CITADEL

TEAM 11

Citadel Data Open West Coast Regional

Topic Question

Understanding the **inter-temporal** behavioral differences in movie reviewers, and how such traits may vary across **different types** of movies.

Background & Motivation

- Unlike the past century, competition in **The Second Golden Age of television is met in both the cinema and on the small screen**.
- Customers are now more keen to "**wait till it's on Netflix**," or their preferred streaming platform rather than going to theaters. Factors like these result in a divergence in viewer profiles across time.
- However, **streaming platforms and Hollywood still have two distinct, primary goals.**
  - Streaming platforms want titles that **attract customers to their platform**
  - Hollywood wants titles that **attract customers to the box office**
- As movie ratings are playing an increasingly prominent role in determining people's watch patterns, **how might we better understand the evolution of consumers' perception of a movie following its release?** Do temporal differences in consumer perception of movies vary across genres? Leveraging on these insights, can we predict whether a movie tends to grow on viewers over time, or receives an initial hype before quickly dying out?

Summary of Analysis & Conclusions

- In our research, we explore **time-based trends in viewers' general perception** of movies in the months following a movie's release.
- By profiling users based on their **average rating given to all movies** in their rating "career", as well as considering the **time elapsed between a movie's release and user ratings**, we discovered that the monthly rating of a movie **can be predicted** by these user profiles as well as the time elapsed from release.
  - In particular, we realized that on average, **a movies' monthly ratings tend to decrease slightly over time**
- As our data showed that rating frequency trends differ across genres, we further analyzed the **impact of movie features**, such as genres and profit margins, on the variability of monthly movie ratings over time.
- Our prediction model did **draw relations between movie features and ratings variability**, but was only able to make **weak predictions** in general. We did notice that predictability generally increased as we considered a longer timeframe, suggesting that **movie features have the potential to predict how specific movies' popularity may vary and evolve over the long term.**
  - These results can also be taken into consideration by movie recommender systems to improve user satisfaction and retention.

Dataset Usage & Data Wrangling

movies.csv

> 58098 rows x 3 columns (movieId, title, genres)

> Dataset containing all movies for analysis, cross-referenced with ratings dataset using movieId; genre features were analyzed

movie_industry.csv

> 6820 rows x 15 columns (budget, name, etc…)

> Dataset containing popular movies and industry metadata, used for feature exploration

ratings.csv

> 28M rows x 4 columns (userId, movieId, rating, timestamp)

> Dataset with all ratings info together with timestamp

genome-tags.csv
genome-scores.csv

> tags: 1128 rows x 2 columns (tagId, itemId)

> scores: 15M rows x 3 columns (movieId, tagId, relevance)

> *title* came in the form of "title (year)" e.g. *Avengers, The (2012)*. Year was extracted and added to a new column *release*

> *title* format did not correspond to other datasets e.g. *Avengers, The* would appear as *The Avengers* in movie_industry. This was resolved through string formatting prior to performing an inner join with movie_industry, so as to obtain industry metadata on selected movies. Similarly, foreign movies sometimes had both original language and translated names; such differences were also resolved through string formatting methods.

> *genres* came in the form of "genre1|genre2|genre3…" as a string. One-hot encoding was done for all movies and genres, producing 16 binary columns representing genres.

> For movies with "0" in the *budget* column, we interpolated their budget using mean profit margins and gross revenue data.

> We further subset movies which are released from 1996 onwards, since ratings data begin continuously from Jan 1996 (1995 has 1 month of data), and we are interested in the duration between movie release and rating.

> *timestamps* were converted from seconds since epoch to pandas datetime format for ease of timing comparison and grouping

Feature Engineering

Ratings Dataset

> With the ratings dataset, we engineered the following features: user_rating, months_delta
> user_rating: the average rating given by the user throughout his movie ratings in the dataset (unique to each userId)
> months_delta: the number of months elapsed from the timestamp of the movie’s first rating in the dataset until the timestamp of the rating (unique to each rating)

Movies Datasets

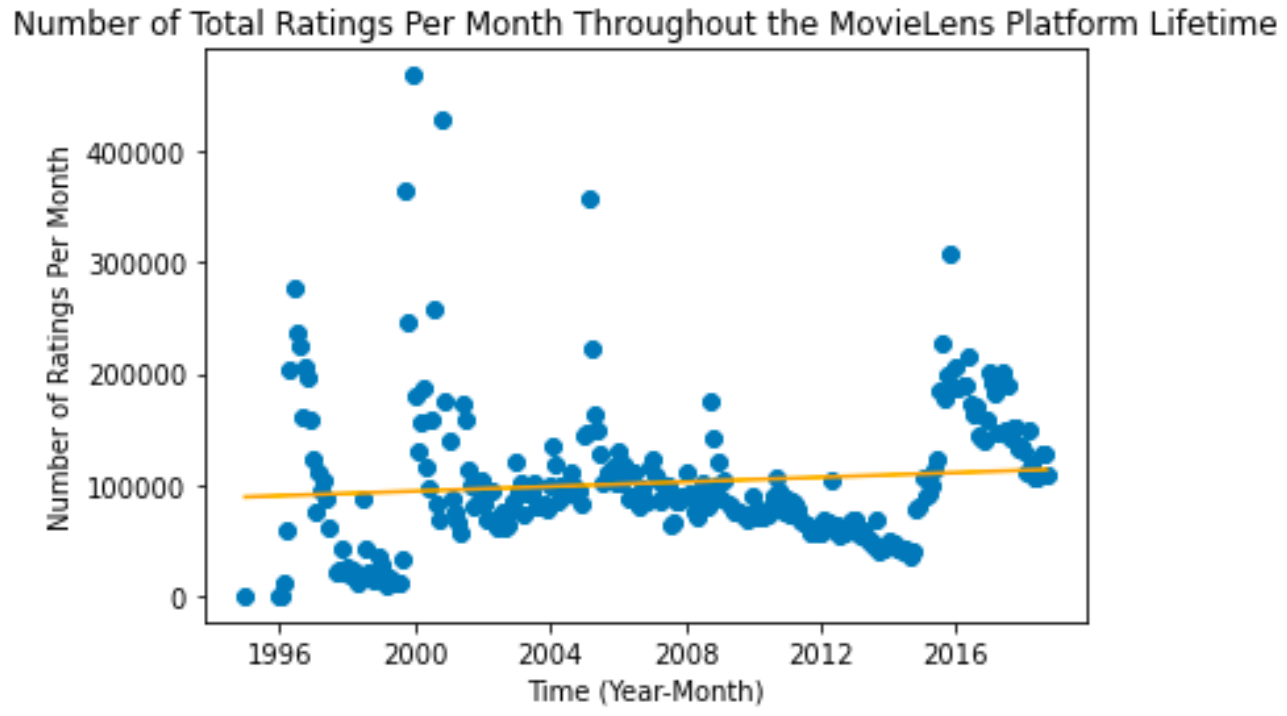
> The ratings features were further grouped by months_delta for each movie, allowing us to calculate the mean monthly ratings of each movie
> In a similar way, usermeanrating was produced as the monthly mean user_rating data for every movie, given the reviewers who reviewed that movie in that specific month
> For movies in movie_industry, we calculated the profit margin of movies as (gross - budget / gross)

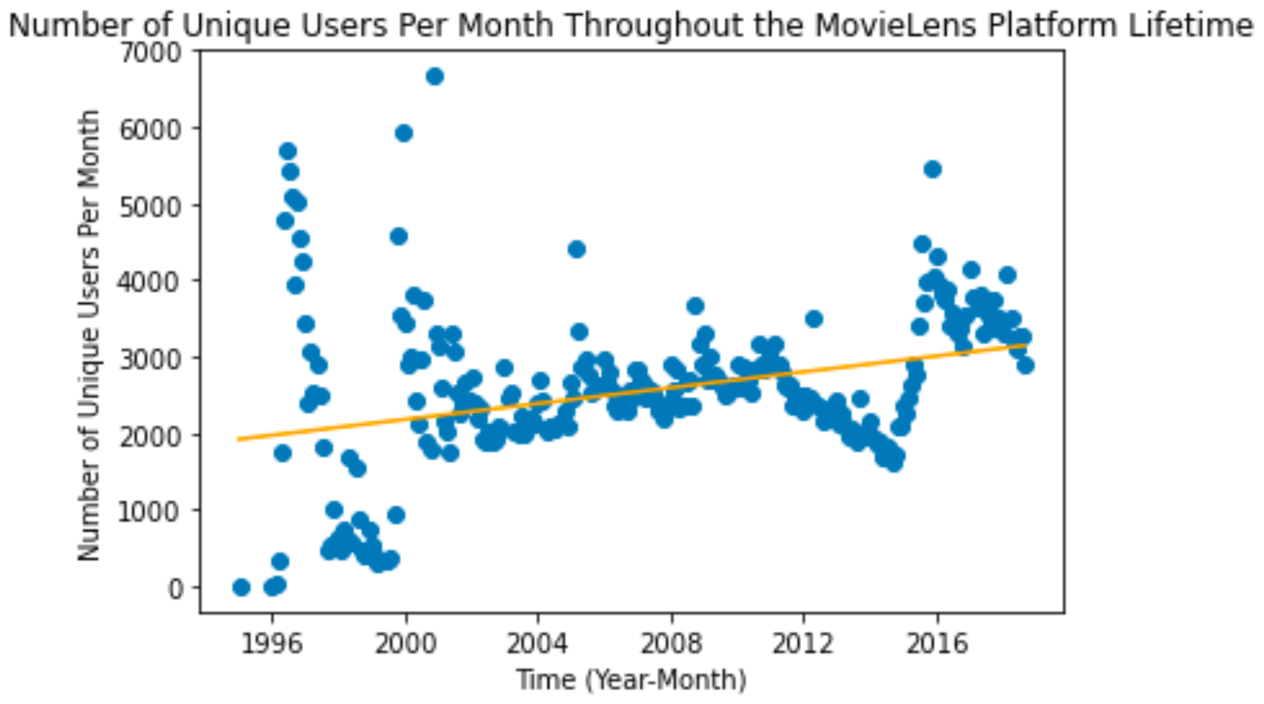
Tags Dataset

> We noticed that the 1128 tags were not vastly different (e.g. “zombie”, “zombies”). As such we used principal component analysis to reduce the 1128 tags to 100 tags, retaining 80% of variance.
>> For movies that did not have genome tag data, we set all tag relevances to 0 to preserve as much data as possible.

