Lights, Camera, Analytics!

10|03|16

Elizabeth Carnahan
Kristopher Korte
Brian Malloy
Noreen Nagji
Karthik Subramanian

Agenda

1. Strategic Challenges
2. Metrics and Mental Models
3. Data and Analysis
4. Insights
5. Recommendation and Next Steps

Strategic Challenges

What factors are relevant in predicting box office revenue of a movie?

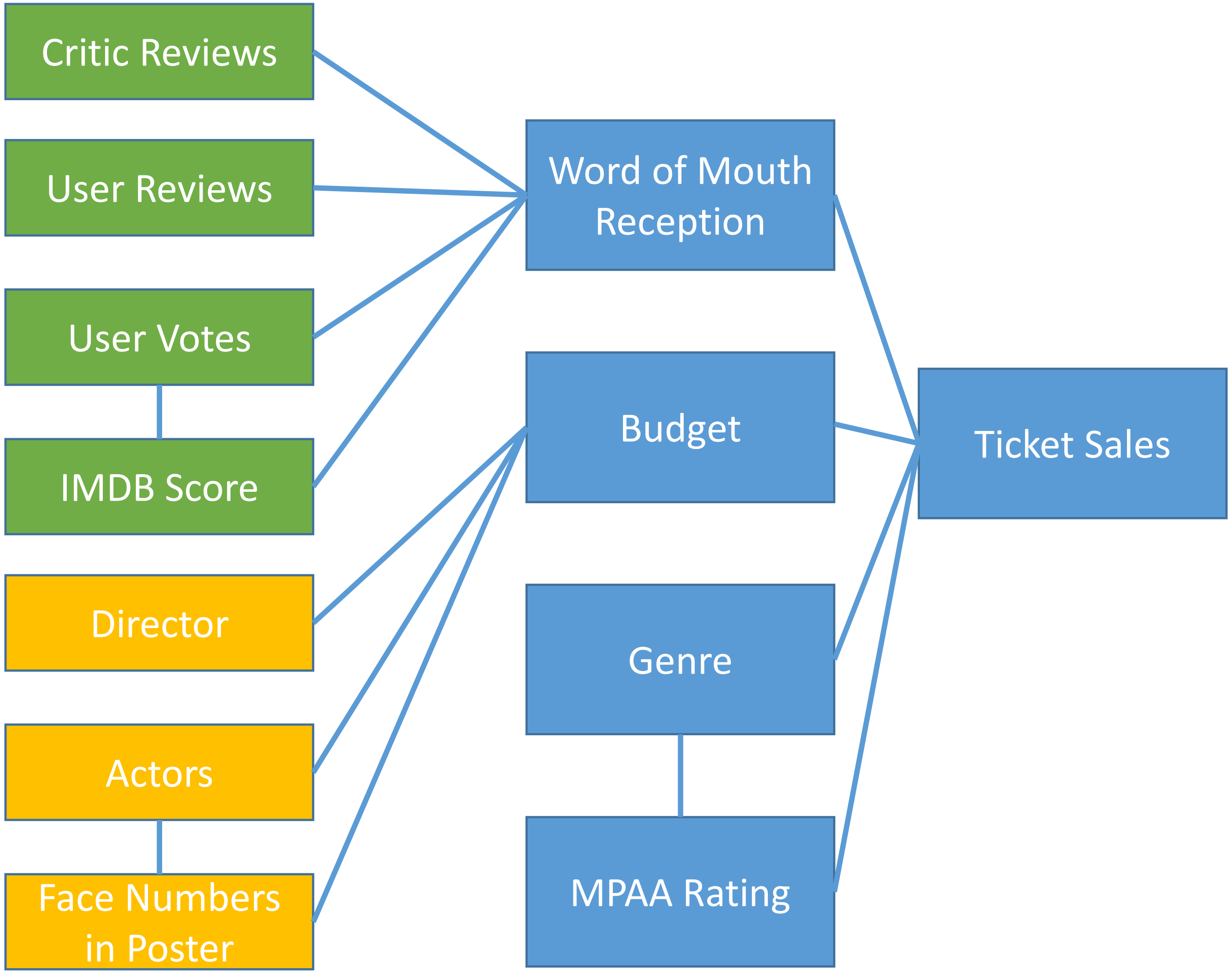
Our hypothesis is that movie production companies can predict revenues through careful selection of specific variables

