



# PREDICTING BOX OFFICE RETURNS OF A MOVIE SEQUEL

Charlie Lew

October 10<sup>th</sup>, 2019



# Background & Objective

- 45% of highest grossing movies past 40 years are sequels
- Studios need to make informed decisions on a sequel's prospective return and whether or not to green light it
- *Predict the box office return of a sequel through the use of regression analysis*

# Data Sources and Tools

- Data sources
  - International Movie Database (IMDb)
    - <https://www.imdb.com/list/ls003495084/>
  - Box Office Mojo
  
- Tools
  - BeautifulSoup for web scrapping
  - Linear Regression with Scikit-Learn



BeautifulSoup



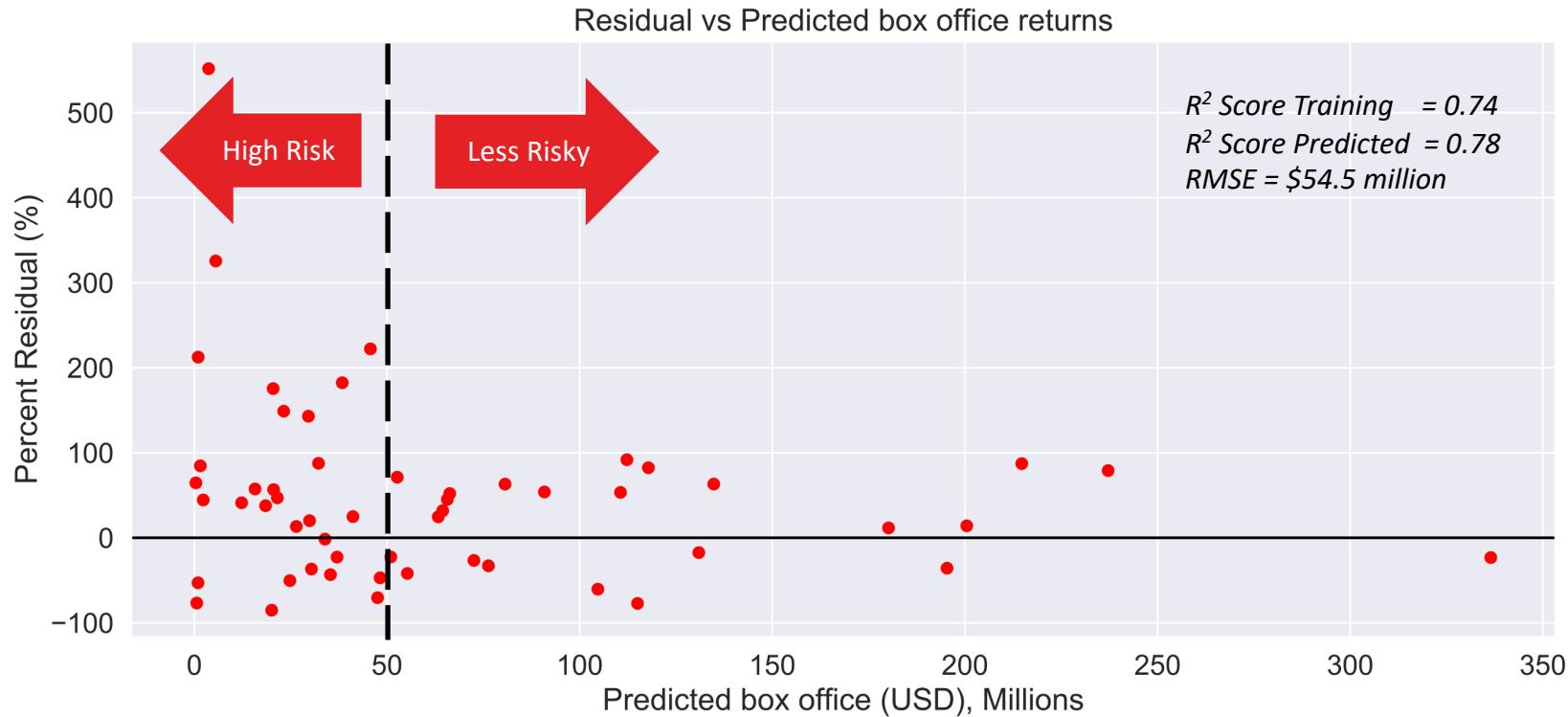
# Features & Observations

- Features
  - Gross domestic box office of original [USD]
  - IMDB Score Ratings
  - Days in between release