Exploratory Data Analysis

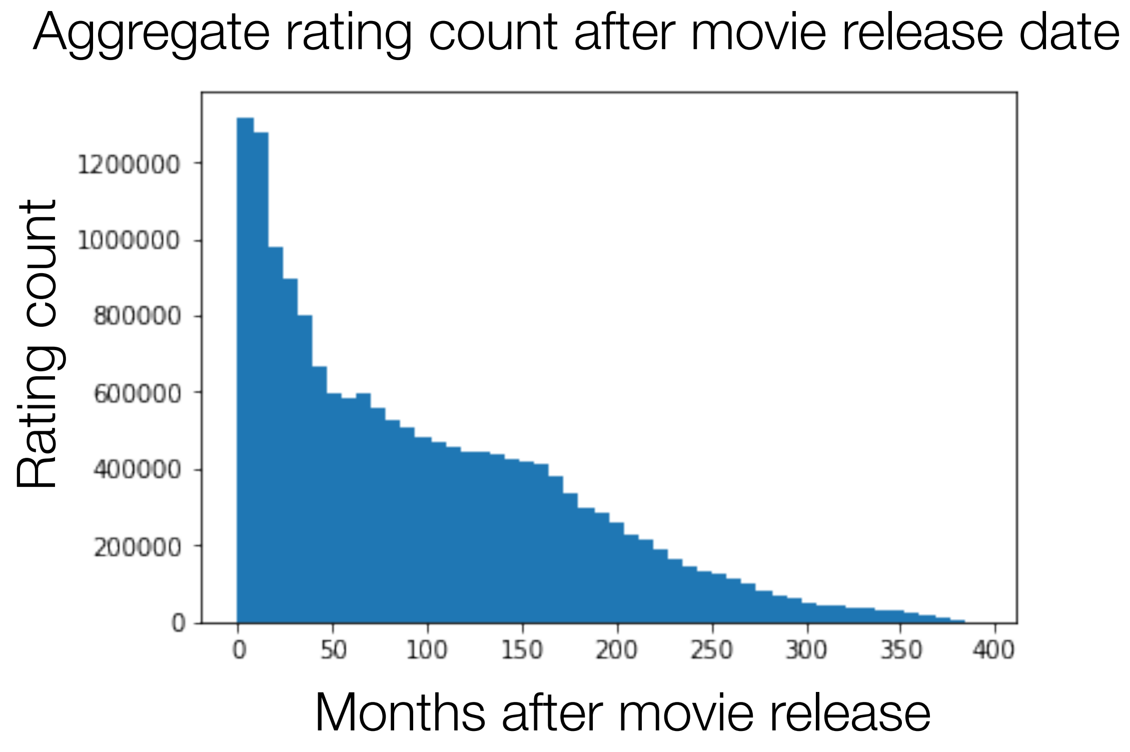
MovieLens Usage Trend





- We first began our data exploration by plotting rating information from the MovieLens dataset, to uncover potential trends, such as when the rating platform becomes more popular over time.
- Seasonality or platform-specific spikes might skew our analysis pipeline. Spikes are noticeable during 1997, 2000, 2005, and 2016
- By plotting rating count and unique users per month over time and performing a simple linear regression, we conclude that there is no significant trend and cyclicality in terms of platform popularity and usage.

"Early Adopters, Late Majority, and Laggards"

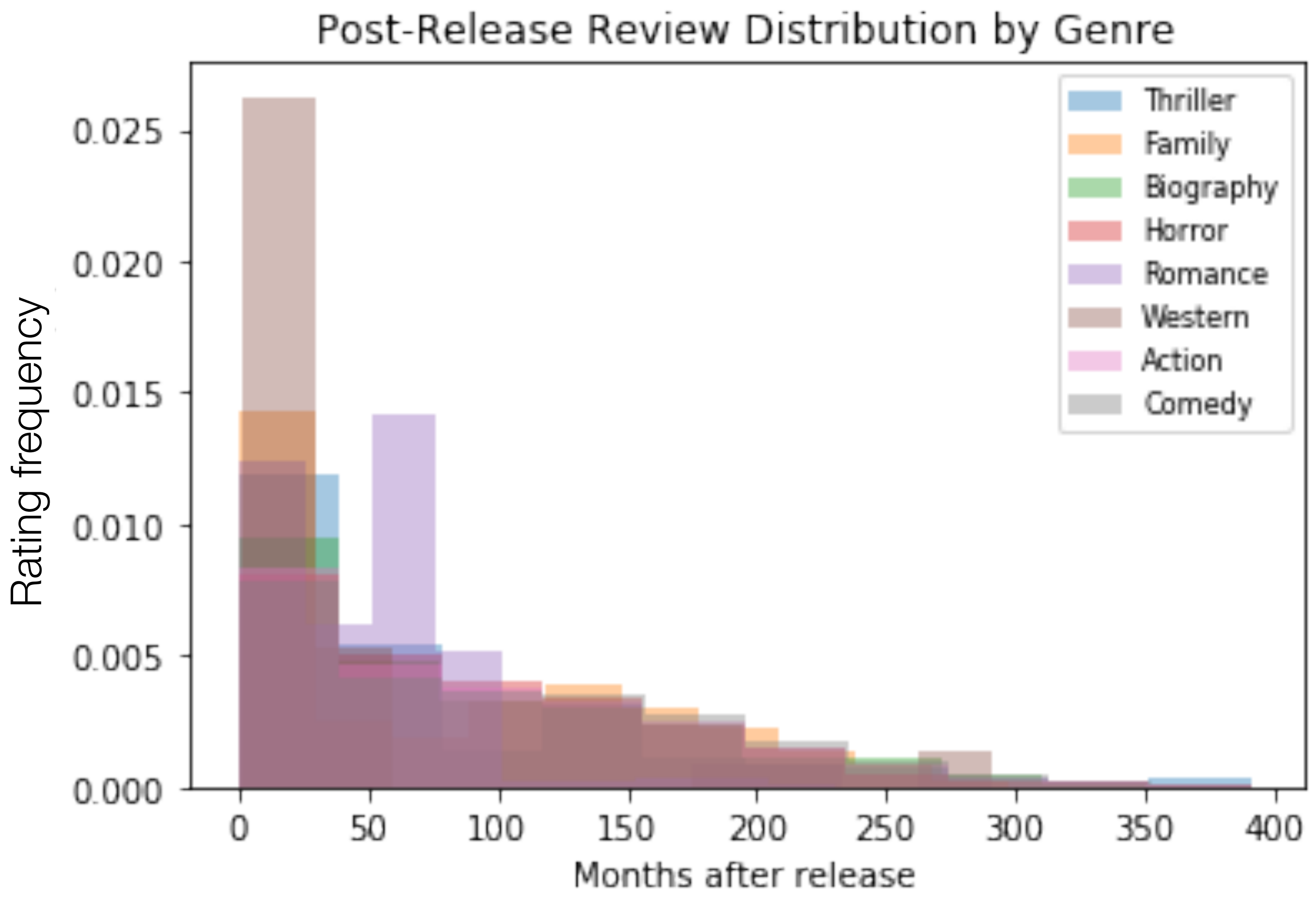


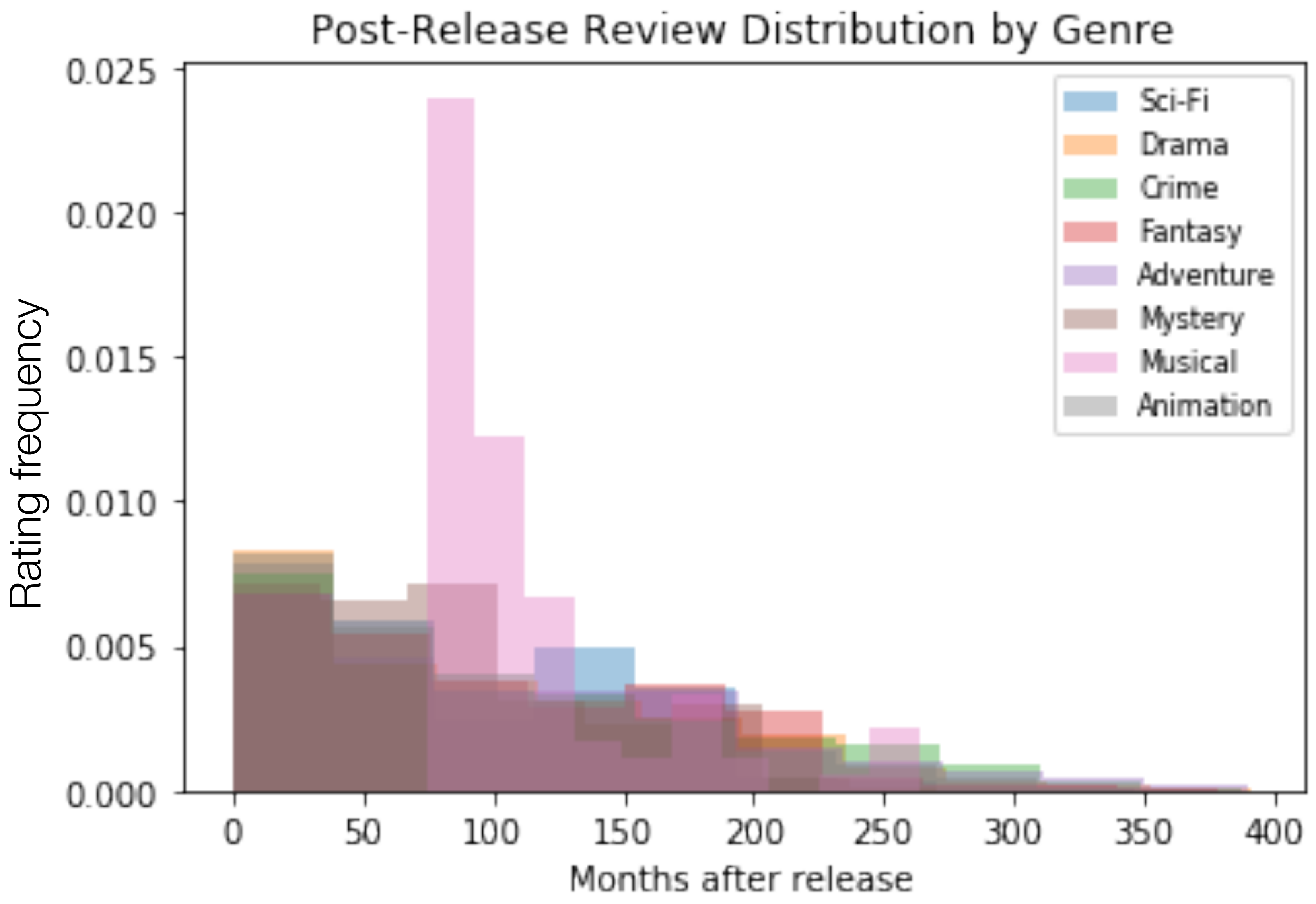
- Using the release date information in movie_industry.csv, we created a subset of 4012 movies released after 1995. For this subset, we found 1,714,6188 corresponding ratings, and using their timestamp information, grouped them based on the months elapsed from the release date of their corresponding movie. The above histogram was then produced.
- We were surprised to find that the ratings submitted within 1 year of a movie's release did not constitute a significant majority of the entire database; instead, a notable percentage of consumers continued watching and reviewing movies many years after the movie's release.
- Based on this, we hypothesized that there may be differences in consumer behavior between early watchers and late watchers, which may potentially arise from contextual differences such as watching a movie in the cinema v.s. on TV/online, being a fan for a specific movie v.s. being a casual watcher.

Exploratory Data Analysis

Cross-Genre Inter-temporal Differences

- Following our general investigation on the time-series distribution of movie reviewers after the release date of a movie, we investigated how such distributions may differ across genres. Using the “genre” column in movie_industry, we classified ratings based on genres, and plotted the distribution density of 16 genres in the 2 plots below.
- Through visual analysis, we can see that most genres get about <1% of their total ratings in the first 2 years of their release, with a gradual decline over the next ~20 years (240 months).
- Interesting outliers include:
  - Western: 2.5% of total ratings are registered within the first 2 years of the movie’s release, followed by an immediate decline
  - Romance: Rating frequency peak at 1.5% after around 4 years of the movies release.
  - Musical: Rating frequency peak significantly at 2.5% after around 6 years from the movie’s release, with a gradual decline for 6 more years.
- These outliers and trends may suggest that inter-temporal consumer behavior may vary across genres. For example, the abnormal frequency peak of the musicals genre may suggest that these movies have high rewatch value, and people do not mind watching them outside theaters.