1	9	3	5	11	2
$2,783,918,982	**$1,274,234,980**	**$1,837,932,841**	**$1,519,479,547**	**$1,163,595,387**	**$2,207,615,668**

Metrics & Mental Models

Mental Model

Prior to running our regressions, we created a mental model of the variables that could significantly affect ticket sales



Data and Analysis

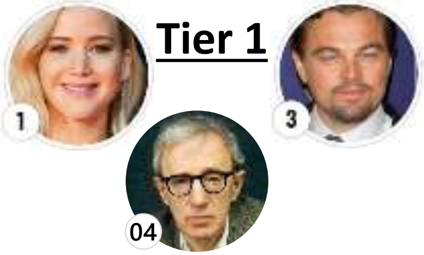
Data

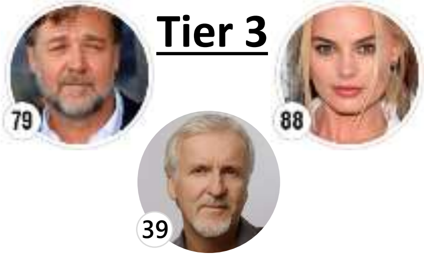
We used complimentary datasets to create dummy variables and adjusted dollar amounts for an accurate comparison

kaggle → IMDb → **5,044 Rows** → **2,695 Rows**

- US Films
- English as Primary Language
- Released in 1990 or Later
- Eliminated duplicates
- Eliminated incomplete rows
- Eliminated films rated X & NC-17

- Created 3 Tiers Based on Vulture Ranks (Actor) & AMC Ratings (Director)





- Adjusted gross revenue for inflation using average ticket price per year of movie release

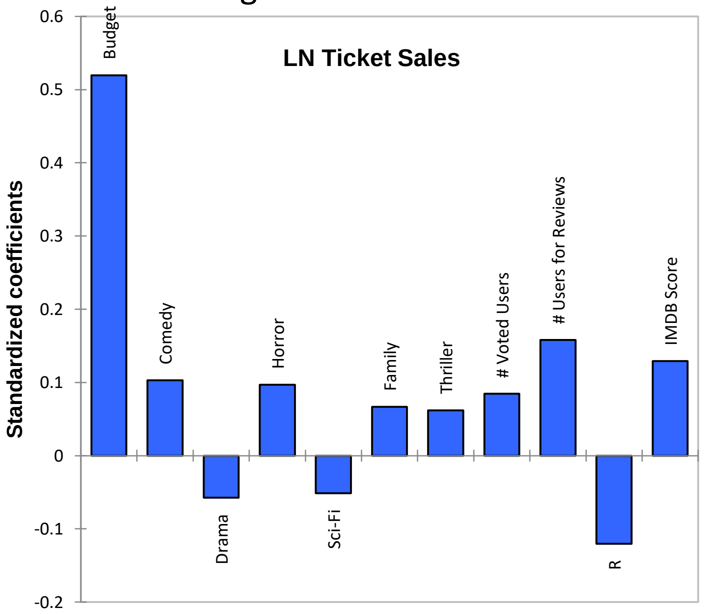
- Ratings
  Ex. G, PG, PG-13, R
- Film Genres
  Ex. Comedy, Drama, Horror
- Actor/Director Tiers
  Ex. Tier 1, Tier 2, Tier 3

Analysis – Linear Regression

We used regression analysis to determine what factors contribute to a movie's box office revenues

Regression → **Model Prediction** → **Goodness of Fit Statistics**

- Test set (20% of data)
- Training set (80% of data)
- Ran a linear regression on log of ticket sales, second time eliminating all insignificant variables

