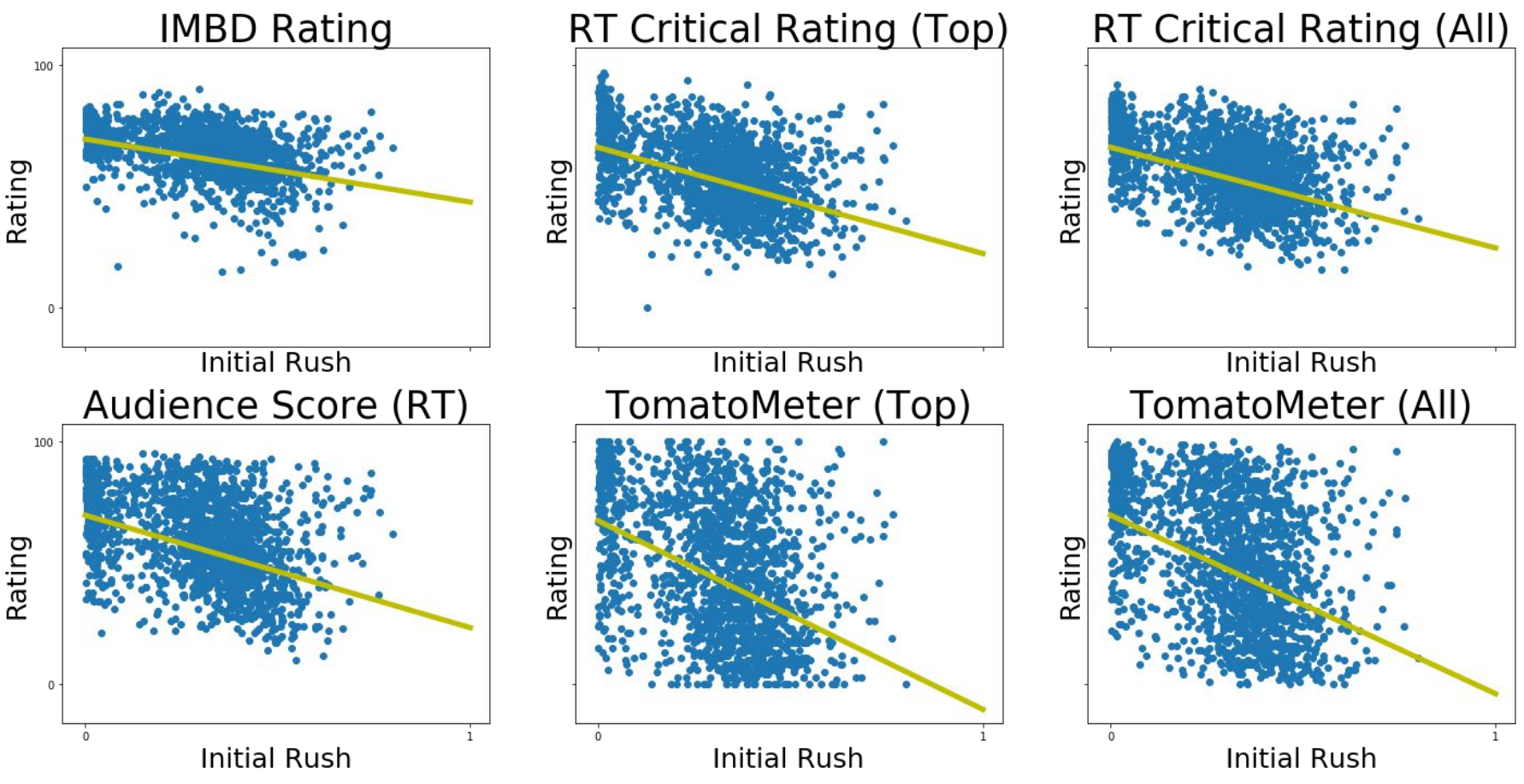
Engineered:

- Runtime Deviation from Mean
- Domestic Gross
- Initial Rush %
- (Budget)
- MPAA Rating (indicators)

Initial Analysis

- Correlation / Pair Plot
  - High variance
- Test-Train for polynomial degree
  - Linear fit (see figure)
  - No expected interactions
- Fit with linear OLS: $R^2 \approx 0.3$
  - Budget ~ Domestic
- Ridge Analysis with cross val
  - No significant improvement in R^2
- Largest Coefficient for Initial Rush