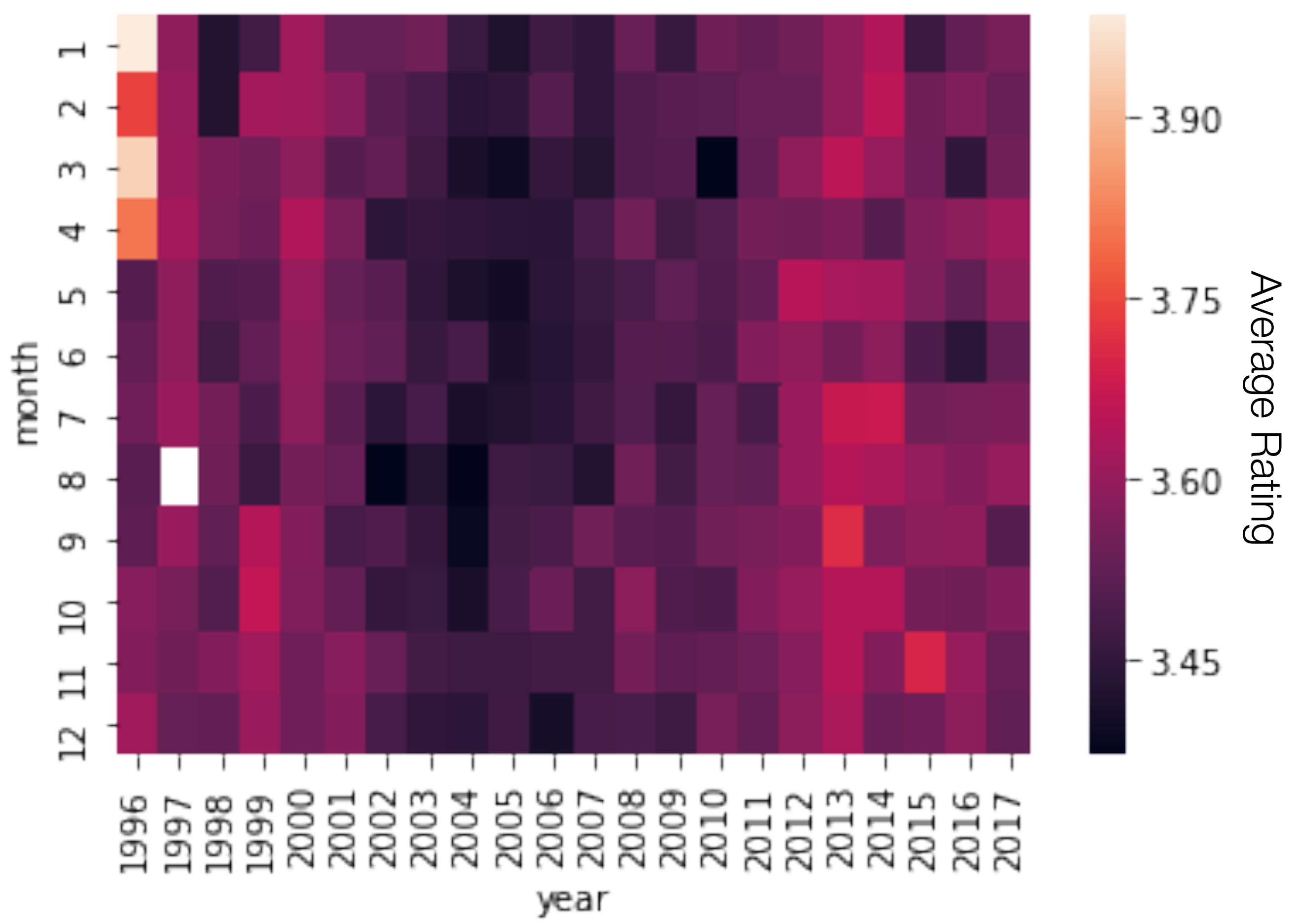




Exploratory Data Analysis

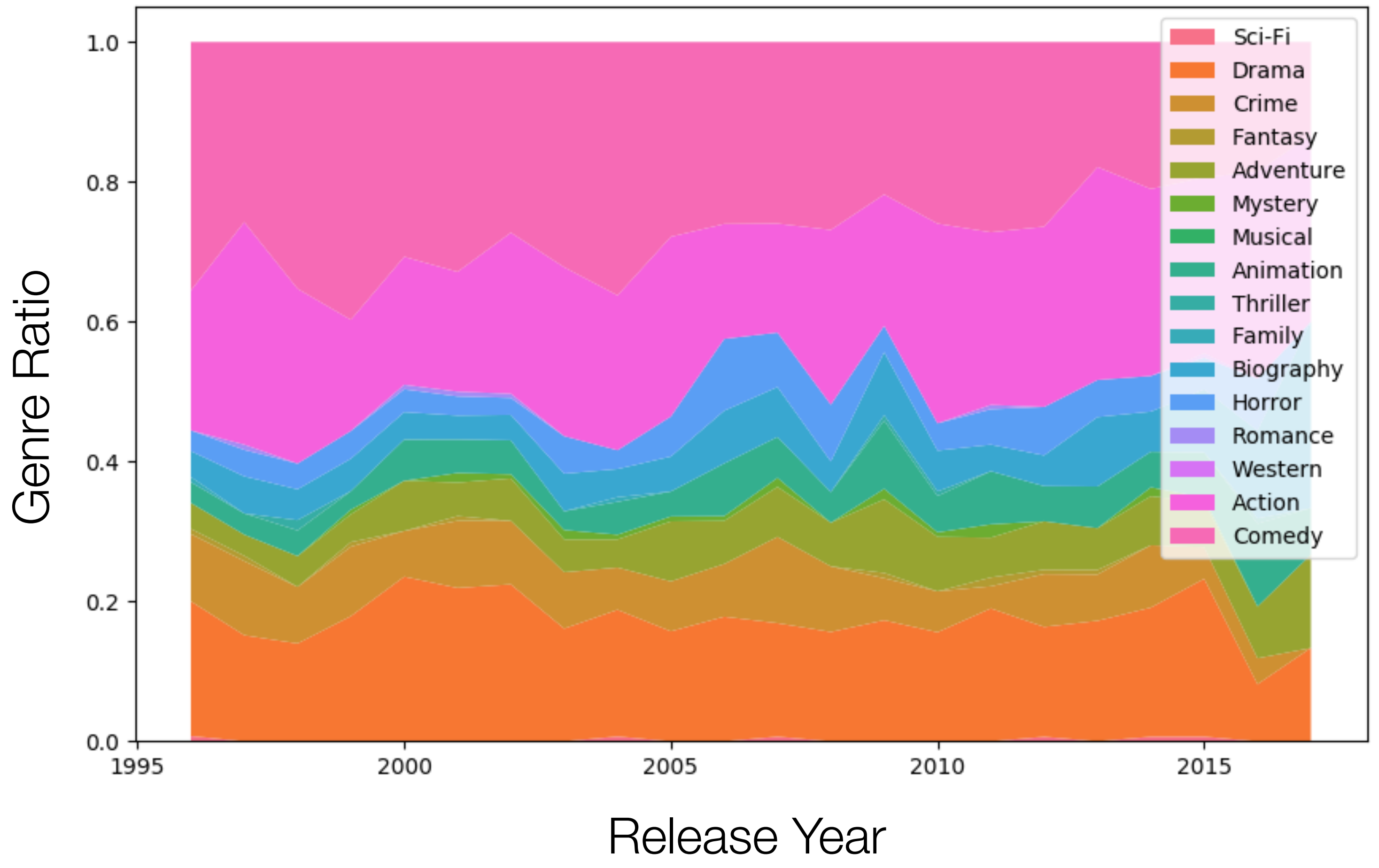
General Exploration

Average Monthly Rating over Time



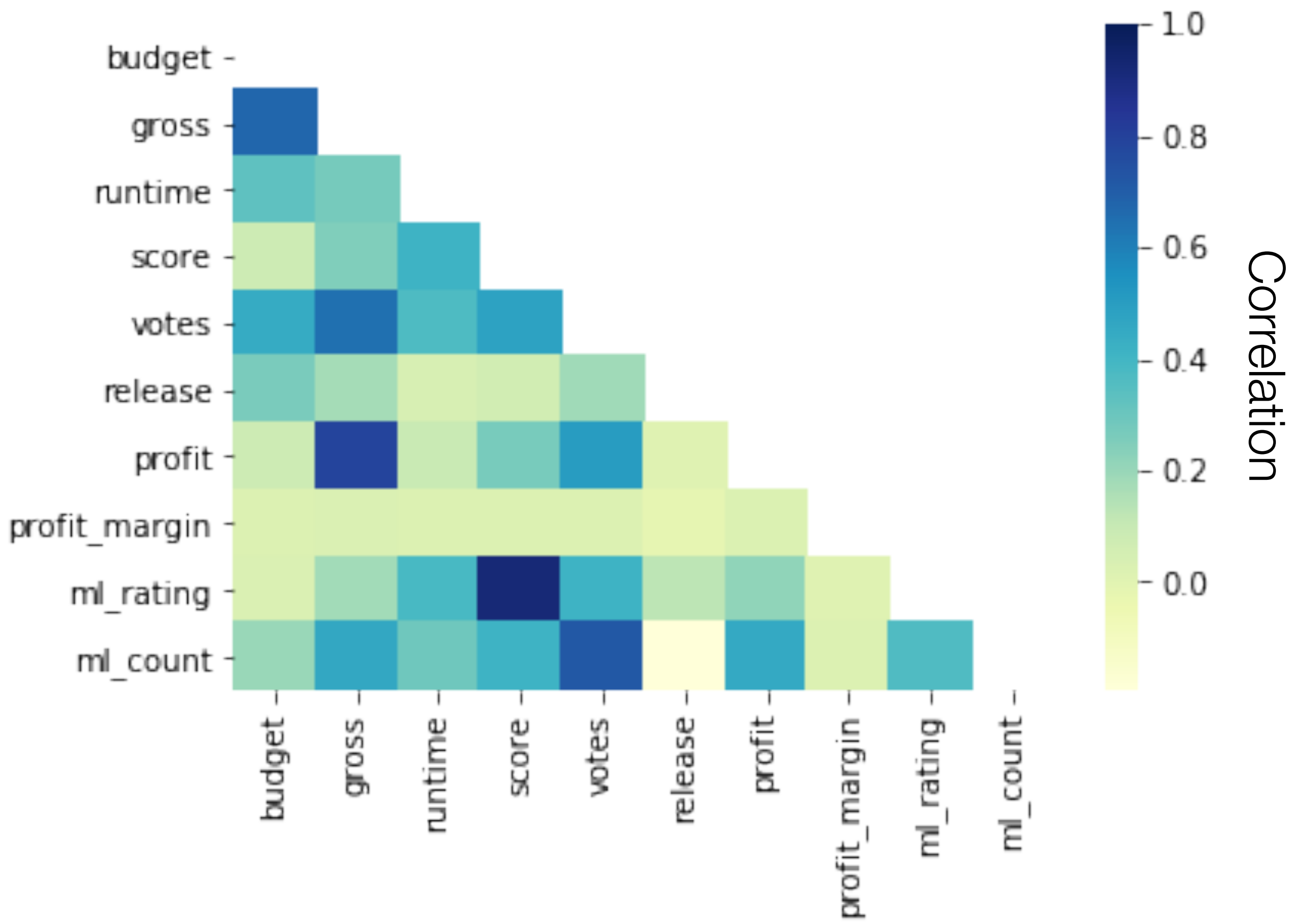
>To identify potential seasonality effects of consumer behavior within the span of our dataset, we plotted the above heat map and found that there is no clear seasonality trend across the years, apart from the minor observation that raters were slightly more critical between 2004 to 2007.

