# *Will It Flop?*

*The pursuit of a predictive model*
*for film-to-stage adaptations*


Nick Wilders
October 9, 2020

# Setting the Stage

**EXPLORATION:** *How does performance of a film affect a stage adaptation?*

# Setting the Stage

**EXPLORATION:** *How does performance of a film affect a stage adaptation?*

# Setting the Stage

**EXPLORATION:** *How does performance of a film affect a stage adaptation?*


# Setting the Stage

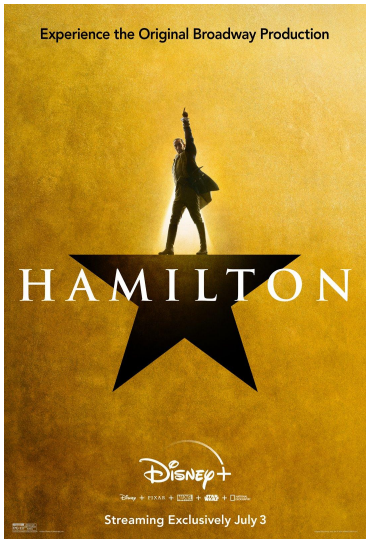
**EXPLORATION:** *How does performance of a film affect a stage adaptation?*



# Setting the Stage

**EXPLORATION:** *How does performance of a film affect a stage adaptation?*



# Setting the Stage

**EXPLORATION:** *How does performance of a film affect a stage adaptation?*

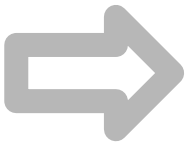



# Setting the Stage

**ASSUMPTIONS:**

- *Including all theatrical productions in a Broadway theatre in NYC (plays AND musicals)*
- ***For the purposes of this study**, it does not matter which comes first*
  - *"Movie first" and "Play/Musical first" bear equal weight*
- *Average gross per week == best measure of commercial success for a Broadway production*
  - *Total gross is biased by length of time production is open*

**FILM** *features (Budget, Domestic Gross, Worldwide Gross, Genre, Rating)*



**MUSICAL** *average weekly gross*

# Setting the Stage

**DATA:** *Web-scraped data from **IMDB***

- *10,000 "musical" movies scraped, combined with basic info for all movies on IMDB*
- *aMerged via title with...*

***Kaggle** data set with weekly grosses for Broadway shows (1985-2020)*

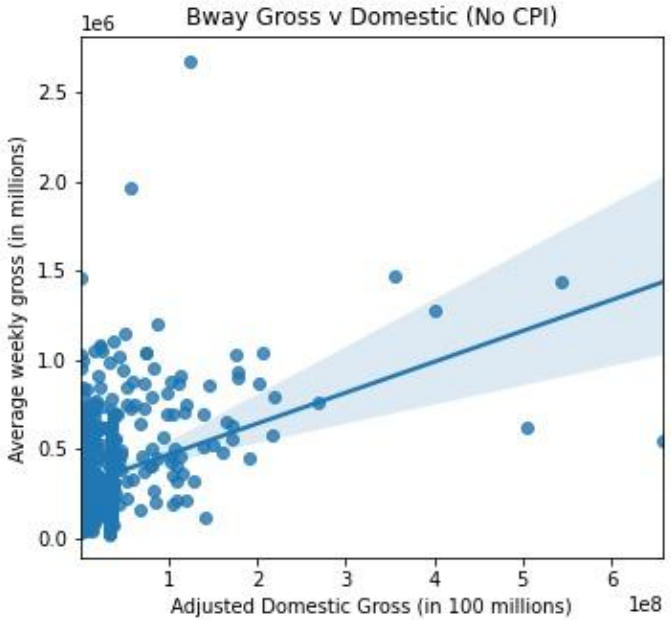
- *Leading to 27,000 title matches*

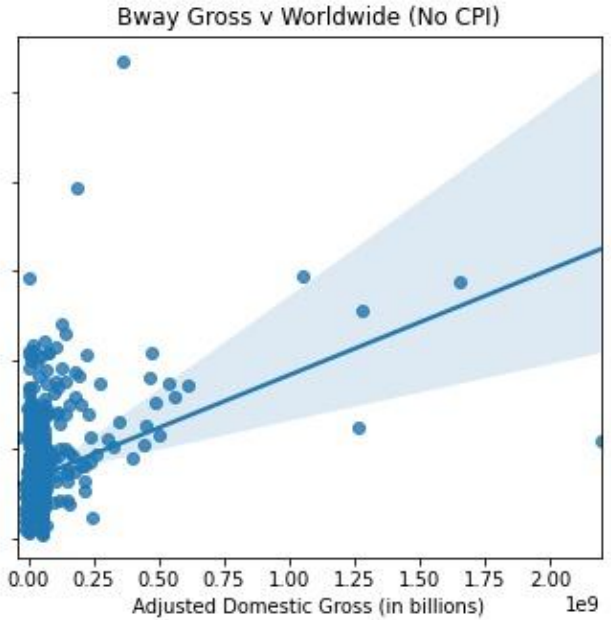
*CPI (Consumer Price Index) Data from **Datahub***

