

*PROD.*

# Predicting IMDB Scores from the 2010s

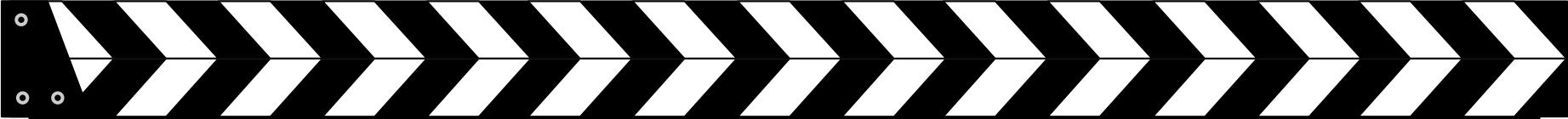
*SCENE:* Metis Project: Regression

*DIRECTOR:* Maxwell Wood

# Project Breakdown

- Client: Netflix/Amazon/Disney+ or other Streaming Service
- Goal: Predict cumulative IMDB user scores for a movie after its finished its run in theaters.





# Data Collected

- All wide released films from Jan 2010 – Dec 2019 (1,673 Movies)
- From Box Office Mojo's Website: **Box Office Mojo**
  - Domestic Gross
  - Total Gross
  - Budget
  - Domestic Distributor
  - Opening
  - Release Date
  - MPAA Rating
  - Genres
  - Running Time
  - Widest Release
  - Opening Money
  - Opening Rank
- From Metacritic's Website: **metacritic**
  - Average Critical Review
- From IMDb's Website: **IMDb**
  - Average review by demographic
  - Number of reviews by demographic

# Meta Critic Data Details

- Metacritic aggregates and averages professional critic reviews, which are written a week before release of the film
- For this project I just take the averaged score as a feature



**100** **The Playlist**  
Todd Gilchrist  
A cinematic, cultural and personal triumph, *The Dark Knight Rises* is emotionally inspiring, aesthetically significant and critically important for America itself – as a mirror of both sober reflection and resilient hope.  
[Read full review ↗](#)

**70** **THE NEW YORKER**  
Anthony Lane  
Christopher Nolan, for all his visionary flair, wants to suck the comic out of comic books; Anne Hathaway wants to put it back in. Take your pick.  
[Read full review ↗](#)

**85** **npr**  
Bob Mondello  
As you might expect from the creator of "Inception" and "Memento," there are surprises both in the story and in the storytelling. But the biggest surprise may just be how satisfying Nolan has made his farewell to a Dark Knight trilogy that many fans will wish he'd extend to a 10-part series, at least. [Read full review ↗](#)

**30** **READER**  
J.R. Jones  
The script, by Nolan and his brother Jonathan, takes a few vague pokes at Wall Street and the financial elite but mainly revives the ponderous psychodrama of the first movie. [Read full review ↗](#)

**58** **inChristianSCIENCE MONITOR**  
Peter Rainer  
Were it not for Anne Hathaway's Catwoman-ish Selina Kyle, there wouldn't be a single character in "Rises" who cracks a smile. I'm not arguing that "Rises" should be "Singin' in the Rain." But its Wagnerian ambitions are not matched by its material. It hasn't earned its darkness. [Read full review ↗](#)

**75** **Movieline**  
Michelle Orange  
The Dark Knight aspires to the epic and reaches it on a number of impressive and less impressive levels. That it is a frequently, unnervingly glorious triumph of brawn over brains is not despite but in spite of Nolan's admirably stubborn - if persistently, risibly serious - insistence that the modern superhero can have it all. [Read full review ↗](#)