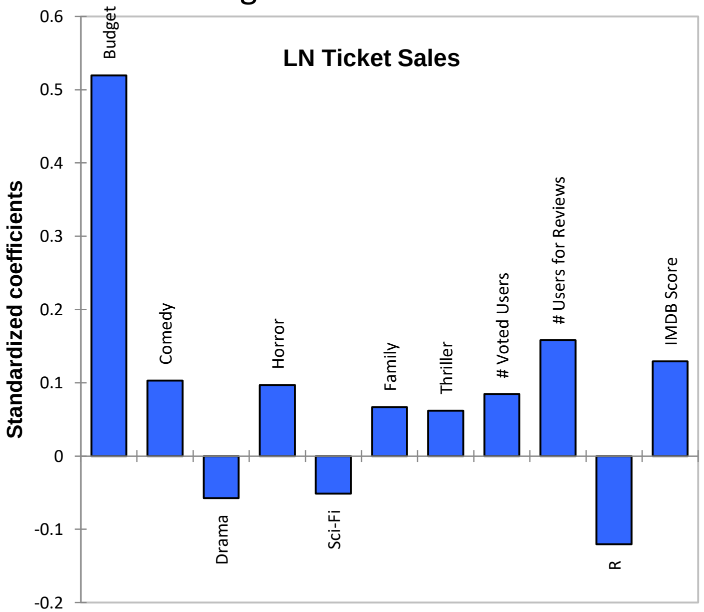


Analysis – Linear Regression

We used regression analysis to determine what factors contribute to a movie's box office revenues

Regression | **Model Prediction** | **Goodness of Fit Statistics**

- Test set (20% of data)
- Training set (80% of data)
- Ran a linear regression on log of ticket sales, second time eliminating all insignificant variables



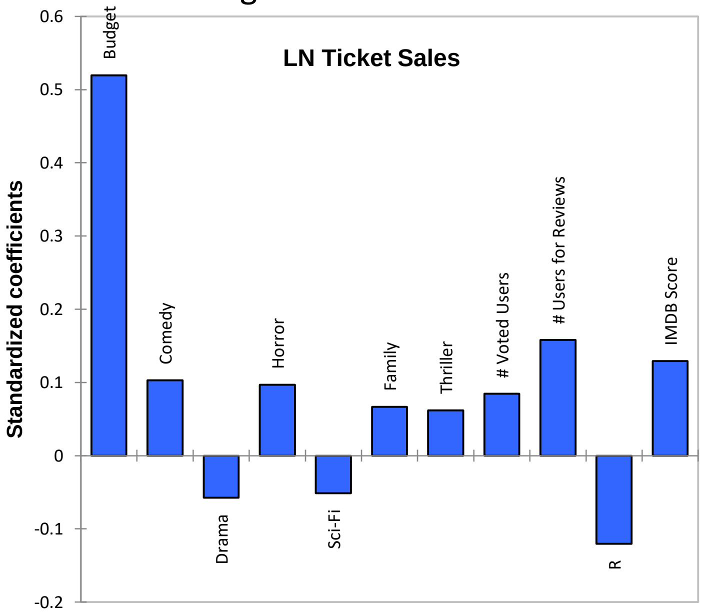
- Used analysis to perform prediction for test data set
- Goal: to predict ticket sales and box office revenues with accuracy

Analysis – Linear Regression

We used regression analysis to determine what factors contribute to a movie's box office revenues

Regression

- Test set (20% of data)
- Training set (80% of data)
- Ran a linear regression on log of ticket sales, second time eliminating all insignificant variables



Model Prediction

- Used analysis to perform prediction for test data set
- Goal: to predict ticket sales and box office revenues with accuracy

Goodness of Fit Statistics

Observations	2145.000
Sum of weights	2145.000
DF	2133.000
R^2	0.509
Adjusted R^2	0.51
MSE	1.891
RMSE	1.375
MAPE	7.24
DW	0.842
Cp	12.000
AIC	1378.900
SBC	1446.950
PC	0.496

Analysis – Logit Regression

We used regression analysis to determine what factors contribute to a movie's box office revenues.