- *Applied to all price indices relative to time period*
- *Leading to 647 Broadway shows with film adaptations, eventually paired down to:*

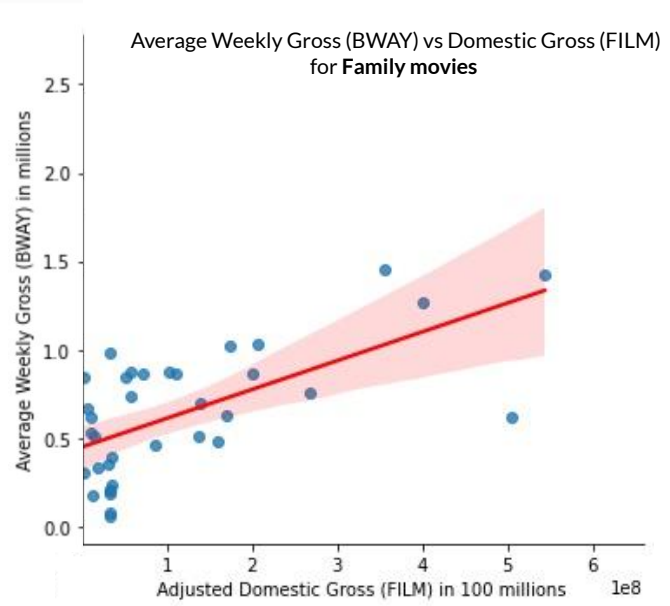
***547** **play / movie combinations***

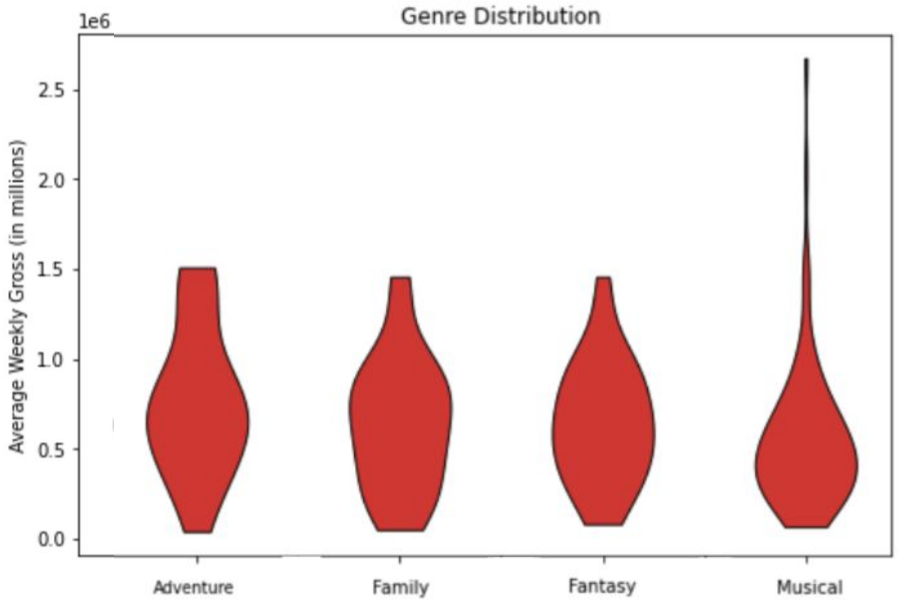
# Cursory data analysis implied a challenging data set - best coorelated was Domestic Gross (FILM) to Total Gross (BWAY)



## Family data was most consistently predicted of all genres, but a small sample size led to inconclusive trends and results.