Movie Category Trend



>Given the genre data from movie_industry, we investigated the distribution of movie genres from 1996 to 2016. We see that comedy, action, and drama are the most frequent movies across all these years, with other genres consistently occupying a smaller percentage.

Movie Feature Correlation



>In understanding the various numeric features in both the MovieLens and movie industry datasets, we observed that the there is high correlation between the MovieLens average rating (ml_rating) and the score in the movie industry data. As such we decided to drop the average MovieLens score. Similarly, gross correlates highly with profit (an engineered feature) and budget, thus we decided to drop this feature too since it is already captured by profit and profit_margin.

Topic Question

Understanding the **inter-temporal** behavioral differences in movie reviewers, and how such traits may vary across **different types** of movies.

> Firstly, we wish to understand how the general perception of a movie changes over time. The key proxy for general perception is the average monthly rating of a movie. We hypothesize that two factors are associated with the monthly ratings of a movie:

- > Time elapsed between movie's release and rating
  - > We hypothesize that the characteristics of reviewers differ across time, leading to self-selection of rating types and biases.
  - > For example, the consumers rating a movie within the first few months of its release may have wanted to watch the movie since its announcement or really enjoy the cinema experience. In contrast, consumers rating a movie many years after its release may only have stumbled upon the movie on Netflix, and/or have viewed it on a smaller screen.
- > General profile of reviewers rating the movie in a given month
  - > We hypothesize that the general profile of reviewers, or how the average reviewer in a group rates an average movie, has a high impact on the monthly rating of the movie. This measures variability across different reviewer types.

> Secondly, having gained initial insights from the impact of time on monthly movie ratings, we wish to further investigate whether such impacts might differ across movie categories and related features, such as genres.

- > We hypothesize that the inter-temporal impact of movie ratings differ across types of movies, and we explored if we could predict such trend differences using movie features.
- > For example, the initial ratings from a franchise-movie (e.g. MCU) might be abnormally high compared to the rest of its time-series data given initial hype and franchise perception, leading to a notable decline over time; in contrast, certain categories of movies may be considered "hidden gems", and only gets popular after an initial dormant period, when it resurfaces on streaming platforms.

Model 1: Multiple Linear Regression to Predict Monthly Average Rating

$$y_t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i$$

Model Explanation

y_t monthly mean rating of a specific movie

x_1 months_delta: number of months since the movie was first rated

x_2 usermeanrating: mean user_rating across users who rated the movie in the given month

- We used multiple linear regression to assess the impacts of two variables on the monthly mean rating of movies
- We used this model for every movie in our movie dataset that was released during and after 1996, had at least 24 months of ratings data (since factors such as theatrical release and movie awards are likely most evident within the first 12 months; our earlier rating count graph also showed that rating count began decreasing significantly after 24 months), and at least 500 ratings, using ordinary least squares.

- Assumption: No multicollinearity between input variables
  - We regressed the two variables and found no significant correlation
- Assumption: Residuals are normally distributed
  - We plotted residuals and found a roughly normal distribution

