

A photograph of the Hollywood sign on a hill, with a radio tower visible in the background. The image has a warm, golden-hour tint.

# movie **BOX OFFICE** revenue app

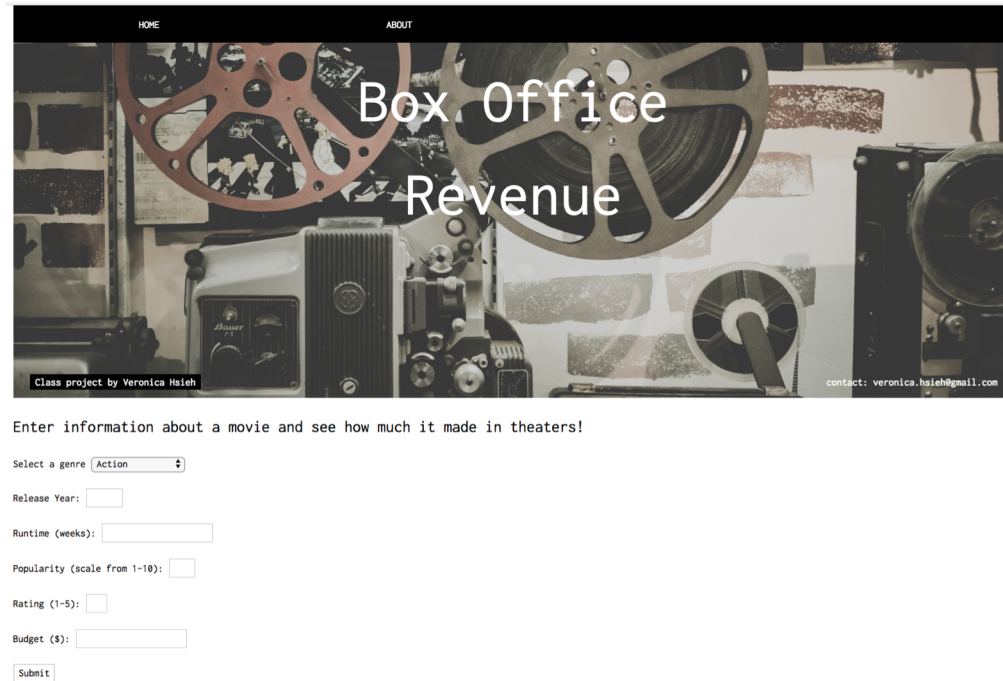
blockbusters unraveled

Monday, March 19<sup>th</sup> 2018

# MOTIVATION FOR APP

## GOAL

> Build a model to accurately predict movie revenue



The screenshot shows a web application interface for predicting movie revenue. The header has navigation links for 'HOME' and 'ABOUT'. The main title 'Box Office Revenue' is displayed in large white text over a background image of movie reels. Below the title, it says 'Class project by Veronica Hsieh' and 'contact: veronica.hsieh@gmail.com'. The main content area prompts the user to 'Enter information about a movie and see how much it made in theaters!'. The form includes a dropdown menu for 'Select a genre' (currently set to 'Action'), input fields for 'Release Year', 'Runtime (weeks)', 'Popularity (scale from 1-10)', 'Rating (1-5)', and 'Budget (\$)', and a 'Submit' button.

## REASON

> Understand what factors lead a movie to be a hit

# DATA



MovieLens dataset from Kaggle

Original dataset had 45,467 observations & 24 columns

After EDA & data cleaning, 5369 observations were used for model building & development

Transformations:

- Log scaling
- Categorical variables to dummies

# PREDICTION MODEL



DecisionTreeRegressor from Sklearn

Movie metadata used:

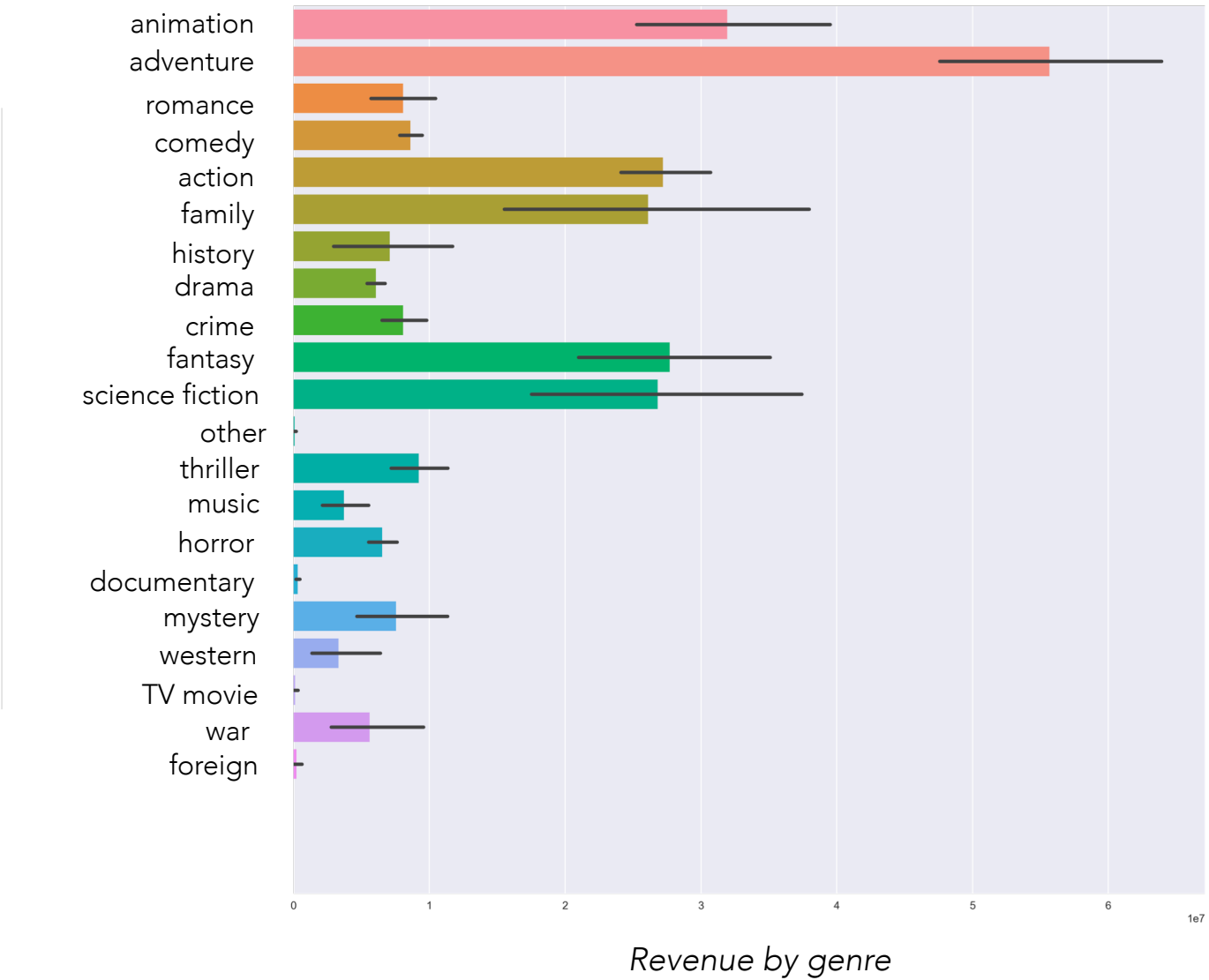
- Genre
- Runtime
- Budget
- Release Year
- Popularity
- Rating