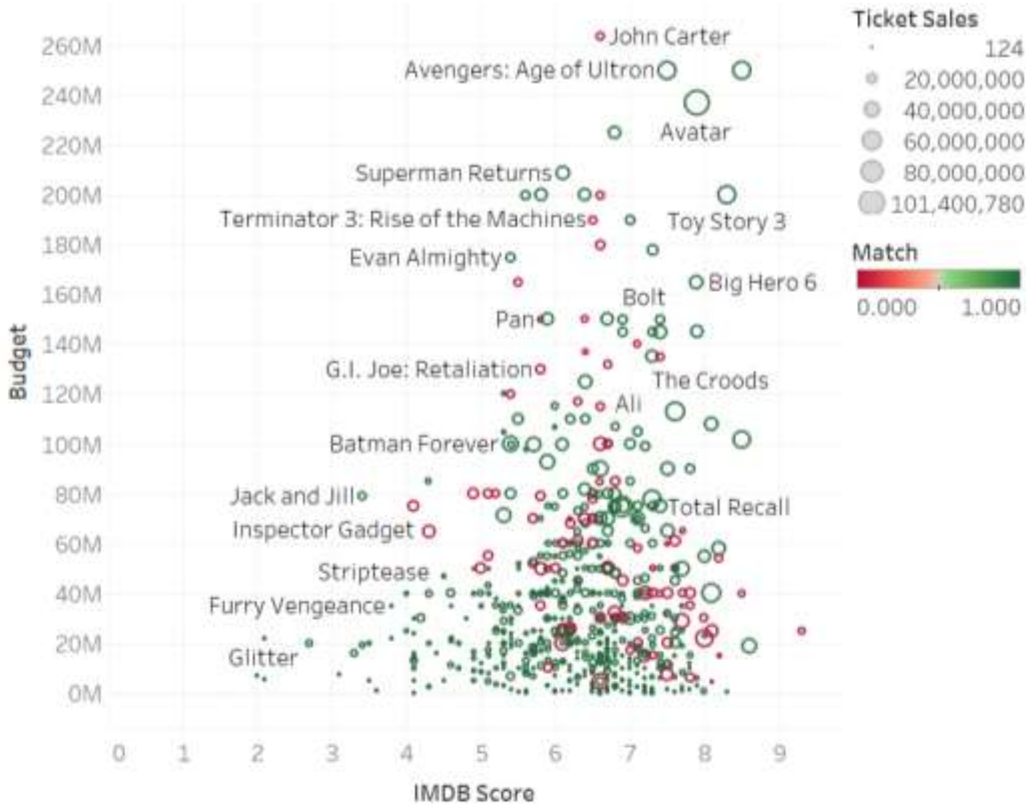
Test set (20% of data)
Training set (80% of data)

Logit regression on the success of each movie

Success = blockbuster movie with $100+ M in adjusted gross revenue

Goal: to predict which movies would be a blockbuster hit

Hit Rate = 84%



Sum of IMDB Score vs. sum of Budget. Color shows sum of Match. Size shows sum of Ticket Sales. The marks are labeled by Movie Title. Details are shown for Movie Title.

Insights


MOVIE NAME
CINEMA
TICKET
ADMIT ONE
B 0026438
B 0026438
TICKET
ADMIT ONE
B 0026438

Sony Picture Studios: 4th Quarter 2015 Releases

In looking at a sample of recent releases, we see how effective the model is in predicting whether or not a movie will be a blockbuster

Movie	Budget	Ln of Budget	# Critics for Reviews	Comedy	Drama	Sci-Fi	Family	# Voted Users	# Users for Review	IMDB Score	Blockbuster Prediction	Match?
HOTEL TRANSYLVANIA 2	$80M	$18.20	156	1	0	0	1	59,884	100	6.7	YES	✓
FREAKS OF NATURE	$33M	$17.31	36	1	0	1	0	6,063	30	5.9	NO	✓
SPECTRE 007	$245M	$19.32	604	0	1	0	0	283,170	1,003	6.8	YES	✓
The Night Before	$25M	$17.03	154	1	0	0	0	34,274	87	6.5	NO	✓
BASED ON A TRUE STORY CONCUSSION	$35M	$17.37	222	0	1	0	0	48,889	143	7.1	NO	✓

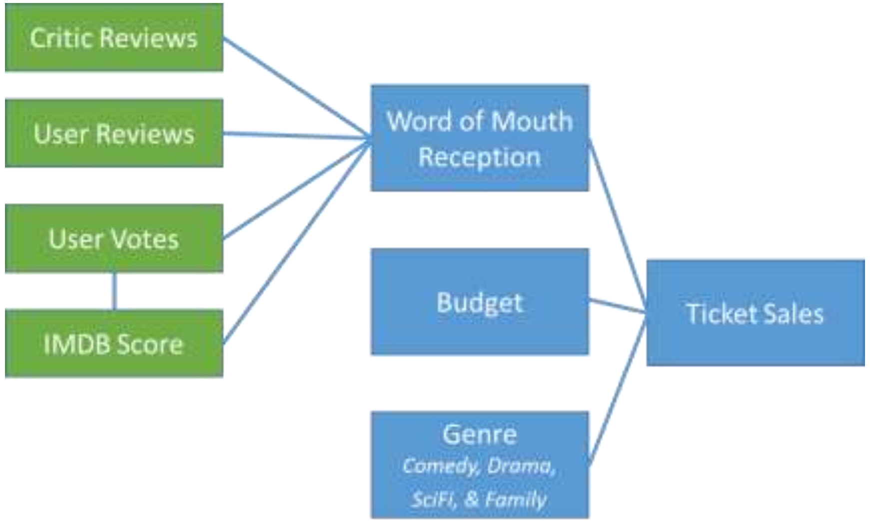
Factors that Govern Ticket Sales / Box Office Revenues