  - Release month
  - Runtime [mins]
- Data points and criteria
  - Movies that did not make the ‘cut’
    - Subsequent sequels (e.g. Superman III, Matrix: Revolutions)
    - Sequels direct to video or DVD
  - Nearly 270 movies

# Linear Regression Coefficients



# Predicted Returns from Linear Regression



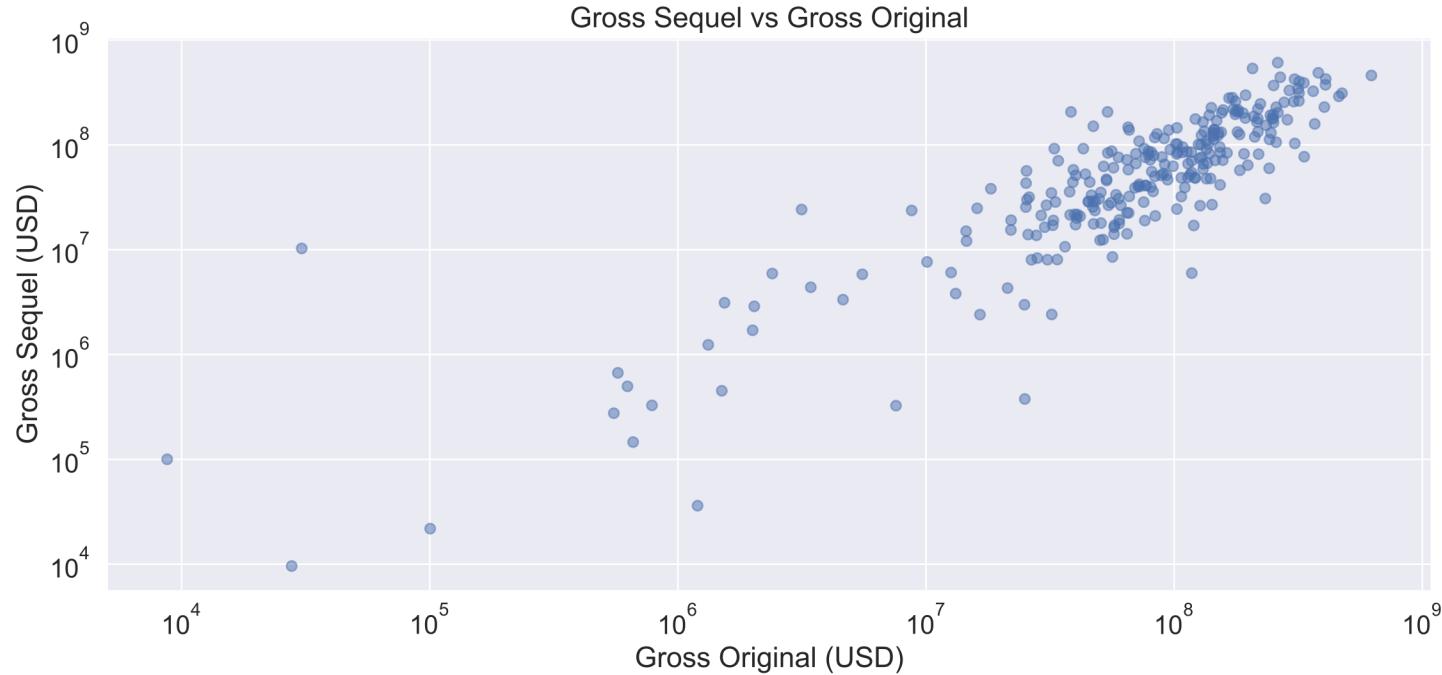
# Key Takeaways & Future Work

- Based on limited data & features
  - Gross box office original has strong effect on sequel box office return
  - Number of days between releases & runtime virtually no effect
  - Small effect of IMDB scores
- Future Work
  - Add more movies (foreign & domestic)
    - Add sub-sequent sequels perhaps
  - Additional features: Cast, Budget, Genre, Historical='Cult Classic?', etc
  - Alternative regressions (polynomial, etc)