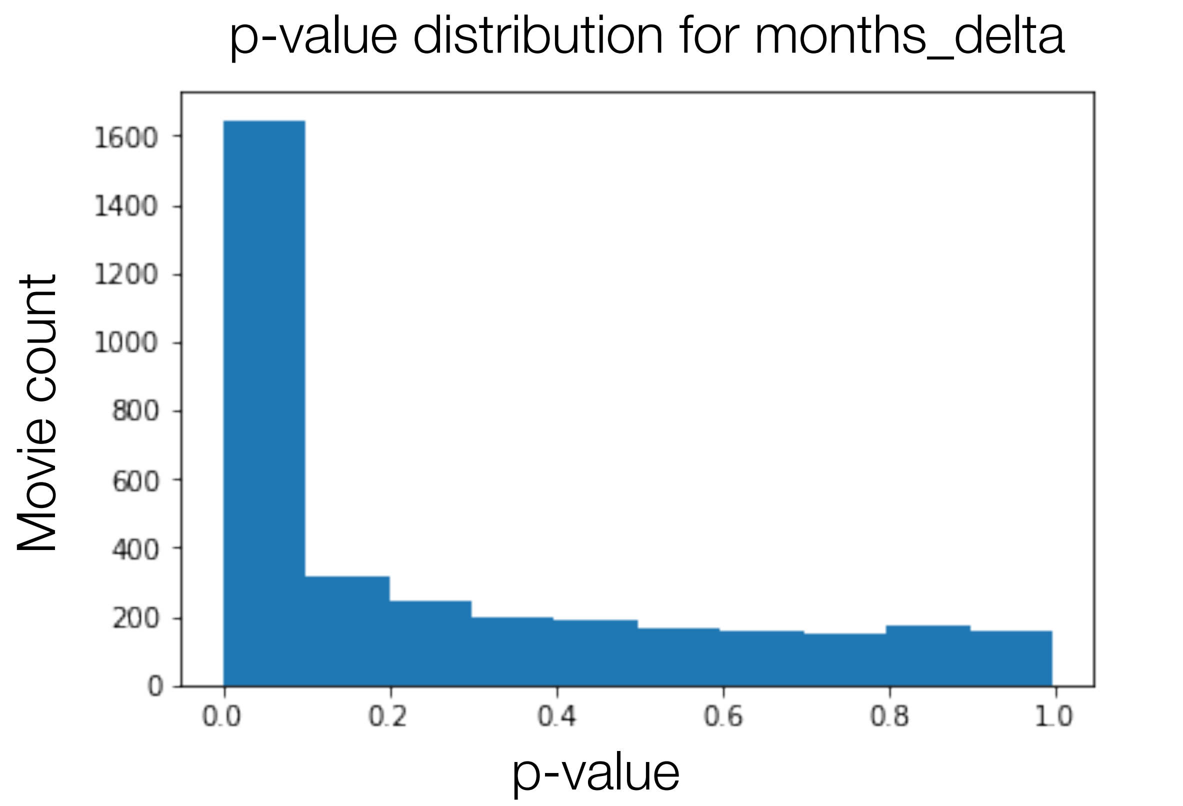
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 rating   R-squared:                       0.346
Model:                            OLS   Adj. R-squared:                  0.341
Method:                 Least Squares   F-statistic:                     71.13
Date:                Sun, 04 Oct 2020   Prob (F-statistic):           1.60e-25
Time:                        00:01:04   Log-Likelihood:                 154.37
No. Observations:                 272   AIC:                            -302.7
Df Residuals:                     269   BIC:                            -291.9
Df Model:                           2
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
months_delta      -0.0013      0.000    -10.868      0.000      -0.002      -0.001
usermeanrating     1.0637      0.114      9.322      0.000       0.839       1.288
ones               0.1992      0.408      0.489      0.625      -0.603       1.002
==============================================================================
Omnibus:                        8.050   Durbin-Watson:                   0.693
Prob(Omnibus):                  0.018   Jarque-Bera (JB):                7.966
Skew:                          -0.411   Prob(JB):                       0.0186
Kurtosis:                       3.162   Cond. No.                     8.02e+03
==============================================================================
```

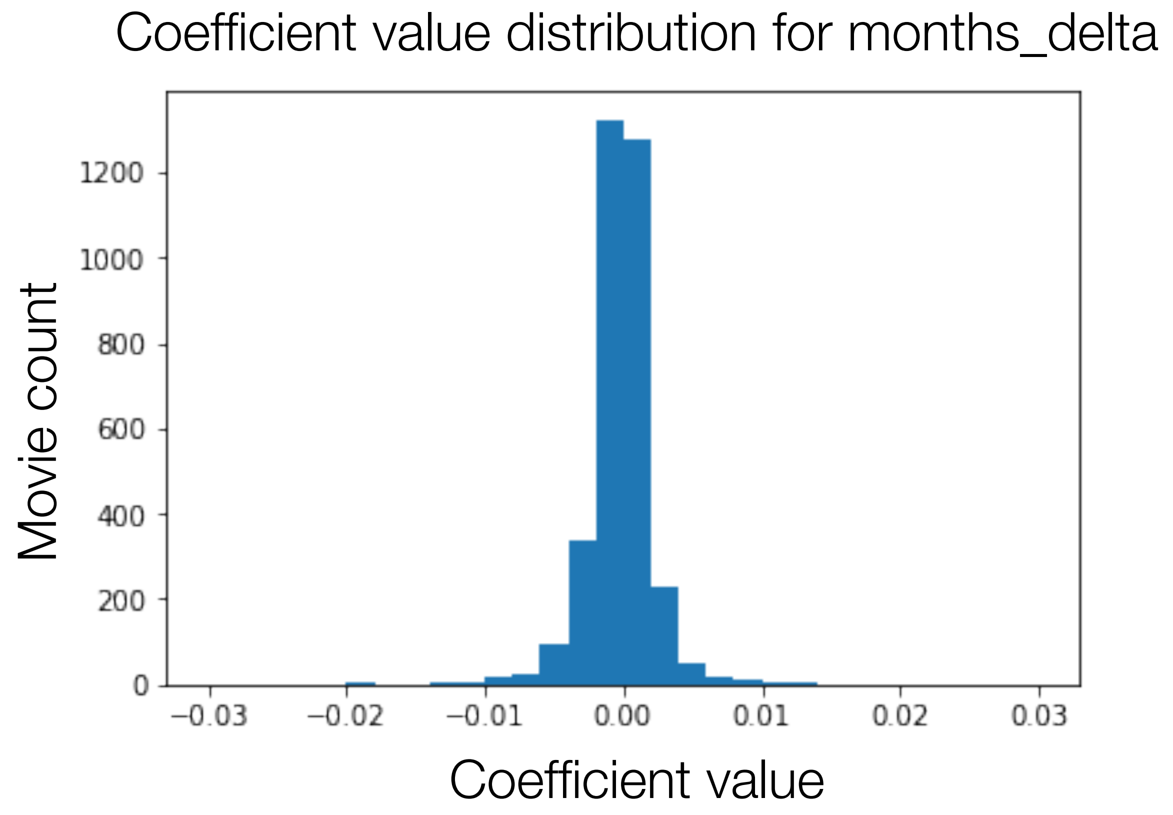
Example OLS results from one movie: Toy Story

- This example results shows that months_delta has a slight negative impact on monthly mean ratings, suggesting that the mean monthly ratings decreased over time
- The >1 usermeanrating suggests that this particular movies is one which raters on average tend to rate higher than their usual rating behavior
- The low p-values associated with both variables suggest high predictability of both factors

Model 1: Multiple Linear Regression to Predict Monthly Average Rating

Predictor: months_delta





> Although each movie has a different set of months_delta data, varying from 24 months to 200+ months, we find that the p-value distribution of months_delta across these movies suggest that the latency of movie ratings do have a significant effect on the monthly mean ratings of a movie.

> There is an almost even split on the direction of this impact, with a slightly negative skew.

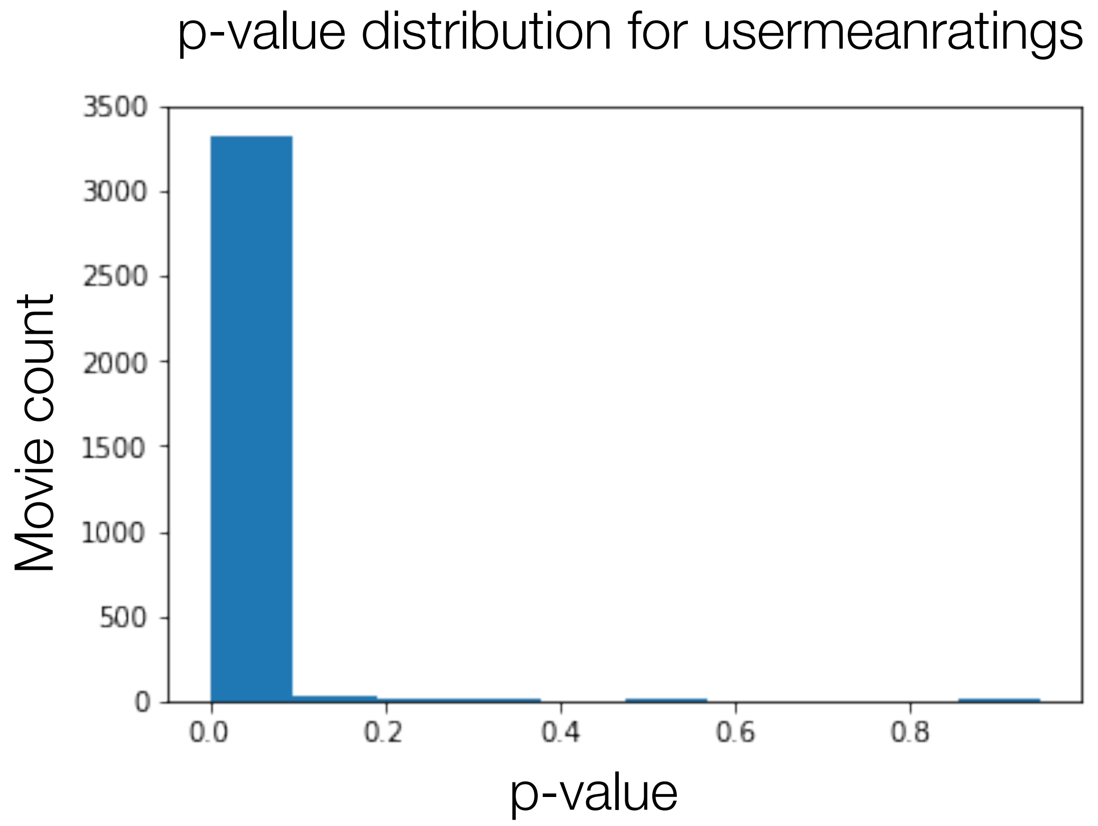
> This suggests that the monthly ratings of a movie tend to decrease, which might be explained by factors such as self-selection of early watchers, and the positive effect of cinema experience.

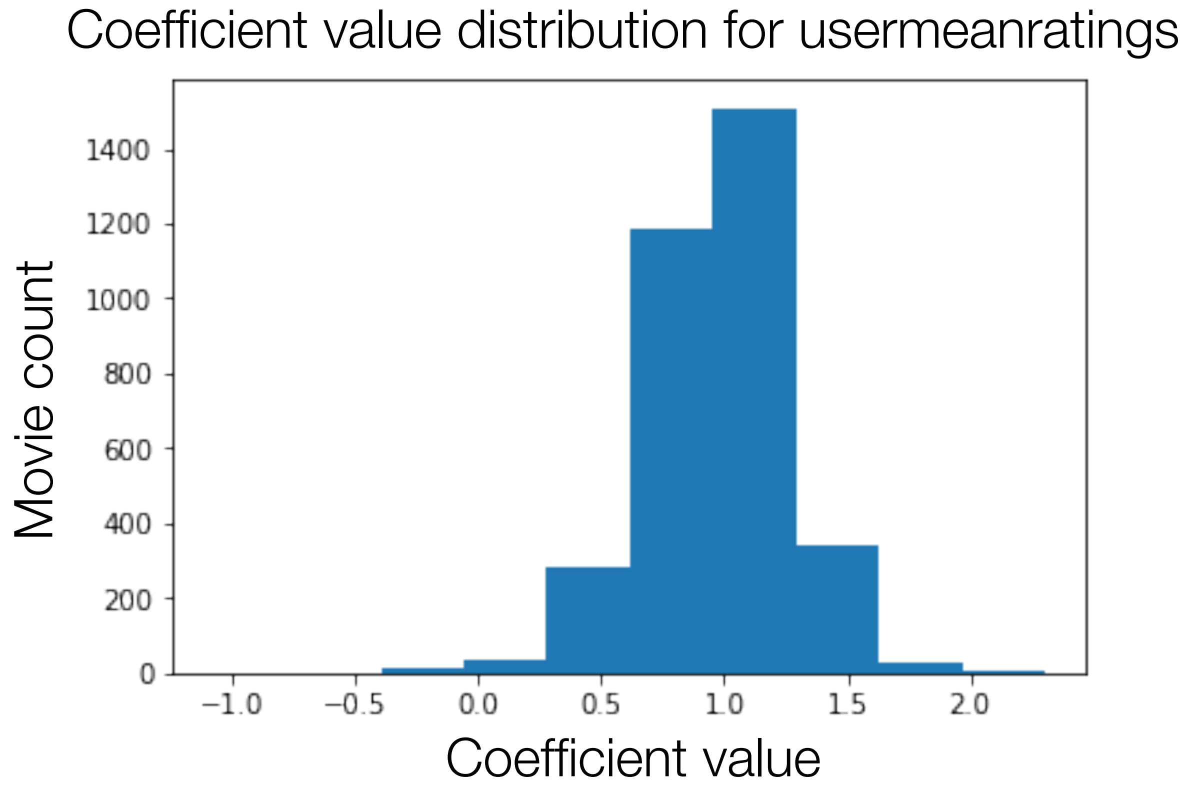
> Nonetheless, the variability of these coefficients warrants further genre and feature based analysis.

> The distribution of p-values from usermeanratings suggests that the average user_rating across all movies given by users rating a specific movie in a month has a highly predictive impact on the monthly rating of that movie.

> That the distribution of coefficients is largely positive and centered around 1.0 suggests that most movies in this subset are within user expectations, with some movies either exceeding or going below expectations.

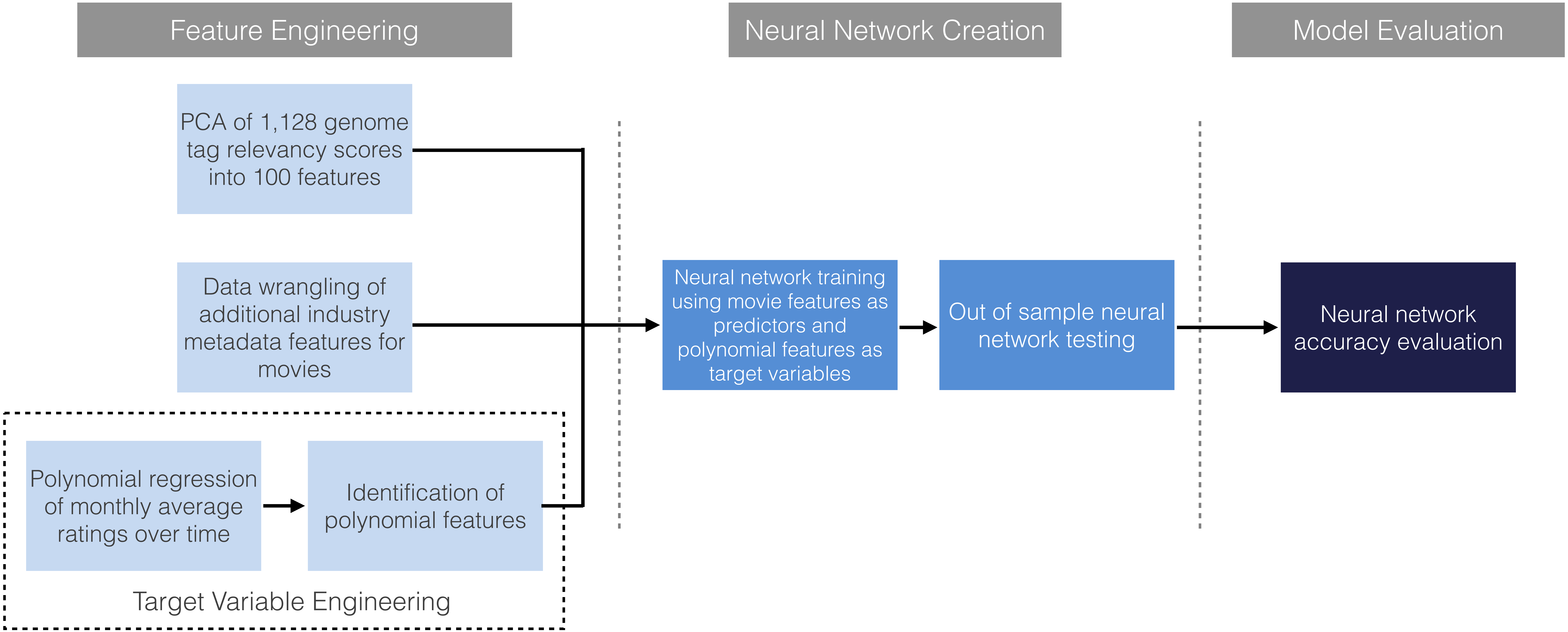
Predictor: usermeanratings





Model 2: Predicting Total Movie Rating Variability Over Time

Model Pipeline Overview



Model 2: Predicting Total Movie Rating Variability Over Time

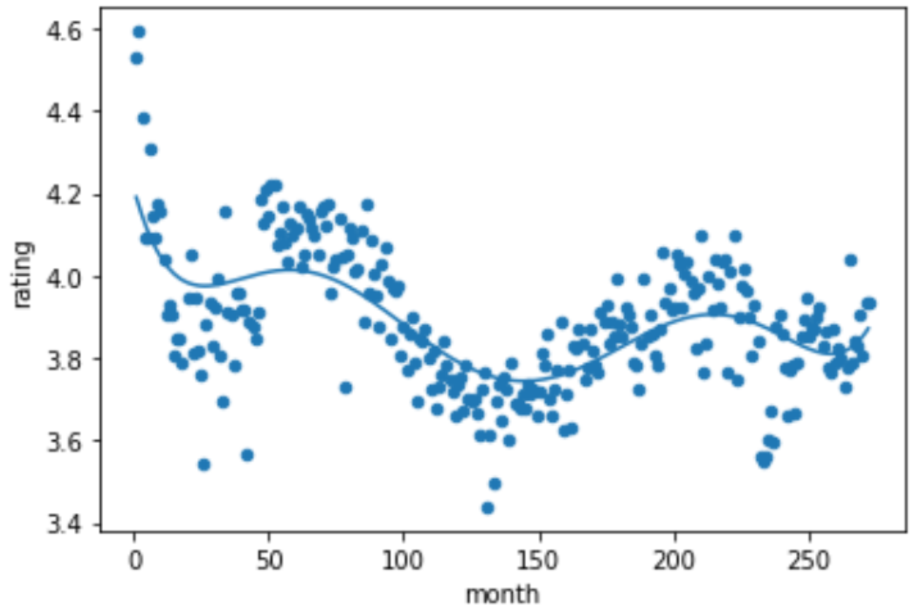
Step 1: Feature Engineering