After identifying the variables that significantly effected the success of a movie, we adjusted our Mental Model accordingly

Significant Variables

- ✓ Movie Budget
- ✓ IMDB Score
- ✓ Movie Genre: *Comedy, Drama, Sci-Fi, and Family*
- ✓ User Engagement on IMDB: *Number of User Votes and Number of User Reviews*

New Mental Model



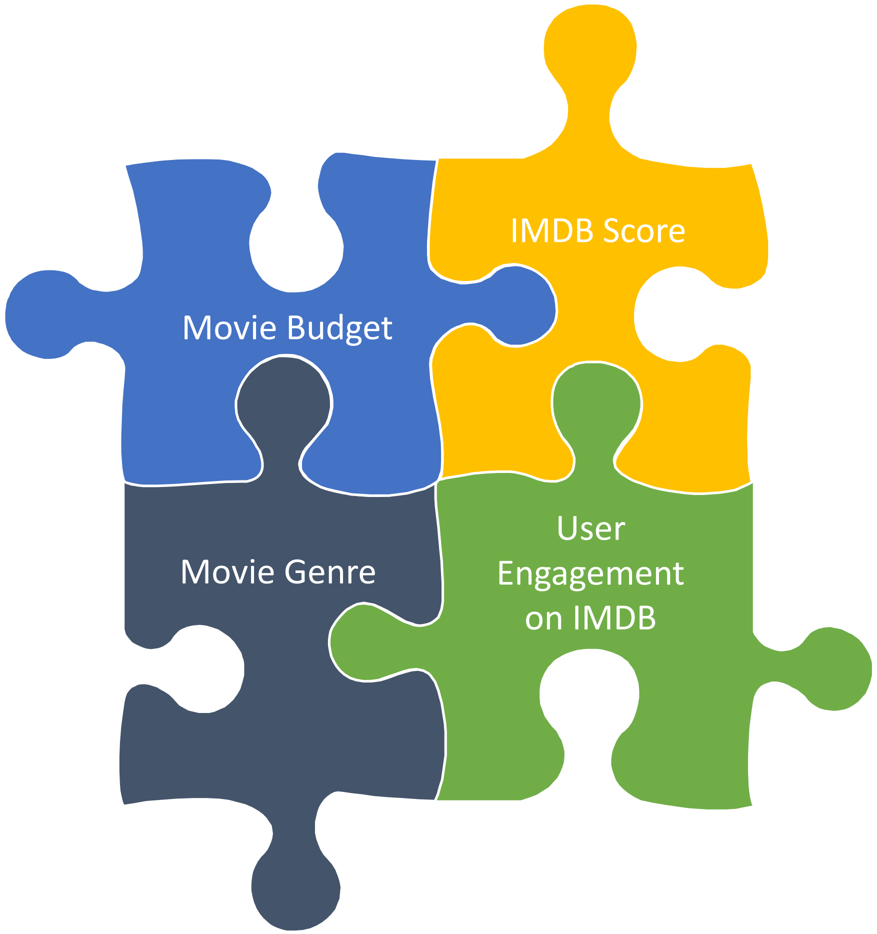
Insignificant Variables

- X Popularity of Cast
- X Popularity of Director
- X Number of Critic Reviews on IMDB
- X Movie Genres (excluding the 4 listed above)
- X Number of Faces in Poster
- X Movie Ratings

Recommendations and Next Steps

Recommendations and Next Steps

To further investigate the actionable variables that our regression highlighted, we have outlined additional strategic challenges



IMDB Score

- What are the components of this rating?
- Are there any early indicators of the IMDB score?
- Can we test these indicators prior to the release of the movie?
- Is there a way to incentivize additional users and critics to review a movie?


Questions?
fresh
POP CORN