

IMDb Movies Data Analysis



▶
Presented By:
Team 9

01

ABOUT THE DATA

04

**VARIABLE
TRANSFORMATION**

02

CONTENT ANALYSIS

05

LIMITATIONS

03

PRODUCTION ANALYSIS

06

SUMMARY

FEAR CAN HOLD YOU PRISONER.
HOPE CAN SET YOU FREE.



TIM ROBBINS MORGAN FREEMAN

THE SHAWSHANK REDEMPTION

CASTLE ROCK ENTERTAINMENT PRESENTS
A FRANK DARABONT FILM "THE SHAWSHANK REDEMPTION" BOB GUNTON WILLIAM SADLER
CELANCY BROWN GIL BELLOWS AND JAMES WINTMORE AS "BROOKS" "THE THOMAS NEWMAN" "RICHARD FRANCIS-BRUCE"
TERENCE MARSH DIRECTOR OF PHOTOGRAPHY ROGER DEAKINS EDITOR LIZ GLOTZER EXECUTIVE PRODUCERS STEPHEN KING
PRODUCED BY FRANK DARABONT AND NIKI MARVIN WRITTEN BY FRANK DARABONT
CASTLE ROCK ENTERTAINMENT COLUMBIA PICTURES
THIS FALL



ABOUT THE DATASET

IMDB movies

What factors contribute to higher rated movies?

- It contains 85855 movies and 22 variables.
- Title, Genre, Country, Average Rating, Actor, Budget
- Ranging from 1894 to 2020.
- We filter for USA movies to do analysis on actors and budget.

02

► **Content Analysis**



AVERAGE RATING

GENRE



**MOVIE
DESCRIPTION**

▲ **WHAT WE ARE WORKING ON**

Genres

How do genres affect ratings?