# IMDb User Score Details

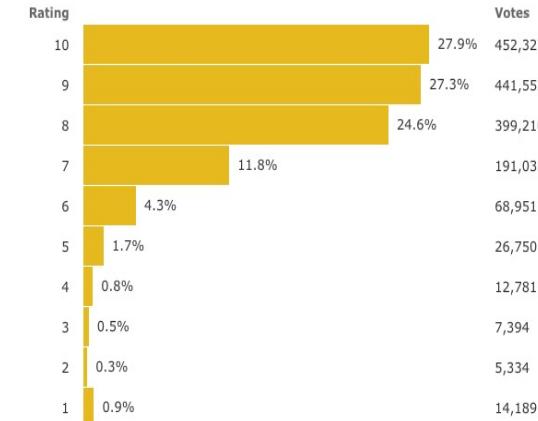
- IMDb Allows Users to rate movies themselves and then provides this data on their website
- For the target I sum all user ratings for a movie
- I also collect data by demographic to explore differences



Rating By Demographic					
	All Ages	<18	18-29	30-44	45+
All	8.4 1,619,524	8.2 766	8.6 298,433	8.4 645,199	8.0 118,450
Males	8.4 983,682	8.3 604	8.6 244,619	8.4 544,483	8.0 100,685
Females	8.3 167,528	7.6 77	8.5 46,387	8.3 90,662	8.0 15,340
Top 1000 Voters		US Users		Non-US Users	
7.9 866		8.3 190,913		8.3 582,240	

## IMDb Users

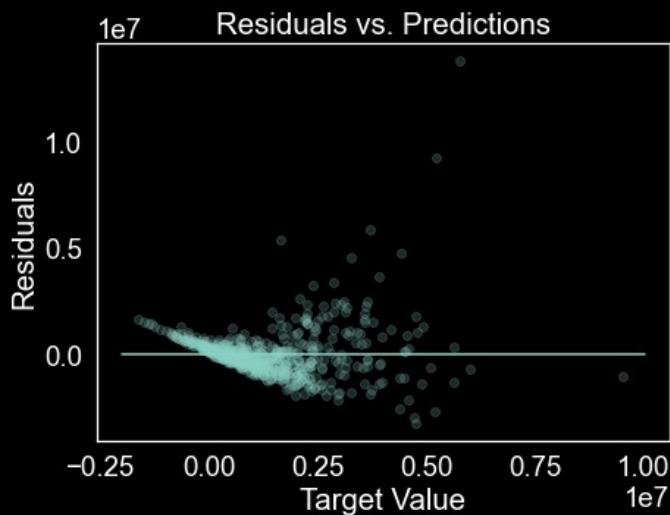
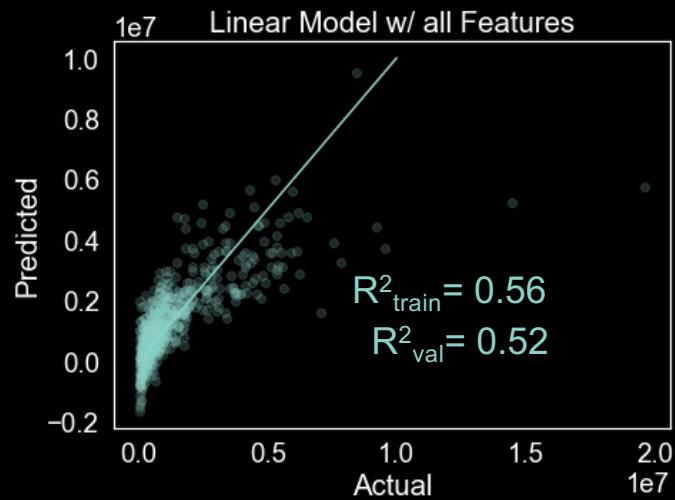
1,619,524 IMDb users have given a weighted average vote of 8.4 / 10