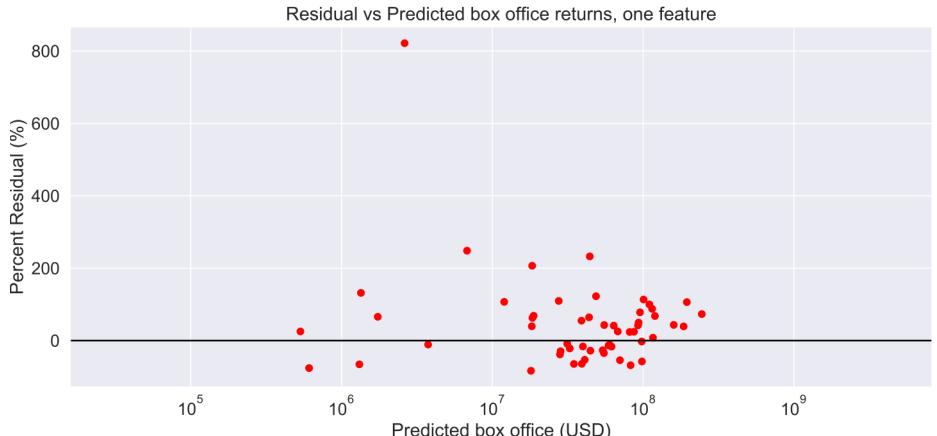
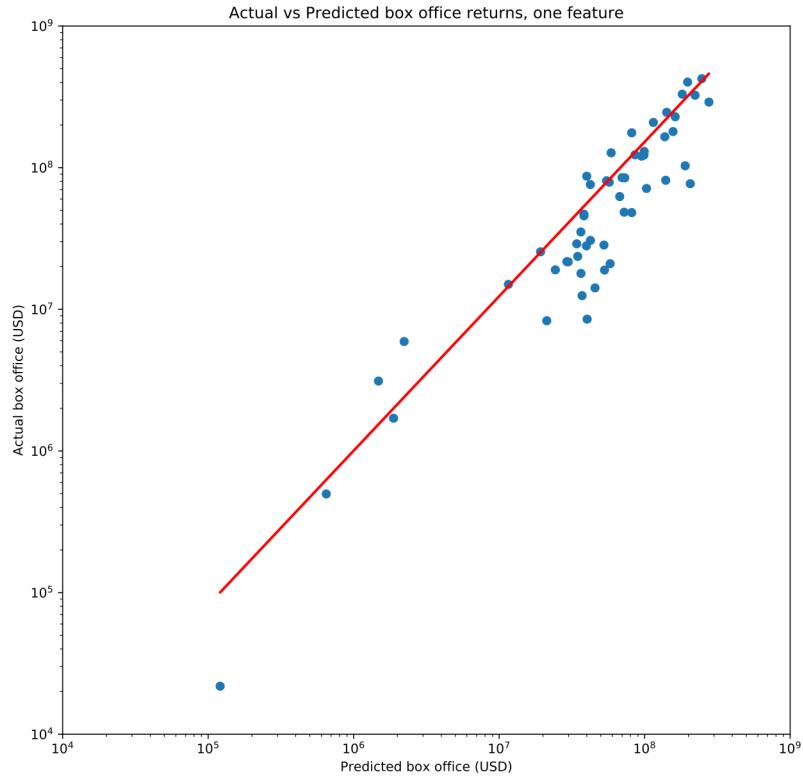
# *Thank You*

# *Appendix*

# Gross Sequel vs Gross Original



# Linear Regression with one feature



# Pair plots