> When investigating average movie ratings as a function of time, we found some intriguing relationships. We contemplated several ways of quantifying the shape of the relationship, including the degree of the polynomial that best fit the data (using regularized linear regression), the total variation in the ratings over time, the location of extrema, average months between extrema, etc.

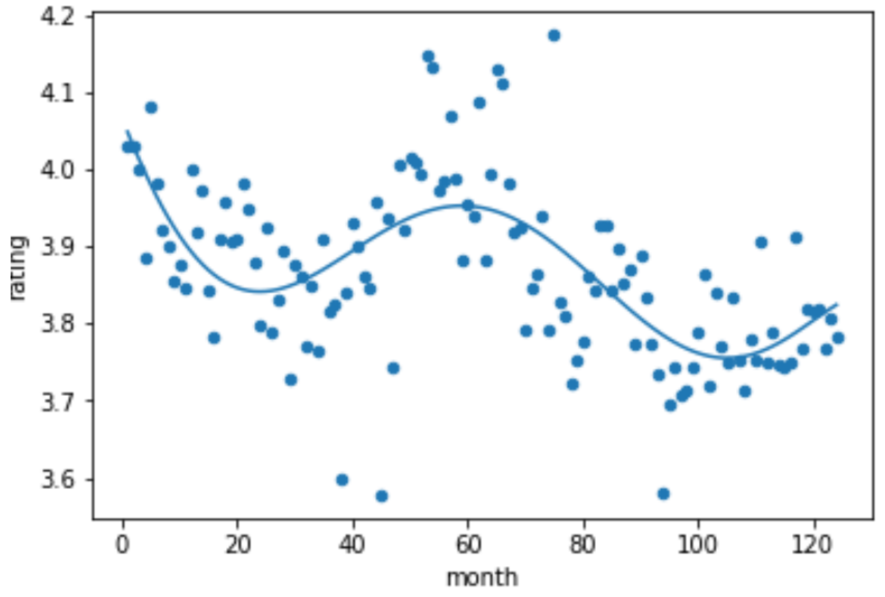
> We decided that the two best ways to capture the intricate relationships in our data were to compute the total variation in the function that best fit the data, as well as the net rating change over time. Because we calculated both of these metrics using the fit polynomial, they are insulated against random variation.

> We fit the ratings data for each movie with a sixth degree polynomial to capture as much of the relationship as possible without overfitting. We chose degree six because we consistently found, visually, that the polynomial relationship between ratings and time for individual movies did not exceed this degree.

Monthly ratings over time for Iron Man

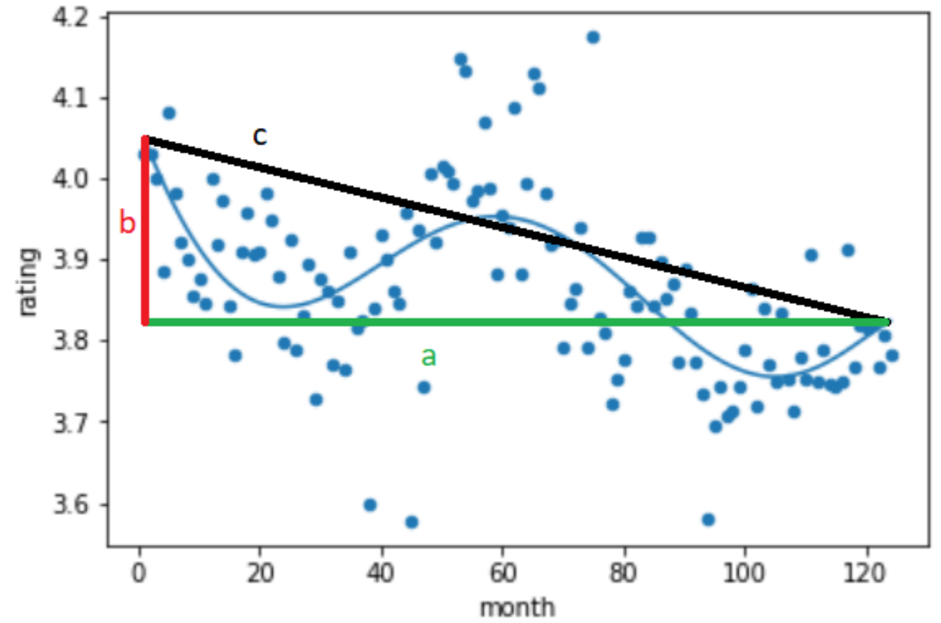


Monthly ratings over time for Toy Story



> We identified two key metrics (target variables) to identify the features of our fitted polynomials:

Feature	Metric	Intuition
Monthly Rating Variation over Time	Arc length of polynomial within a specified month range	Characterizes the volatility of consumers' perception towards a movie over time
Long Term Rating Trend (over n months)	f(n) - f(1), where f is the fit polynomial	Characterizes the direction and magnitude of change in consumers' perception towards a movie over time



In the sample graph above, b denotes our metric for long term rating trend

$$\int_1^n \sqrt{1 + (\frac{dy}{dx})^2 dx}$$

Formula for polynomial arc length

Model 2: Predicting Total Movie Rating Variability Over Time

Step 2: Artificial Neural Network

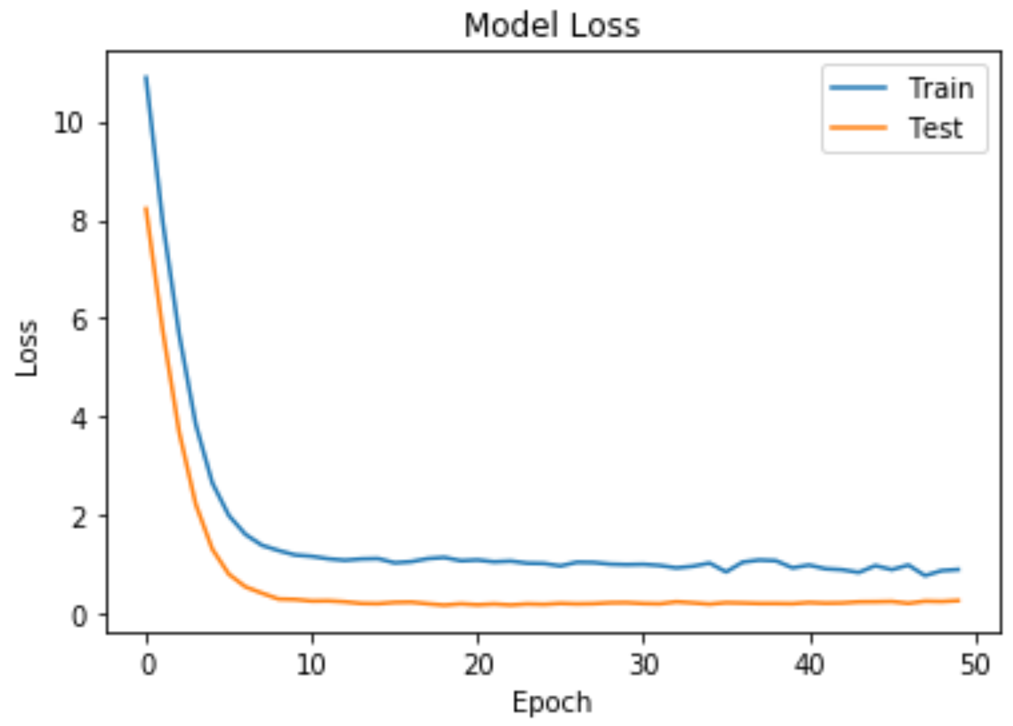
- With the two target variables identified, we chose a number of movie features from both the genome tag database and movie industry metadata database as predictors.
  - Our predictor variables are:
    - Industry metadata: Genres, budget, profit margin, score, votes, runtime, release year, profit
    - Genome tag data: 100 principle components generated from 1,128 genome tags
  - In particular, we dropped gross revenue and MovieLens net ratings due to high correlation with budget and movie industry ratings respectively. We also dropped categorical data such as writer, director, star actors, and production company due to high variability across movies. We also dropped features, such as country, that intuitively would not produce any predictability.
- We chose to use neural networks to capture the unknown nonlinear relationships between our features and output variables.
- We ran two neural networks for each time frame; one for rating variation and one for net rating change. Our time frames ranged from 4 months, 12 months, 24 months, and 90 months after the movie’s release. That is, we ran our models while considering months between a movie’s release month and 4 months after release, between a movie’s release month and 12 months after release, etc.

Model 2: Predicting Total Movie Rating Variability Over Time

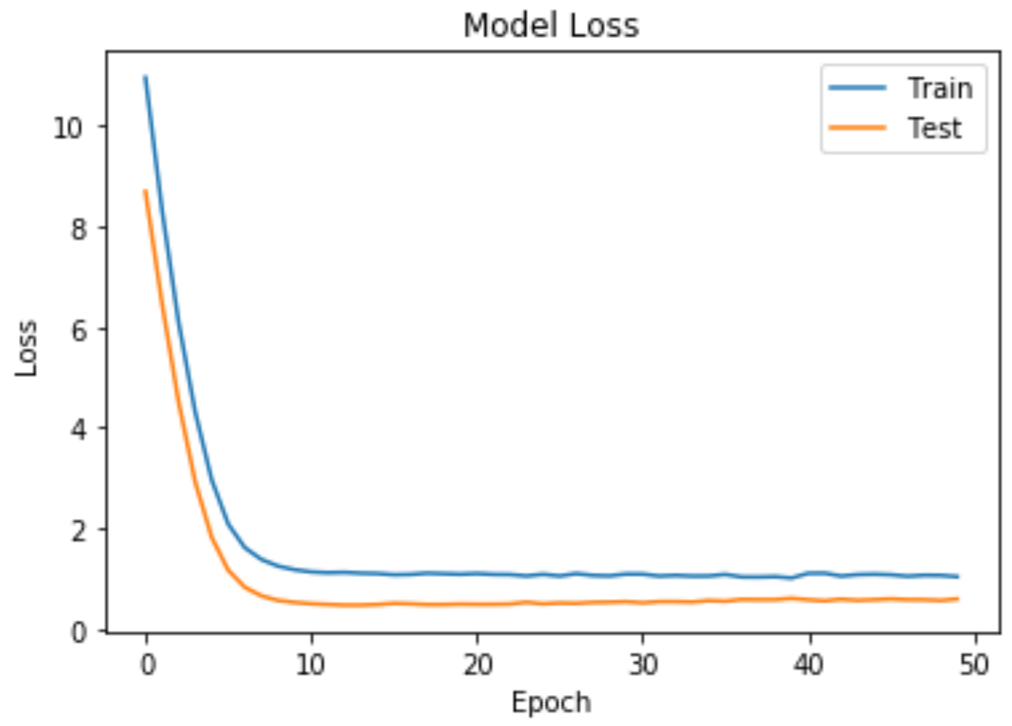
Results (4 months & 12 months)

- For both 4 months and 12 months timeframes, our models losses converge at around 10 epochs.
- The root mean square error (RMSE) values reported are in number of standard deviations away from the true value that our models deviate from.

Model Loss for Monthly Rating Variation over 4 Months

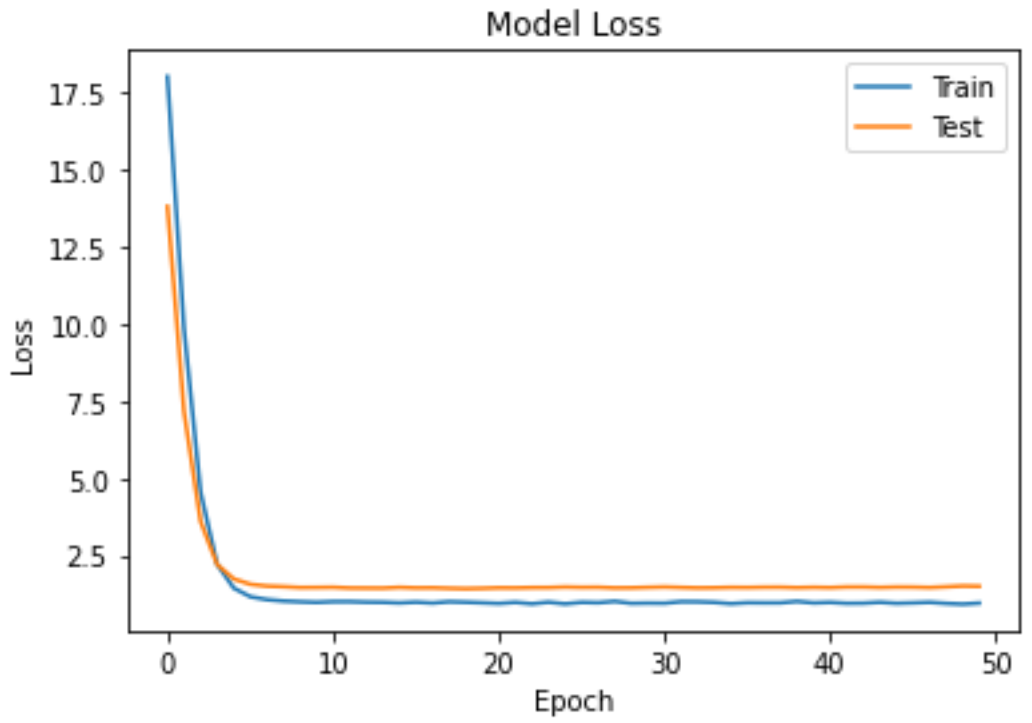


Model Loss for Long Term Rating Trend over 4 Months