Arithmetic mean = 8.4 Median = 9

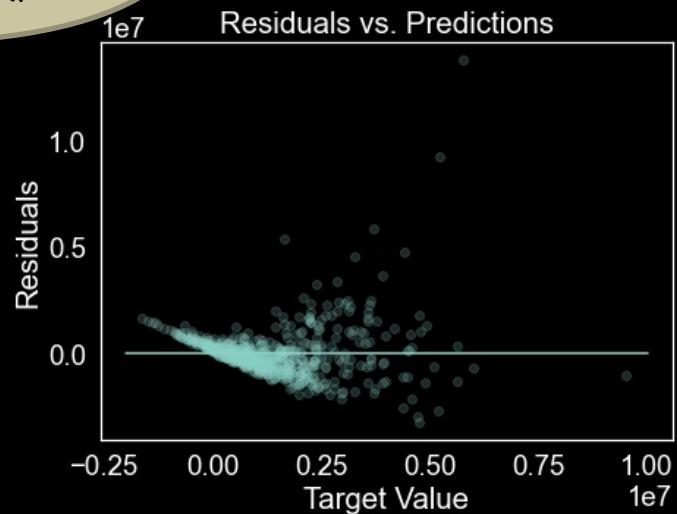
# First Attempt!

- Data Cleaned of Nan's and Duplicates
- Data Split: 60 Train: 20 Validation: 20 Test
- Categorical Features Turned Into Dummies
- Continuous Variables Normalized



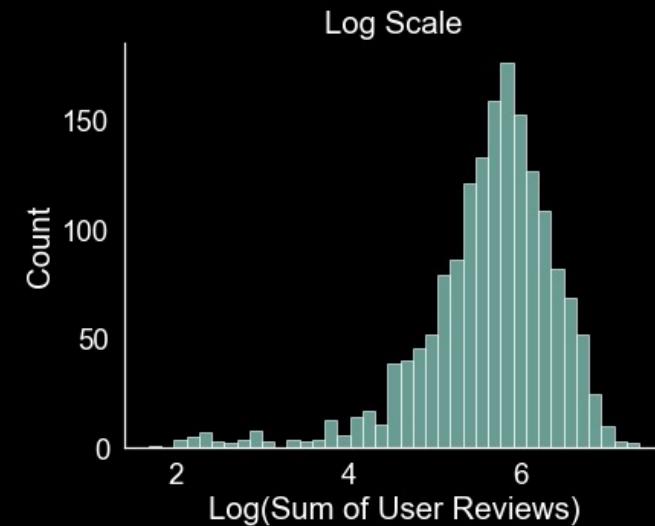
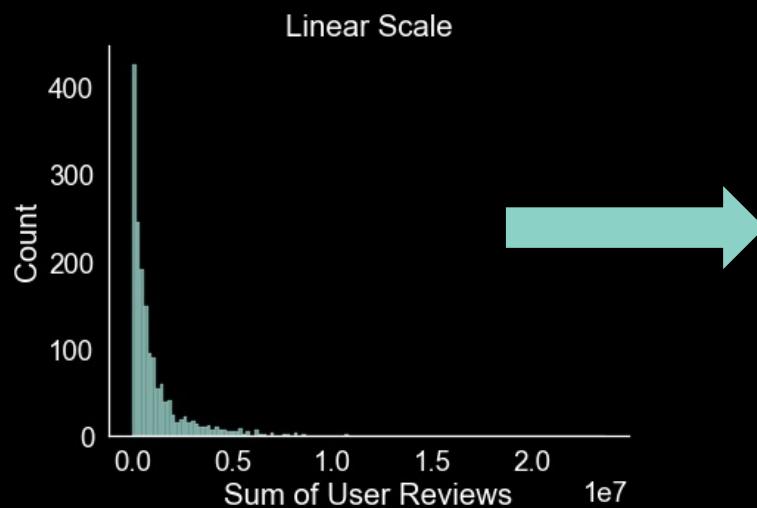
# First Attempt!

HOLY-  
HETROSKEDEASTICITY  
BATMAN!