24

Genres

28511

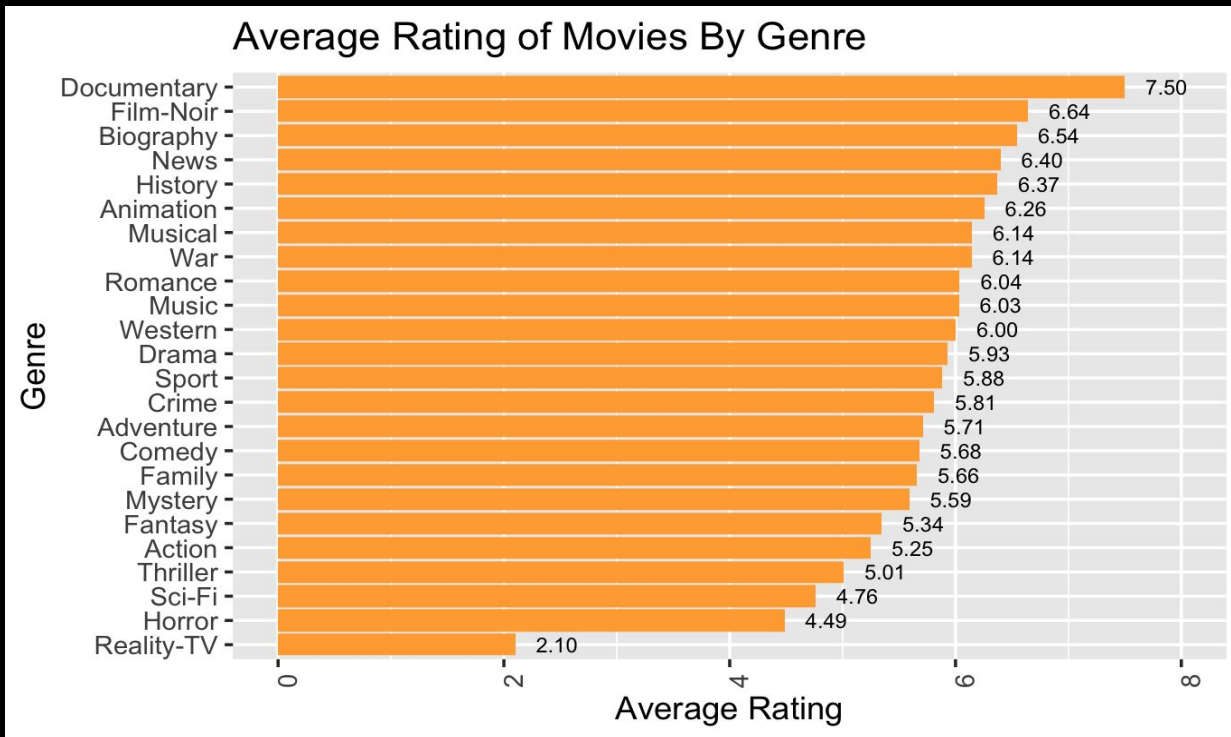
USA movies

277

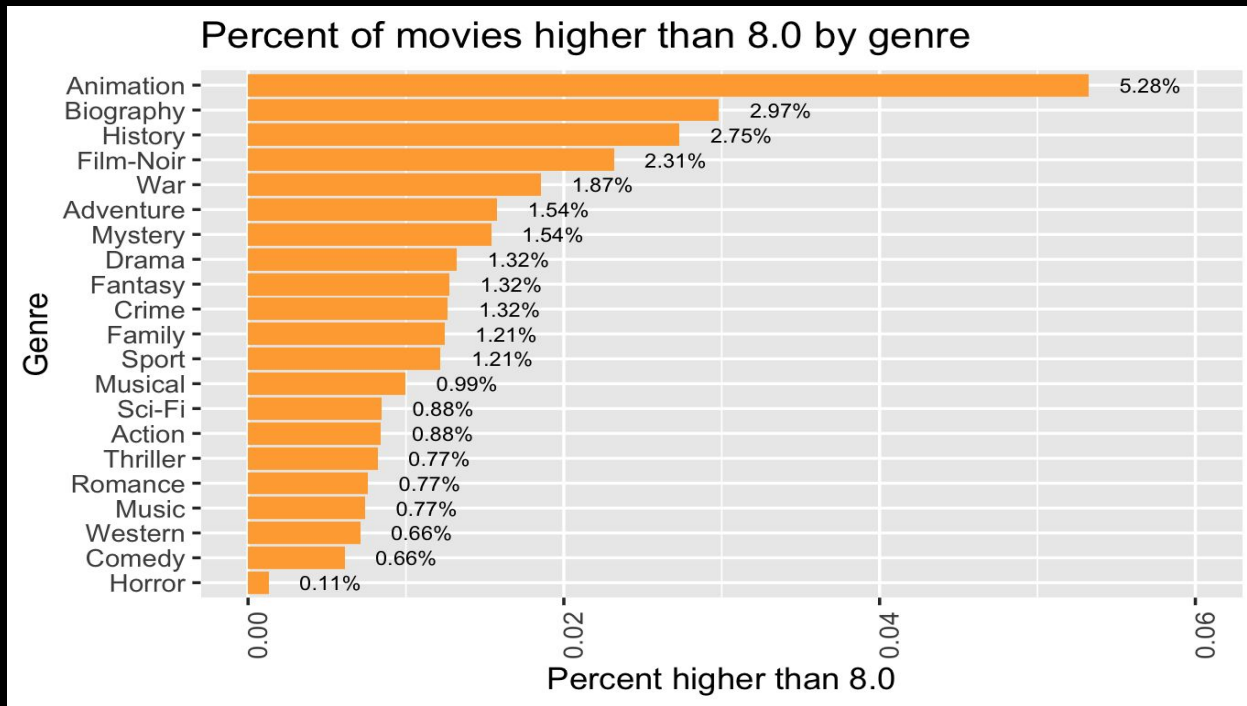
USA movies rated
higher than 8.0



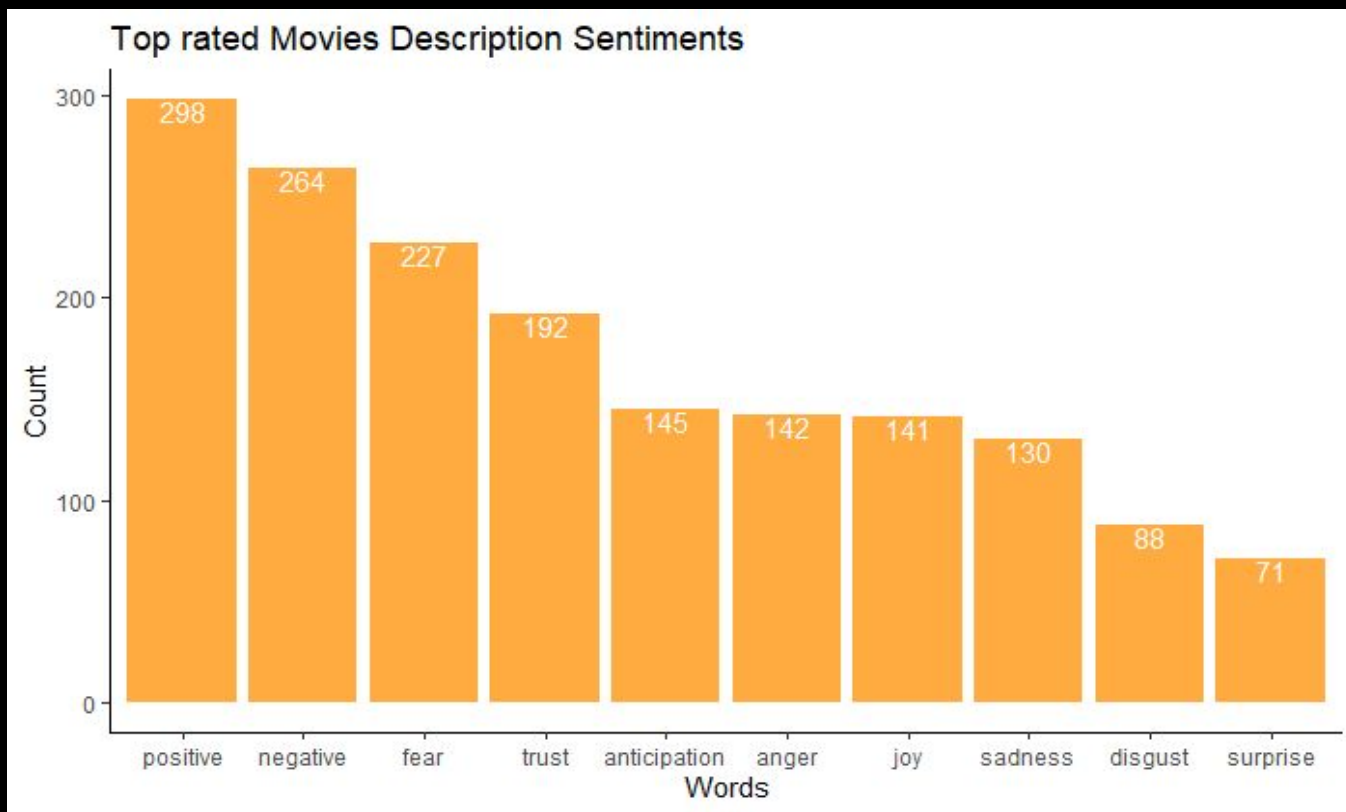
Average Rating of all Movies by Genre



Percentage of high-rated movies by genre

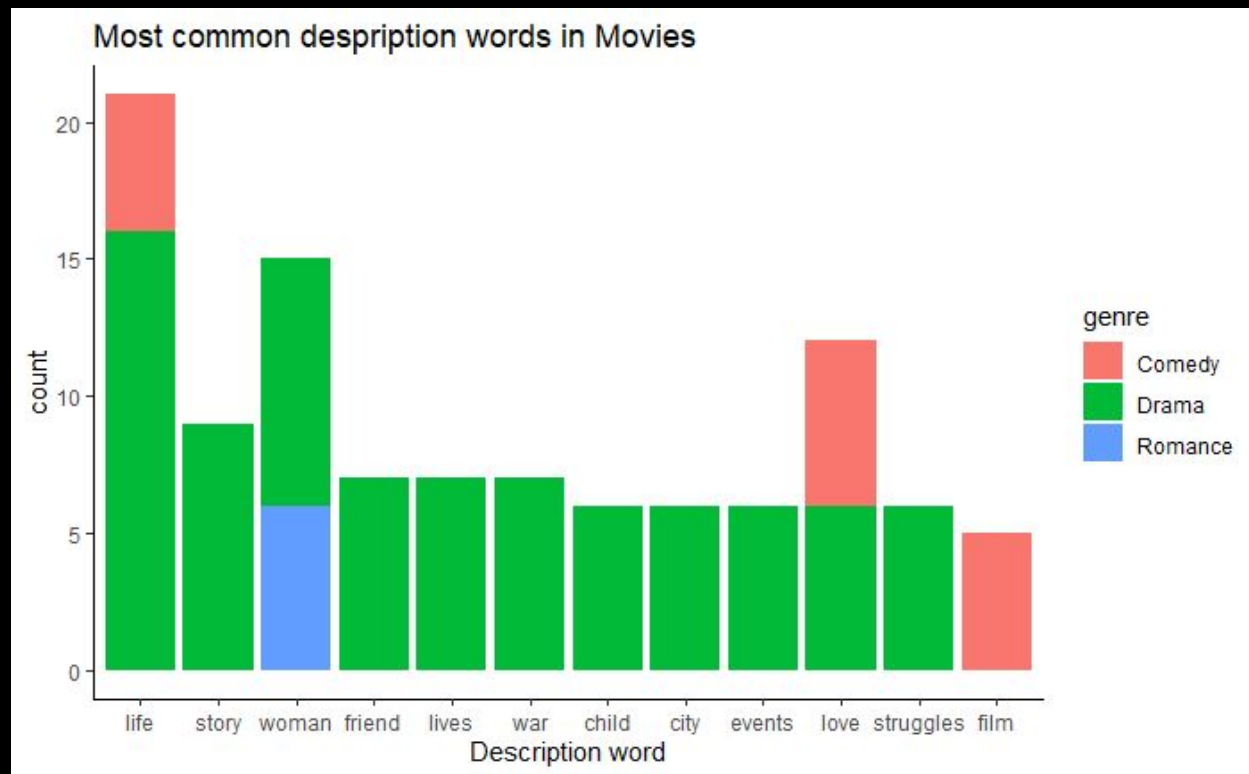


Sentiment Analysis on Top-Rated Movie Description



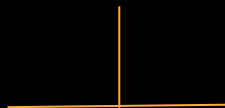