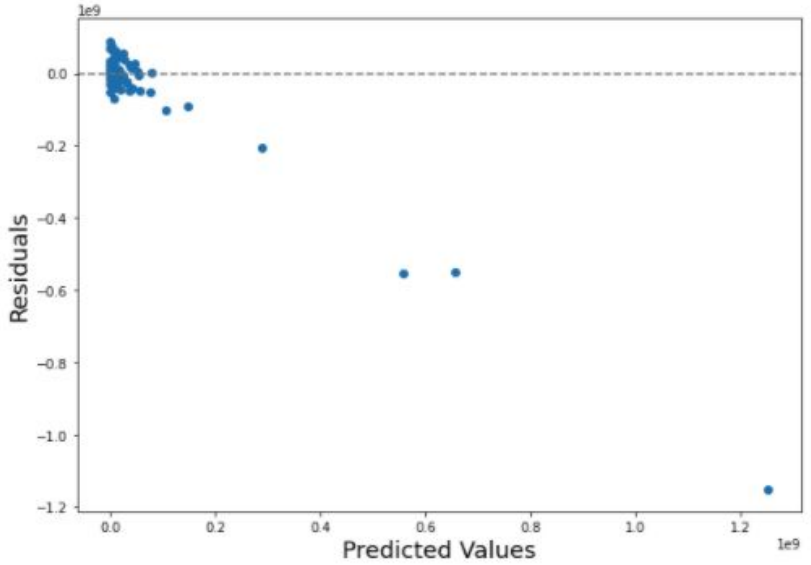



# The first model had massive outliers, and was missing the prediction up to $1mil

## MODEL 1

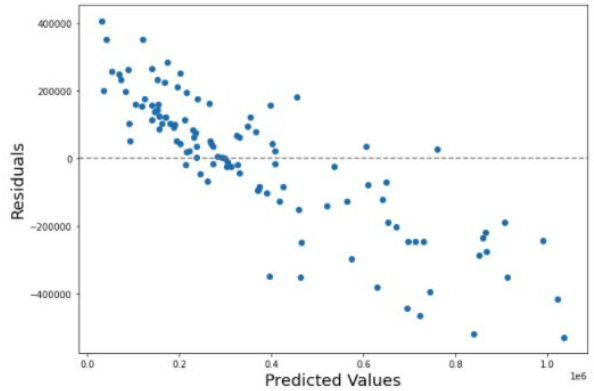
- ***Clear*** downward trend
- Long-running shows, higher total gross over time
- Outlier culprits:
  - *Phantom Of The Opera* (30 yrs +)
  - *Chicago* (20 yrs +) / *Cats* (18 yrs)

**TEST SCORES**-
LassoCV = - .007
RidgeCV = -.071



# I developed the model further by changing my target to average weekly gross, and incorporated CPI to adjust for inflation.

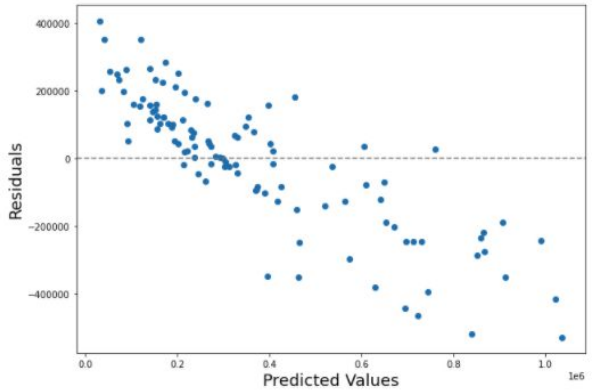
MODEL 2: Adjusted - Overall BWAY
Gross to Weekly Gross




**TEST SCORES**-
LassoCV = .39 (+ .40 increase)
RidgeCV = -.34 (-. 27 decrease)

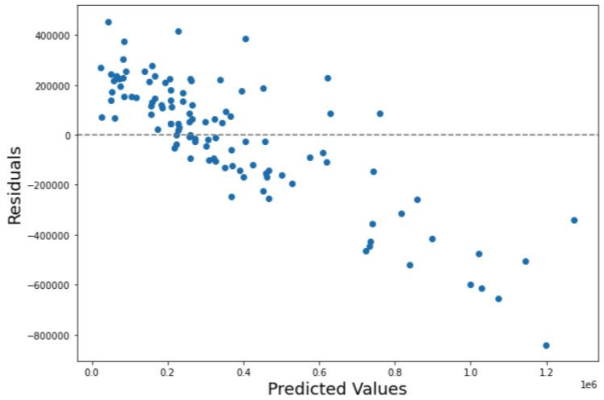
# I developed the model further by changing my target to average weekly gross, and incorporated CPI to adjust for inflation.

MODEL 2: Adjusted - Overall BWAY Gross to Weekly Gross




**TEST SCORES**-
LassoCV = .39 (+ .40 increase)
RidgeCV = -.34 (-. 27 decrease)

MODEL 3: Adjusted - Added Consumer Price Index for inflation




**TEST SCORES**-
LassoCV = - .25 (+.26 increase)
RidgeCV = -.27 (-.2 decrease)

# If I had to pick a model right now, it would be model 2.

- Capable of predicting average weekly gross:
  - Within $200,000 accuracy
  - With ~40% confidence

# If I had to pick a model right now, it would be model 2.

- Capable of predicting average weekly gross:
  - Within $200,000 accuracy
  - With ~40% confidence

| | |
|---|---|
| ***PARASITE*** | $3.6 million per week (more than *Hamilton)* |
| ***SHAWSHANK REDEMPTION*** | $904,970 per week (on par with 60% of shows) |
| ***FINDING NEMO*** | ***$4 million per week!*** |

# The downfall of this model is a small sample size, and a lack of information available for older movie musicals.

- 100 + “dud” pages cleared (merge based on title)
- 20 manual adds that eluded title merge
- Small sample size at n = 547

## The downfall of this model is a small sample size, and a lack of information available for older movie musicals.

- 100 + “dud” pages cleared (merge based on title)
- 20 manual adds that eluded title merge
- Small sample size at n = 547

## With more time and manual entry, I believe this model could be trained up to r^2 = .60 with much more confidence.

# Other factors this model could account for (notwithstanding sample size limits) include:

- Which came first? Stage vs Movie adaptation
    - Comparison of multiple adaptations (*Hairspray*, *Footloose*, Shakespeare,  etc)
- Influence of cast members / creative team (composers / lyricists included)
- Further scaling methods to decrease residuals
- Audience sentiment (Rotten Tomatoes, etc.)

# Thank you!

Any questions?