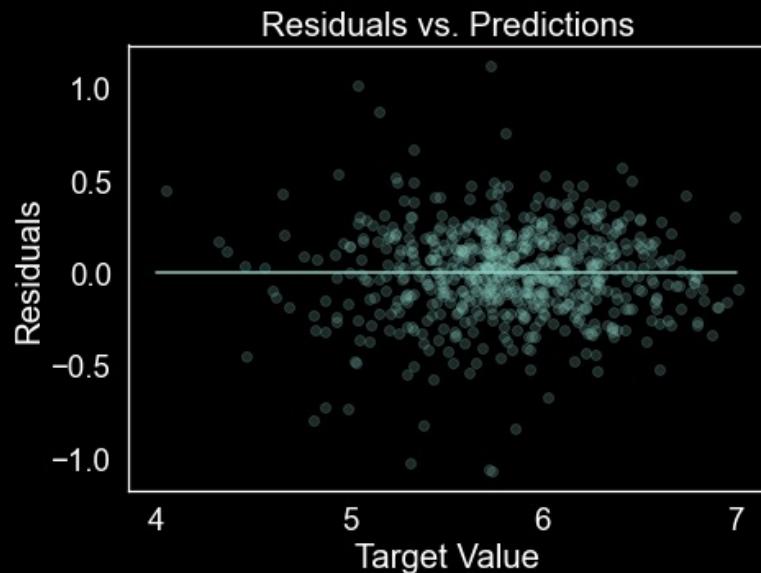
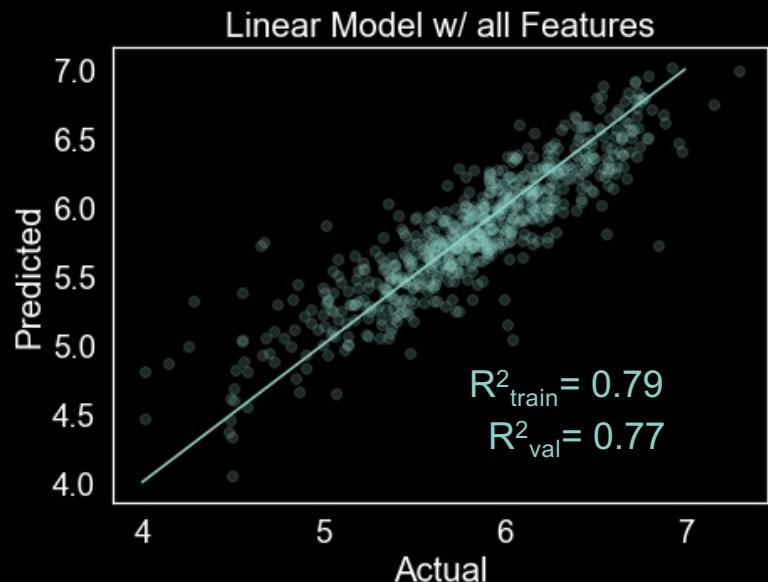