▲ Sentiment Analysis

Most Common Description Words By Genre in Top Rated Movies





Drama



LIFE



LIVES



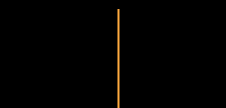
STORY



WOMAN



Comedy



LOVE



GIRL



LIFE



TRAMP



Romance



WOMAN



LOVE



WIFE



AMERICAN



COMMON WORDS IN GENRE



03

► **PRODUCTION
ANALYSIS**



AVERAGE RATING

**PRODUCTION
BUDGET**



ACTORS

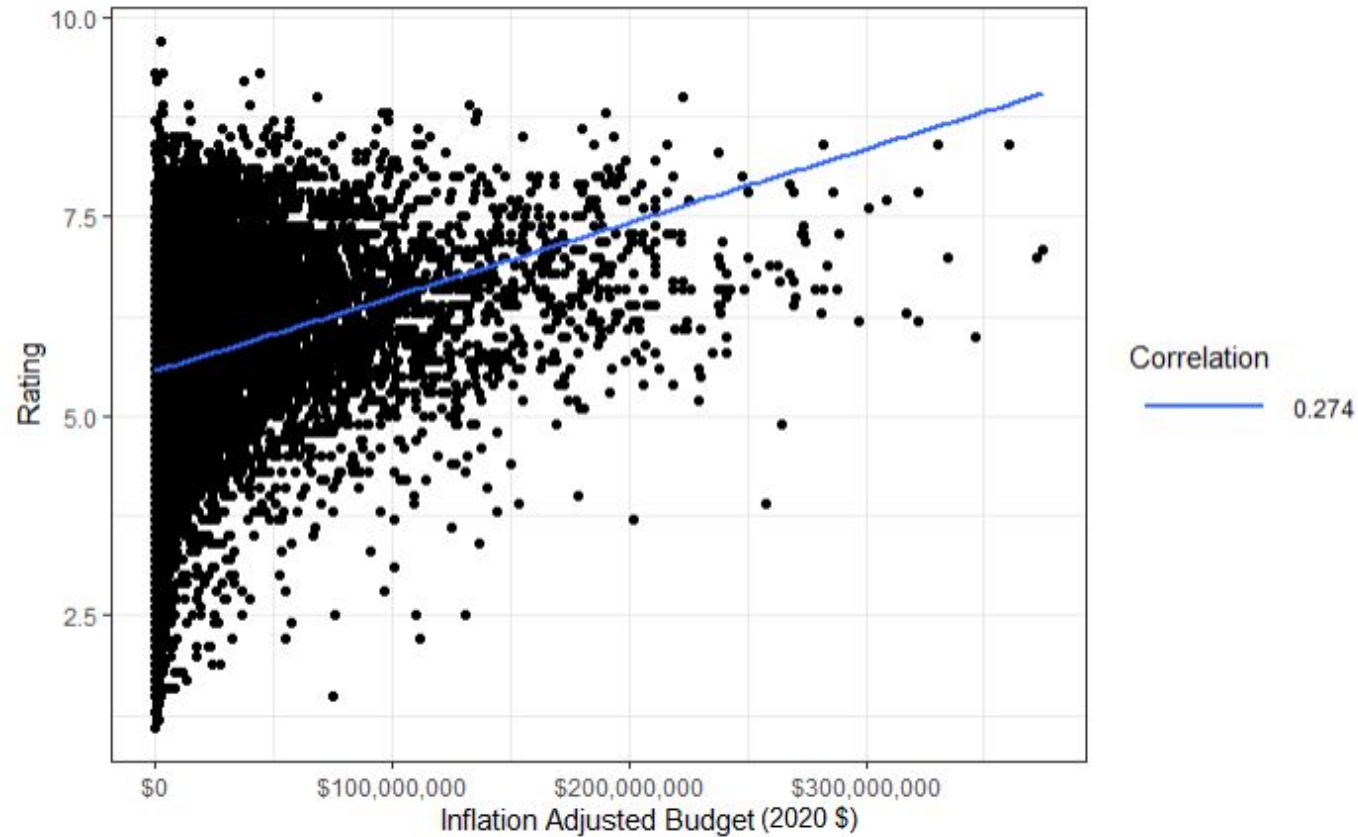
▲ **WHAT WE ARE WORKING ON**

Top 10 Inflation Adjusted Production Budgets

Movie	Year	Adj. Budget (2020 \$)	Rating
Pirates of the Caribbean: At World's End	2007	\$374M	7.1
Cleopatra	1963	\$372M	7.0
Avengers: Endgame	2019	\$360M	8.4
Superman Returns	2006	\$347M	6.0
Star Wars: Episode VIII - The Last Jedi	2017	\$335M	7.0
Avengers: Infinity War	2018	\$331M	8.4
Titanic	1997	\$323M	7.8
Spider-Man 3	2007	\$322M	6.2
Justice League	2017	\$317M	6.3
Tangled	2010	\$309M	7.7

▲ Budget Analysis

Relationship Between Budget and Movie Rating



▲ Budget Analysis

Relationship Between Budget and Movie Rating



Budget Analysis



Average Movie Rating by Budget Group

5.69 6.33 6.64 6.80

< \$50M

11,476 Movies

**\$50M -
\$125M**

1,446 Movies

**\$125M -
\$225M**

415 Movies

> \$225M

62 Movies

Budget Analysis



Actors/Actresses

How do casting decisions affect rating?