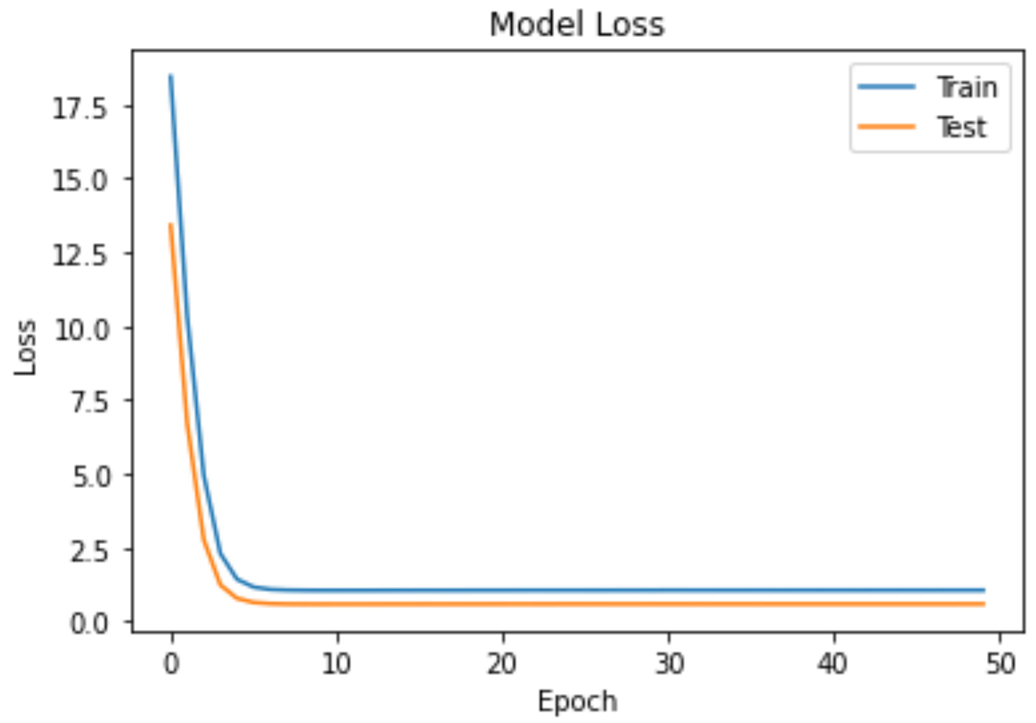


Final Train RMSE for Model One: 0.872
Final Train RMSE for Model Two: 0.966
Final Test RMSE for Model One: 1.182
Final Test RMSE for Model Two: 1.032

Model Loss for Monthly Rating Variation over 12 Months



Model Loss for Long Term Rating Trend over 12 Months



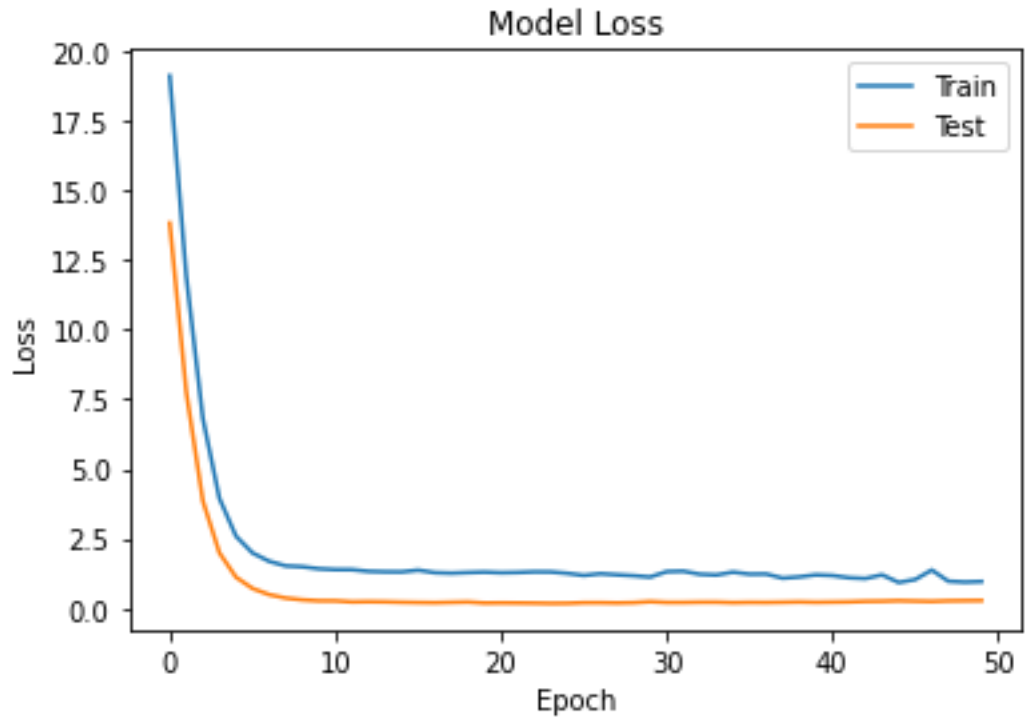
Final Train RMSE for Model One: 0.709
Final Train RMSE for Model Two: 1.011
Final Test RMSE for Model One: 1.122
Final Test RMSE for Model Two: 0.888

Model 2: Predicting Total Movie Rating Variability Over Time

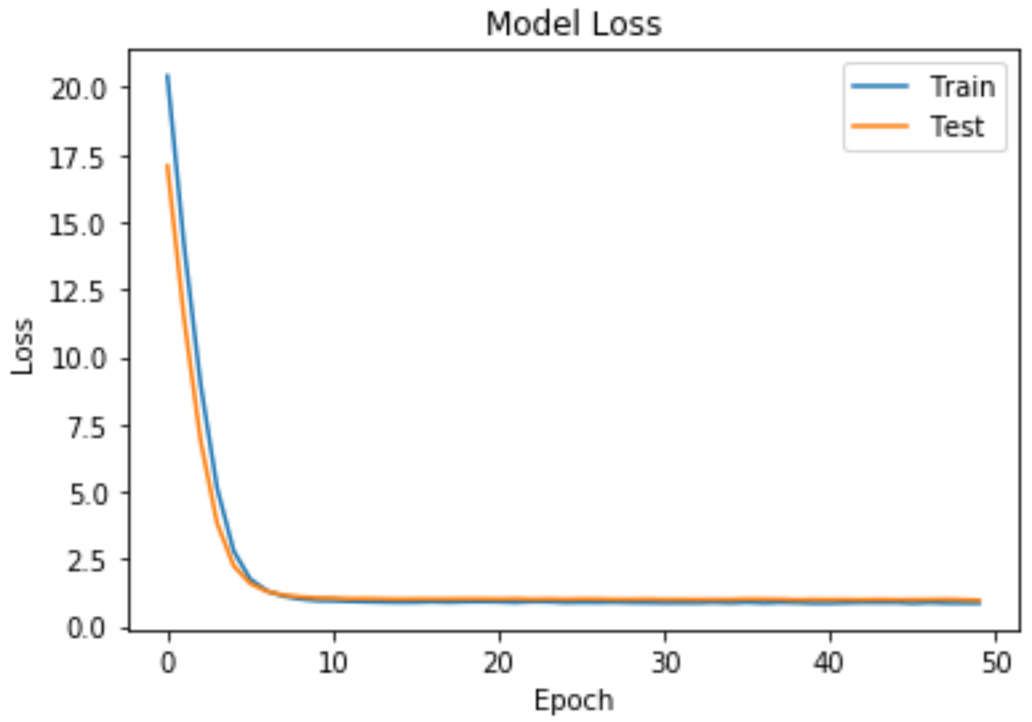
Results (24 months & 90 months)

- For both 24 months and 90 months timeframes, our model loss also converge at around 5 epochs.
- Although the RMSE values are still near 1 standard deviation, we can see a slight trend in that the model accuracy increases as we increase the timeframe to a longer-period.

Model Loss for Monthly Rating Variation over 24 Months

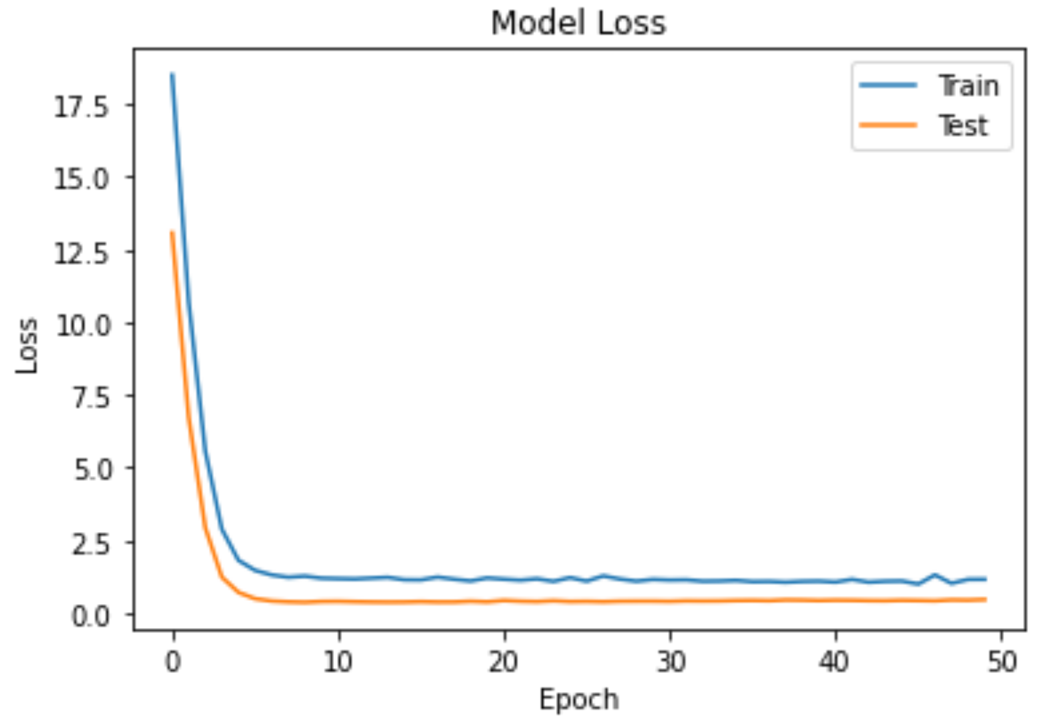


Model Loss for Long Term Rating Trend over 24 Months

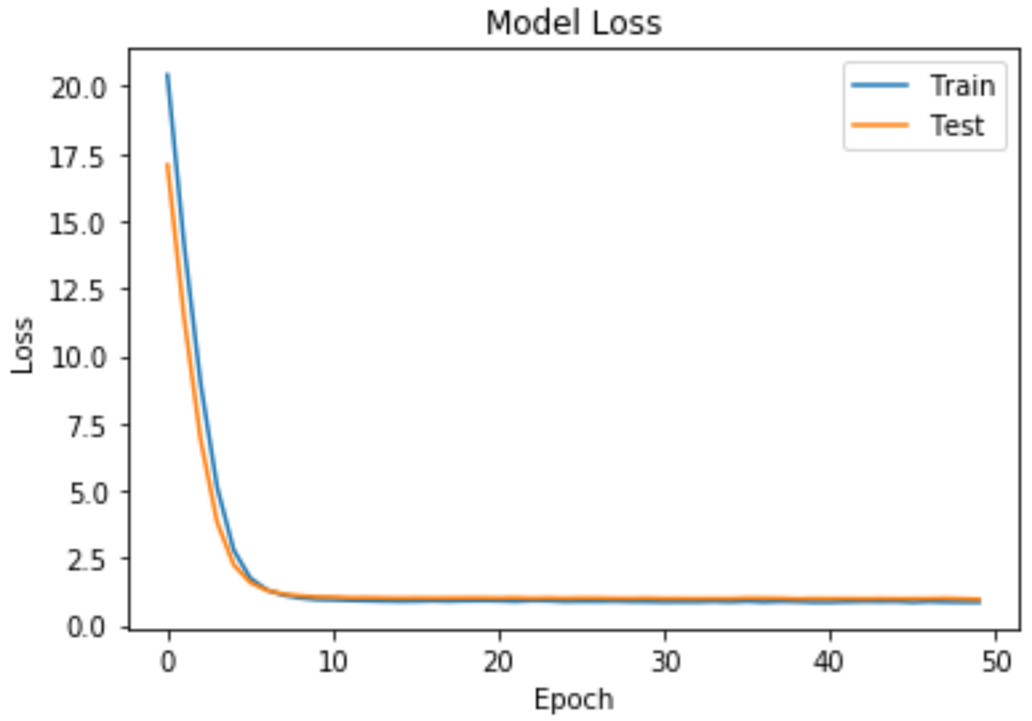


Final Train RMSE for Model One: 0.742
Final Train RMSE for Model Two: 0.942
Final Test RMSE for Model One: 0.885
Final Test RMSE for Model Two: 1.078

Model Loss for Monthly Rating Variation over 90 Months



Model Loss for Long Term Rating Trend over 90 Months



Final Train RMSE for Model One: 0.993
Final Train RMSE for Model Two: 0.830
Final Test RMSE for Model One: 0.655
Final Test RMSE for Model Two: 1.087

Model 2: Predicting Total Movie Rating Variability Over Time

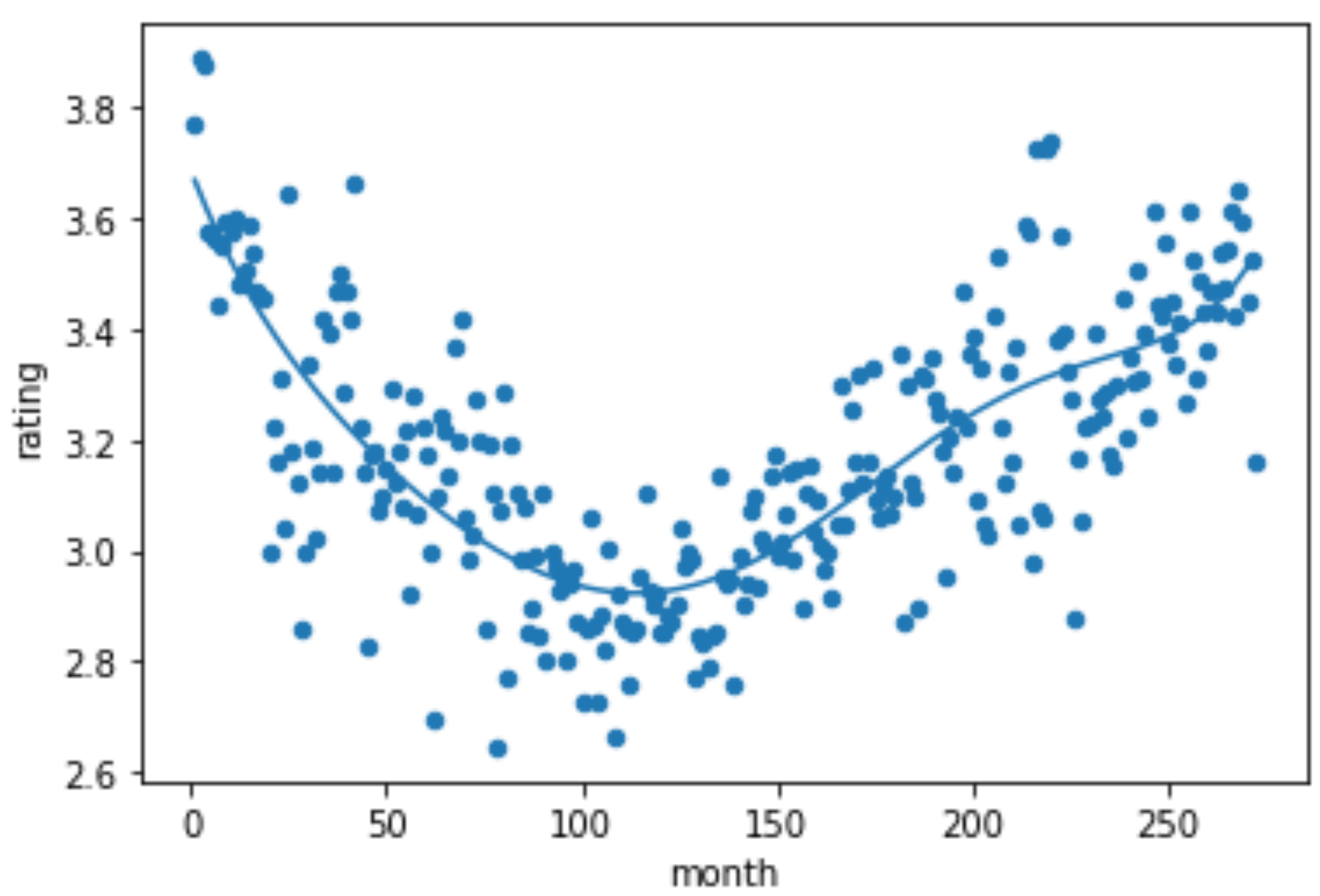
Conclusions

- As a whole, our artificial neural network model is able to make relatively weak predictions (< 1 standard deviations) for both the rating variability of movies over time and net rating change over time using features including profit margin, runtime, release year, and subject tags.
- However, a trend we observed through varying the time frame length is that the rating variation model performed significantly better as we increased the timeframe from ~1 year to ~8 years, suggesting that the features are more predictive of a movie's long term rating variability as opposed to a short period after a movie's release.
- The features inputted to our model was movie metadata (e.g. budget, genre, tags, etc.). Given the high amount of diversity between movies, our model indicates that features that are specific to the movie itself will create a more accurate model. This movie-specific data includes but is not limited to: caption transcripts, plot summaries, and written critic reviews. In addition, understanding the breakdown of specific metadata could uncover important trends. For instance, understanding the marketing sub-portion of budgets or the sub-genres of movies could substantially increase our model's accuracy.

Model 2: Predicting Total Movie Rating Variability Over Time

Key Insights

Monthly ratings over time for Jumanji



We see a distinct shape in this graph for ratings over time for Jumanji. This relationship exemplifies many such relationships we saw for individual movies.

- We showed that there were distinct and significant trends in the relationship between a movie's ratings and the time elapsed since its release.
- Our features were unable to capture why these trends exist, but other features we did not use may be able to predict them well.