# Log Transform Data

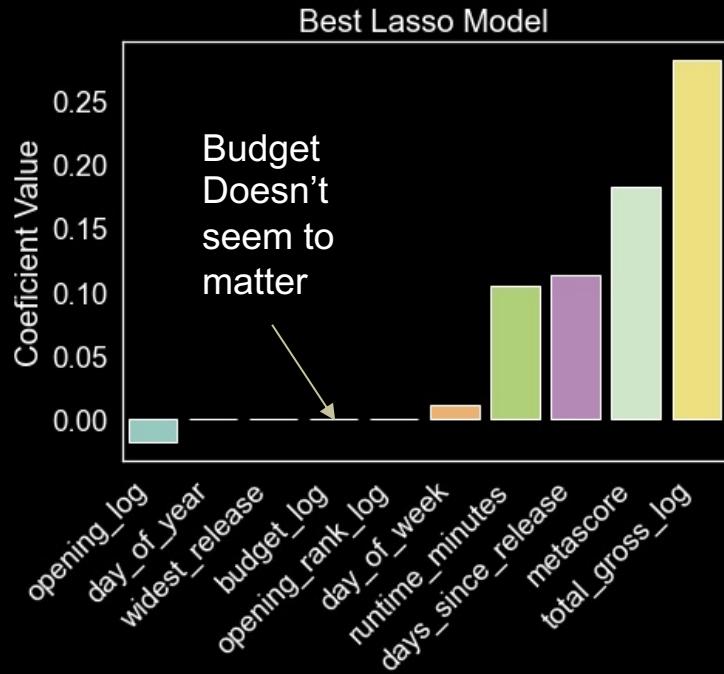
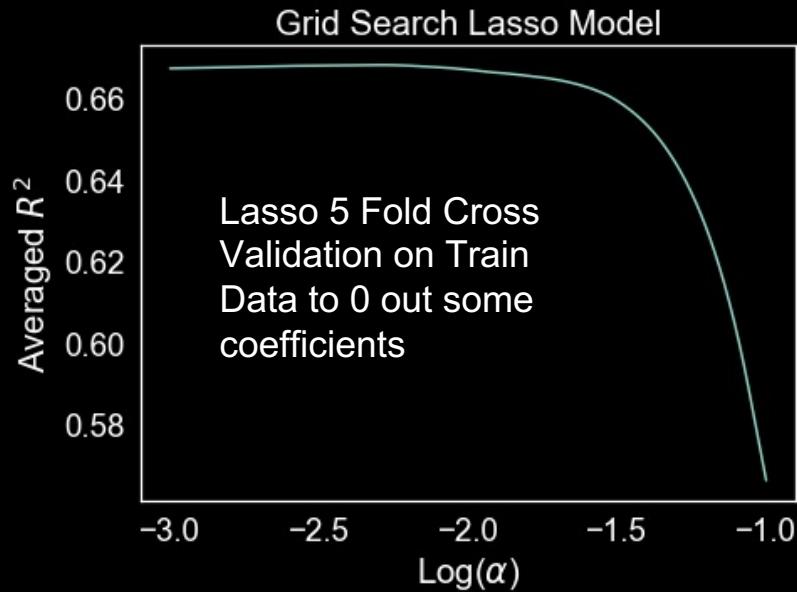
- Problem: Lots of features have log-normal distributions including the target
- Solution: Take the Log of these Features and Target