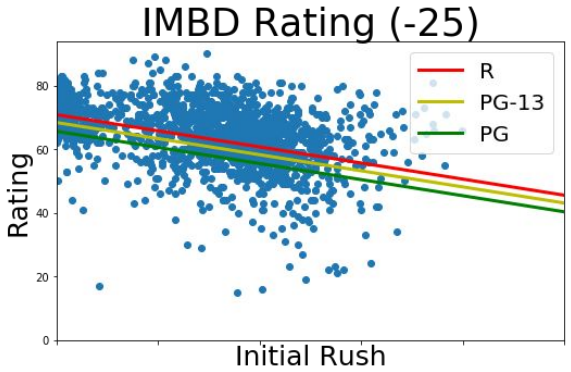


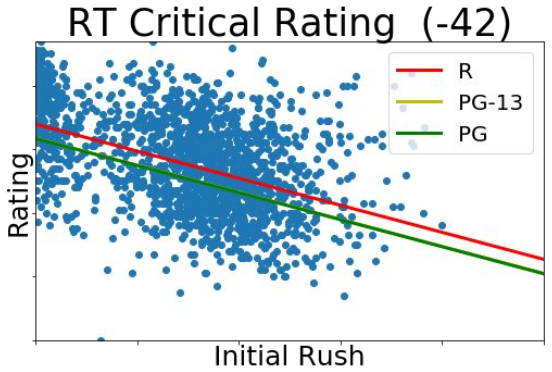


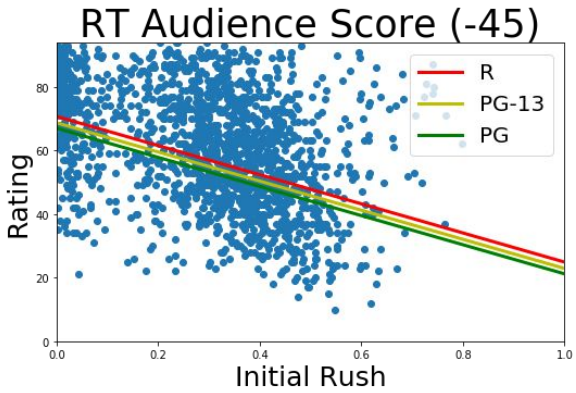
Six Ratings with Initial Rush % Trend

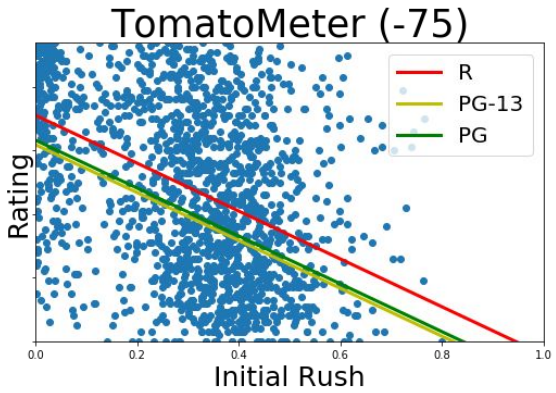
Secondary Analysis

- Top ~ All
- Explore trends among MPAA ratings
  - Too few G and NR
- Increasing Beta