- Given a model that can accurately predict rating variability over time and net rating change over time, we can derive additional insights. Such a model can be used to advise decision-making for streaming platforms, such as identifying types of movies that gain renewed interest among consumers over the long-term, and the consistency of consumer perceptions over time. The model can also be used to understand consumer behavior towards movies, predicting the self-selection of viewer types as a function of elapsed time.
- These results can then be taken into consideration by movie recommender systems to improve user satisfaction and retention.
- For movie producers, such a model can advise on movie features that decrease rating variability over time. For Hollywood Production movies, most operate with very low profit margins and high volatility. Having a model such as ours can advise on movie features that decrease rating variability over time — a much-needed trait to secure funding.

Model 2: Predicting Total Movie Rating Variability Over Time

Future Research

- One obstacle we faced is the size of our dataset. Although we began with ~58,000 movies with user ratings, only ~33,000 movies were released after 1996 (when the ratings began), ~13,000 movies had genome tag relevancy scores, and ~7,000 movies had industry metadata. Ultimately, our usable dataset post-wrangling and merging was scaled down to ~4,000 movies, which hindered our analysis.
- For further analysis, we could scrape more industry metadata to increase the size of our training and testing datasets.

Monthly Average Rating Across Movies

> Across various category and types of movies, we found that the time elapsed since the release of a given movie has a high significance for predicting the movie's monthly average ratings. Similarly, the rating profile of viewers in that given month is also a highly significant predictive factor.

> The mean of the effect sizes for time elapsed is slightly negative, suggesting that on average, movies experience decreasing monthly ratings over time.

Predicting Monthly Rating Trend Using Movie Features

> We created an artificial neural network model aimed at predicting monthly rating trends based on movie features, which was able to make weak predictions within 1 standard deviation. This suggests a degree of predictability between movie features and rating trends such as variability and general perception change.

> In particular, we found that predictability increased slightly when we increased our time horizon, suggesting that such a prediction model is more accurate in determining long-term trends.

Appendix

Considered Research Topic: Netflix/Oscar Effect

- We also attempted to investigate the phenomenon known as the "Intervention Effect", where viewership of movies drastically changes when an event brings certain movies to the spotlight, such as a movie being released on a streaming platform or a particular actor being nominated for an award. For example, this is observed when a movie from long ago is reintroduced on Netflix, resulting in a huge spike in viewers and reviews for that movie overall.
- Although this idea initially sounded promising, after working with the provided datasets and supplemental Netflix datasets, we realized that we did not have sufficient data to work with as the intersection between the Netflix dataset and our overall dataset was very small (<1000 movies), especially after considering normalization factors arising from the timeframe of our ratings dataset. A similar issue arose when considering the Oscar dataset.
- Furthermore, for the "Oscar Effect", we felt that the one-year timeframe between a movie's release and an Oscar nomination event further limits the ratings data for a pre-post comparison. This initial observation then motivated us to consider a longer timeframe, hence leading us to our main research topic.