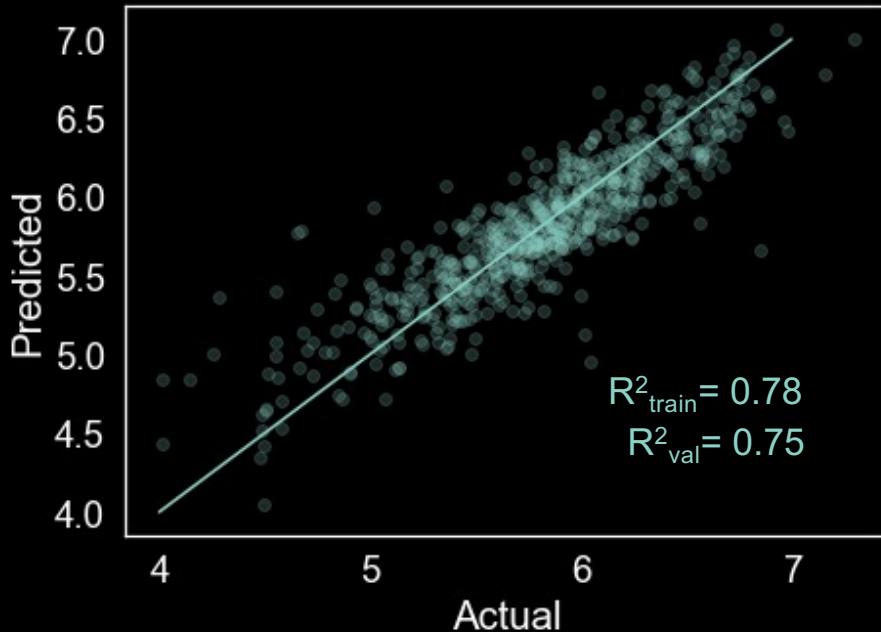
# Retry Linear Regression



# Cut Down Continuous Features

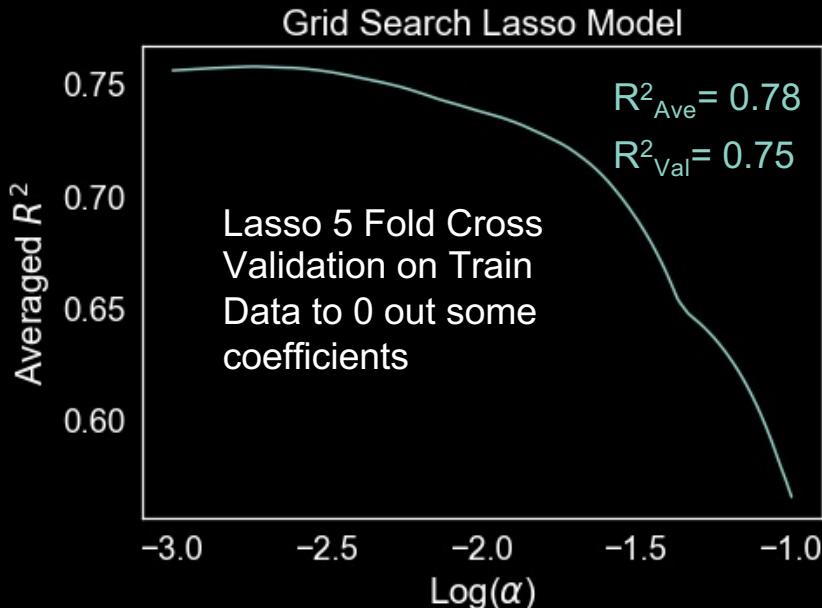


# Add in Genre and MPAA Dummies



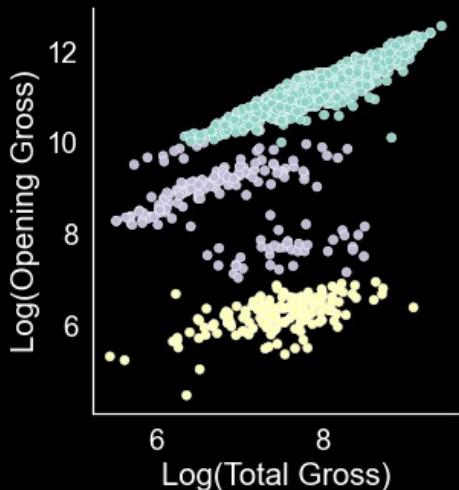
Features	Coefficient		
6	Action	0.030978	
22	War	0.016109	
10	Romance	0.013565	
7	Thriller	0.012317	
25	Film-Noir	-0.000000	
4	Drama	-0.008523	
8	Adventure	-0.009756	
21	Sport	-0.010432	
16	Animation	-0.016617	
15	Horror	-0.033165	
18	History	-0.044411	
19	Musical	-0.047431	
20	Music	-0.136799	
11	Family	-0.215010	
23	Documentary	-0.370852	
2	total_gross_log	0.301921	
28	R	0.260805	
1	metascore	0.201955	
27	PG-13	0.175319	
17	Biography	0.132660	
12	Sci-Fi	0.128699	
3	days_since_release	0.122348	
24	Western	0.120659	
9	Fantasy	0.076502	
14	Mystery	0.069215	
13	Crime	0.055730	
0	runtime_minutes	0.048300	
5	Comedy	0.037260	
26	G	0.035572	

# Lasso Again!



	Features	Coefficient
3	total_gross_log	0.311932
16	R	0.253902
2	metascore	0.192855
15	PG-13	0.165994
9	Sci-Fi	0.129574
4	days_since_release	0.124166
12	Biography	0.104687
7	Fantasy	0.070400
11	Mystery	0.068715
10	Crime	0.059221
1	runtime_minutes	0.049545
5	Comedy	0.041444
6	Action	0.037702
0	opening_log	-0.022887
13	Music	-0.147005
8	Family	-0.235426
14	Documentary	-0.364746