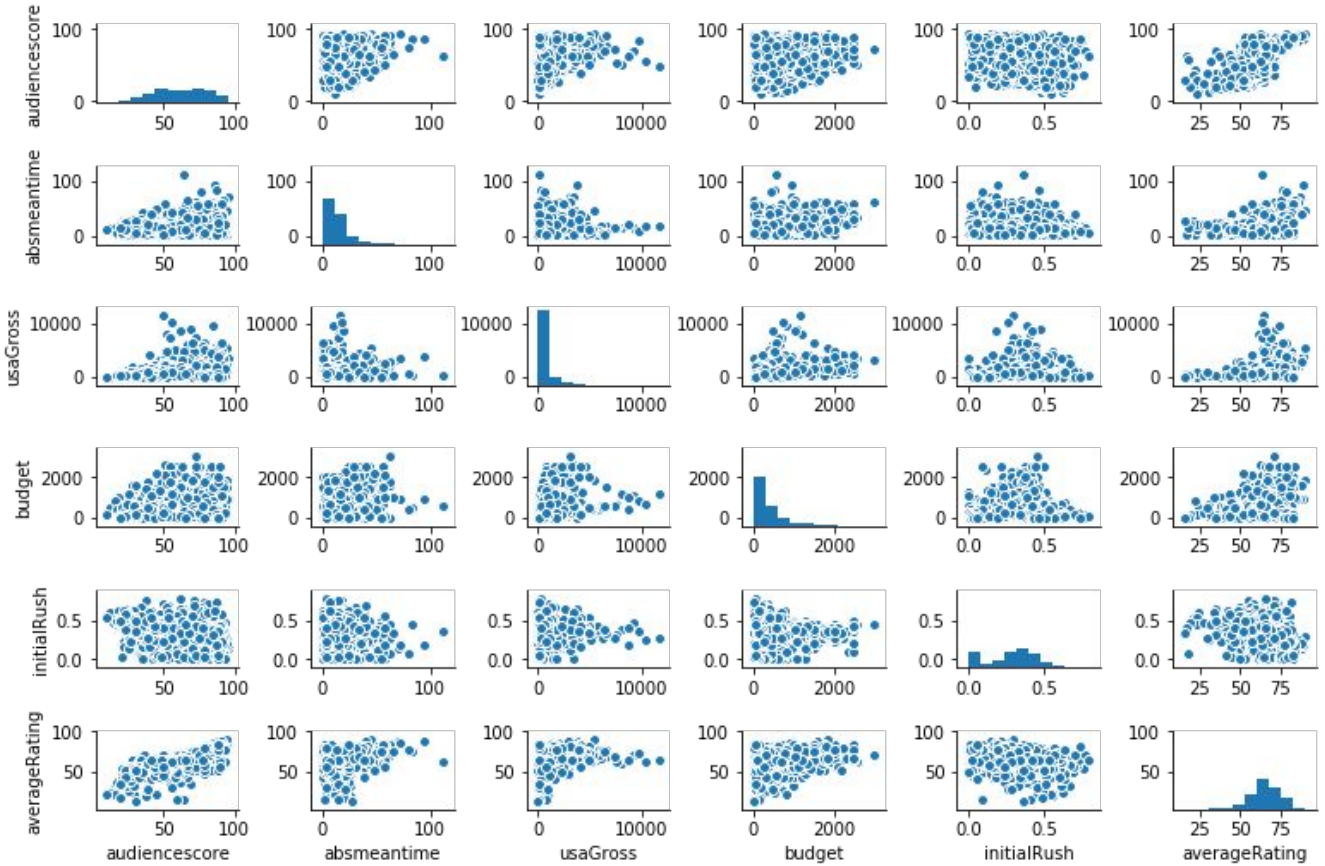




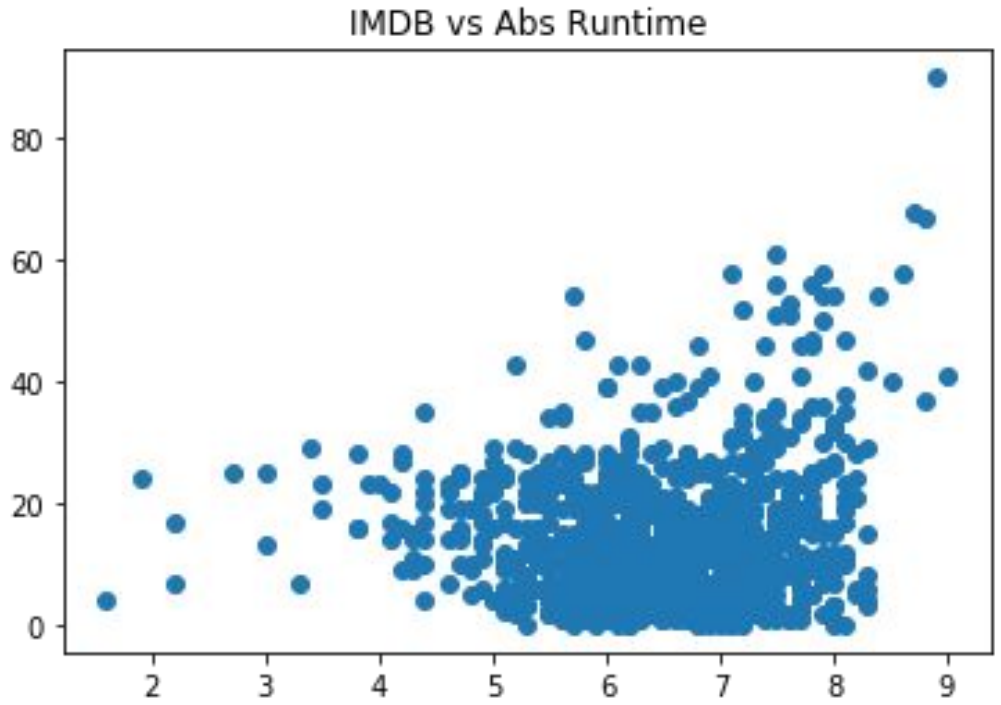


- Poor predictive power due to high variance
  - Small R^2 and small p-values
- Clear overall trends identified
- Some variations between ratings quantified
- Further strategies:
  - More features (obviously)
  - Initial over month
    - Wide/Limited
  - Cause/Effect

Thank you



Appendix #1: Pair Plot



Appendix #2: Runtime Distribution