132,123

Actors / Actresses

74,663

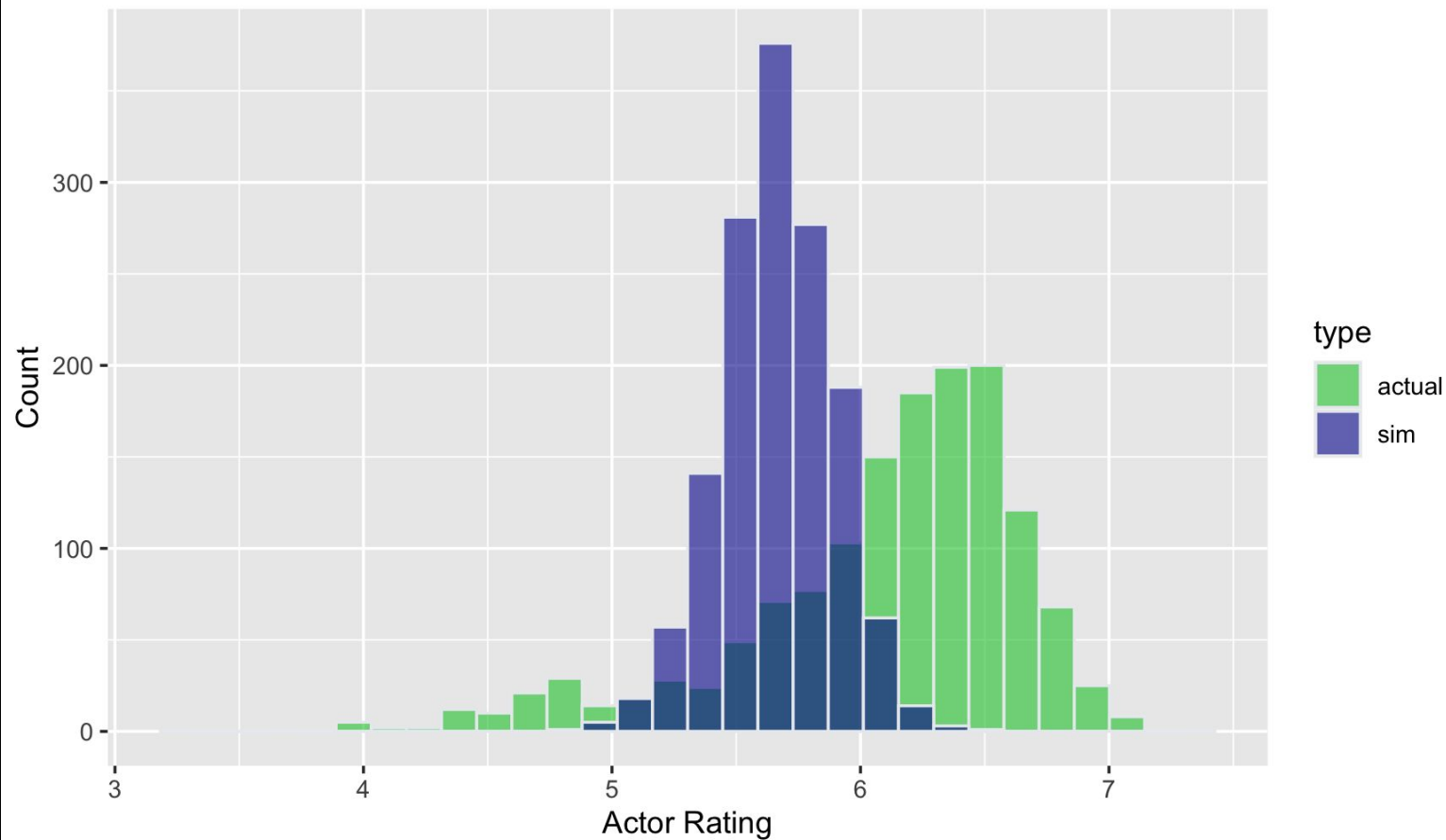
English Language
Actors / Actresses

1,424

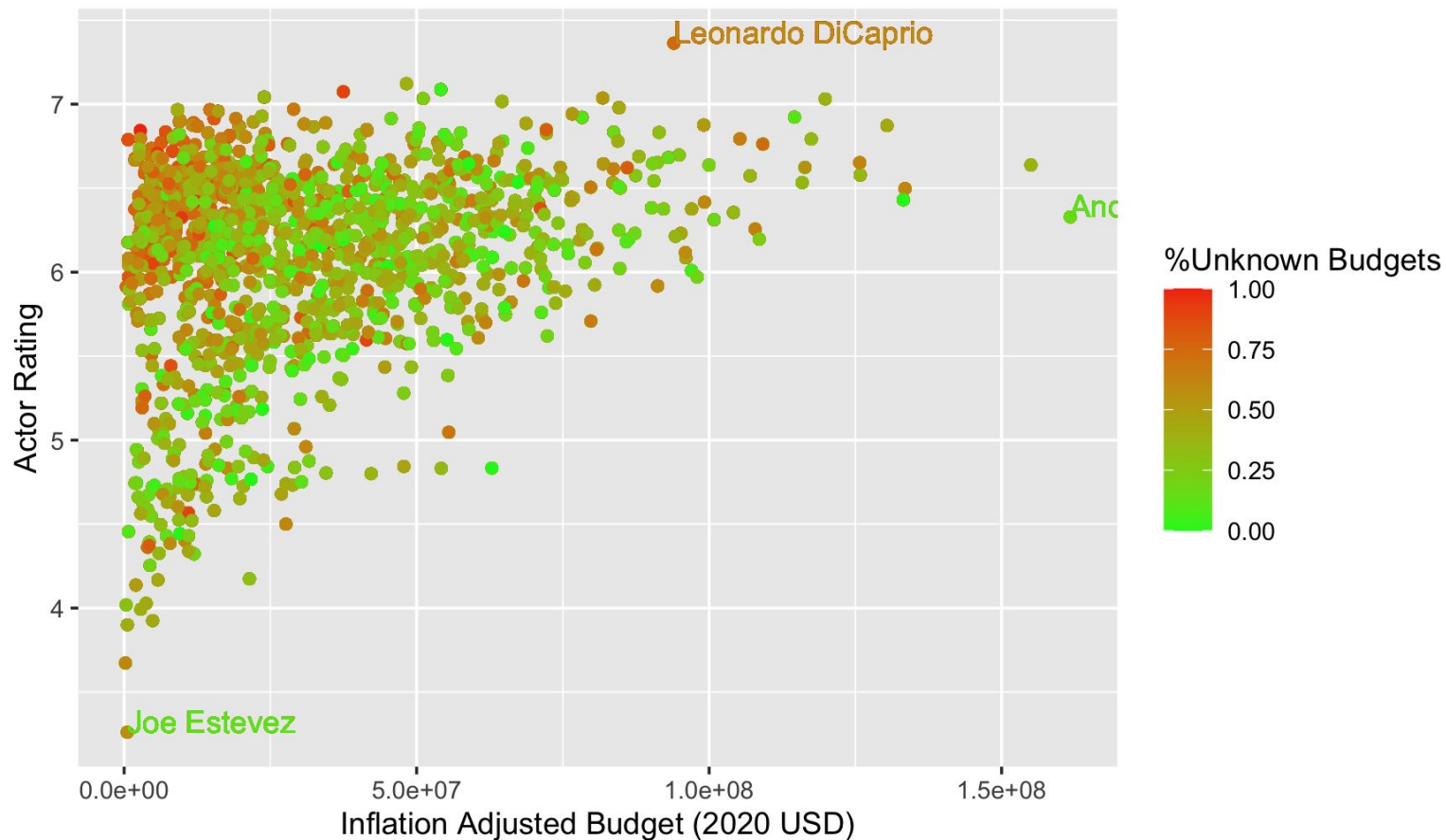
English Language
Actors / Actresses
who have been in
>= 20 movies



Distribution of Actor Ratings vs a Simulated Random Distribution



Actor's Rating vs their Average Movie Budget



04

▶ **VARIABLE
TRANSFORMATIONS**

Actor Rating

```
```{r}
actor_ratings = tp0 %>%
 filter(category == "actor" | category == "actress") %>% #filter non-actor crew
 select(imdb_title_id, imdb_name_id) %>%
 left_join(mv0, by = "imdb_title_id") %>% #bring in movie rating info
 filter(str_detect(language, "English")) %>% #only look at English movies
 group_by(imdb_name_id) %>%
 summarize(avg_avg_rating = mean(avg_vote), n = length(avg_vote)) %>%
 filter(n >= 20)
```
```

Movie Budgets

```
```{r}
budgets <- movies %>%
 mutate(budget = as.numeric(gsub("\\$", "", budget)), duration = as.numeric(duration)) %>%
 select(imdb_title_id, original_title, year, budget, avg_vote, duration, worldwide_gross_income) %>%
 arrange(desc(budget))
```
```

```
```{r}
budgets <- budgets %>%
 mutate(adj_budget = adjust_for_inflation(budget, from_date=as.numeric(year), to_date=2020, country = "US"))
 %>%
 arrange(desc(adj_budget))
```
```



05

▶ **LIMITATIONS**

- The number of votes can have large impact on rating
- Drama movie has taken up the most percentage
- Language is a barrier for analyzing datasets
- It is hard to set up a threshold of top-rated score when dealing with large datasets
- Many films had missing budget information.



LIMITATIONS



06

▶ **SUMMARY**



Sentiment Analysis

Plot descriptions for the top rated movies often contain words like "life", "love", "friend", and "woman", and the most common sentiment is "positive"

Genre Analysis

Documentaries have the highest average rating among all genres, while animated movies have the highest percentage of movies with a rating greater than 8.0



Budget Analysis

As production budgets increase, so do average movie ratings. Budgets >\$250M produced an average movie rating of 6.8, whereas budgets <\$50M produced an average of 5.69

Actor Analysis

There is a strong correlation between actors and movie ratings. Leonardo DiCaprio has the highest average rating, who is credited in 27 movies with an average rating of 7.36