# Why is Opening Negative?



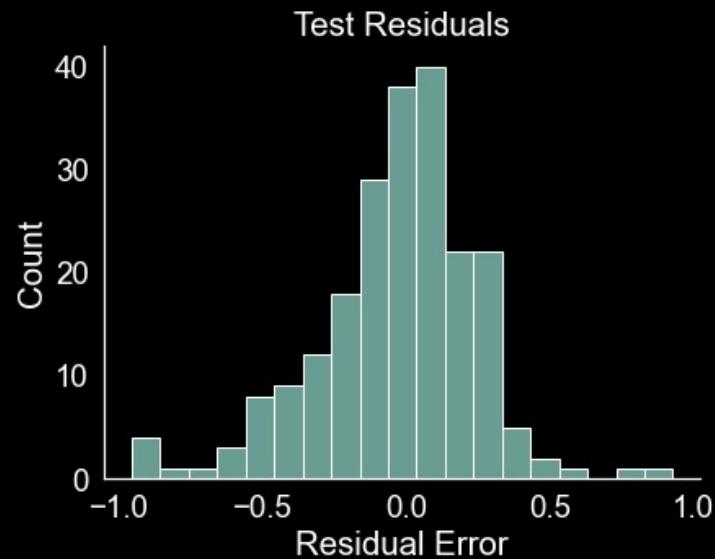
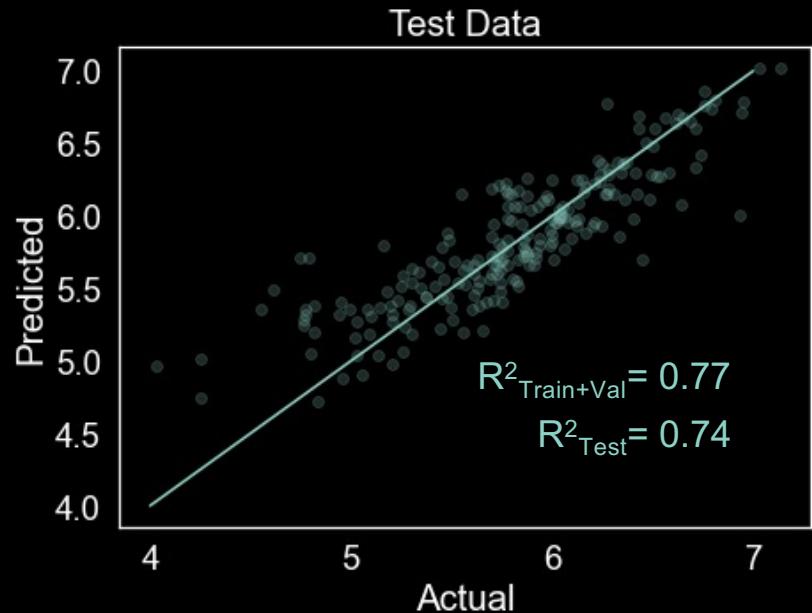
Features	Coefficient
3 total_gross_log	0.311932
0 opening_log	-0.022887

Studios like to open their Oscar worthy films in “Select Theaters” before they’re released to general public. These films often have

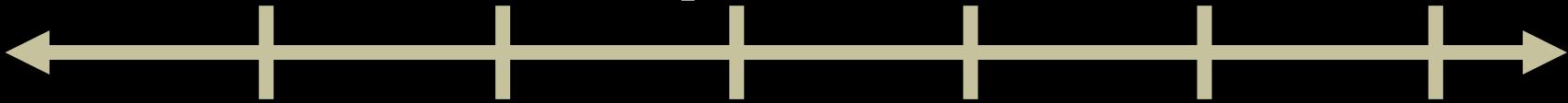


Date ^	DOW	Rank ♦	Daily	%± YD	%± LW	Theaters ♦	Avg ♦
Dec 9	Friday	15	\$299,307	-	-	5	\$59,861
Dec 10	Saturday	15	\$298,688	-0.2%	-	5	\$59,737
Dec 11	Sunday	14	\$283,109	-5.2%	-	5	\$56,621

# Final Scoring!



# Random Sample of Predictions



Actual Score

5.26

5.74

6.03

6.15

6.27

6.49

Predicted Score

5.23

5.48

5.98

6.25

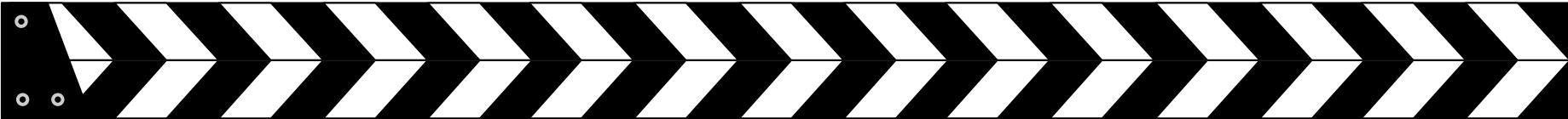
6.78

6.28

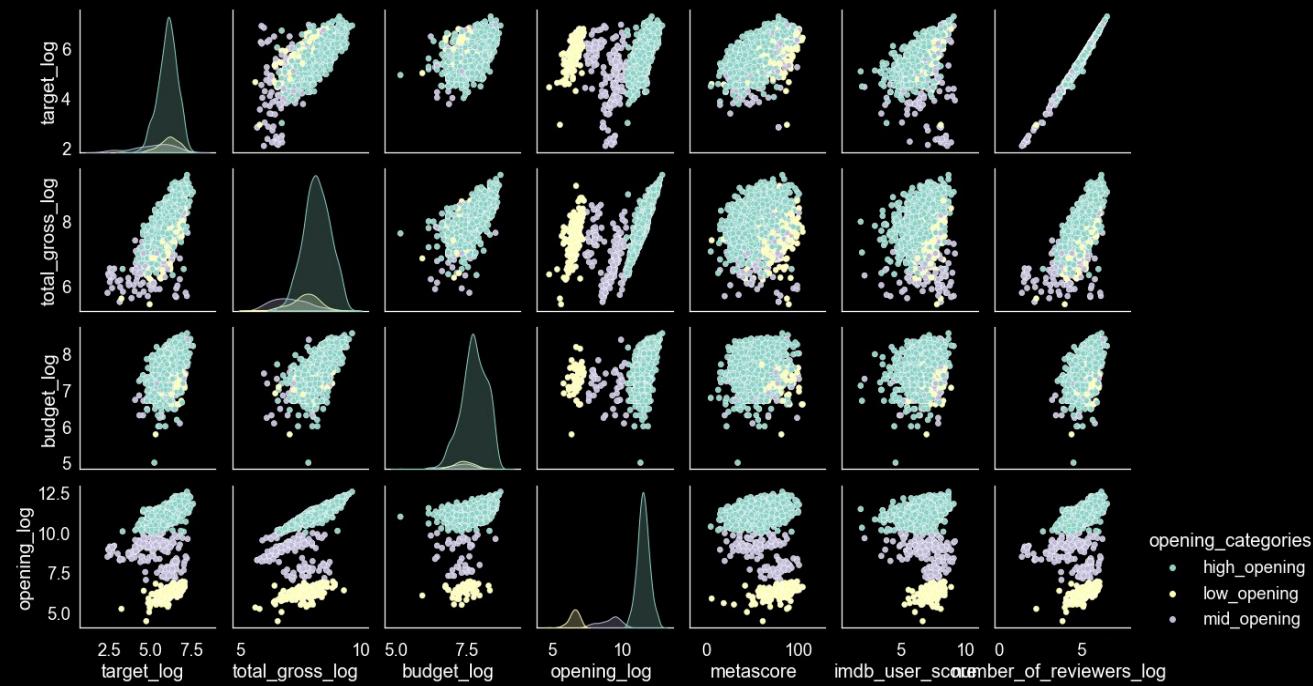


Questions?





# Appendix



# Appendix