Train-Test split: 80/20

Evaluation Metric: R-squared

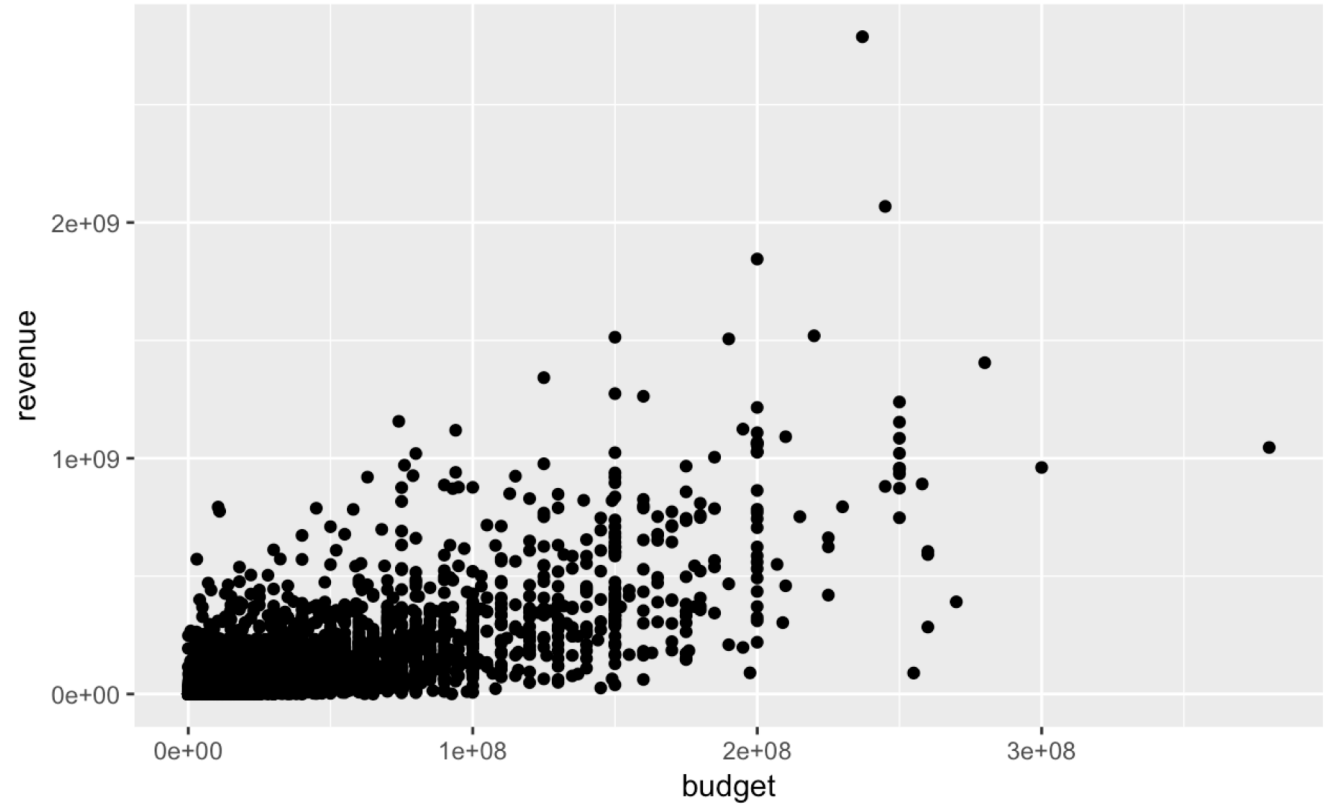
Movie goers enjoy watching genres that take them outside of reality

INSIGHTS



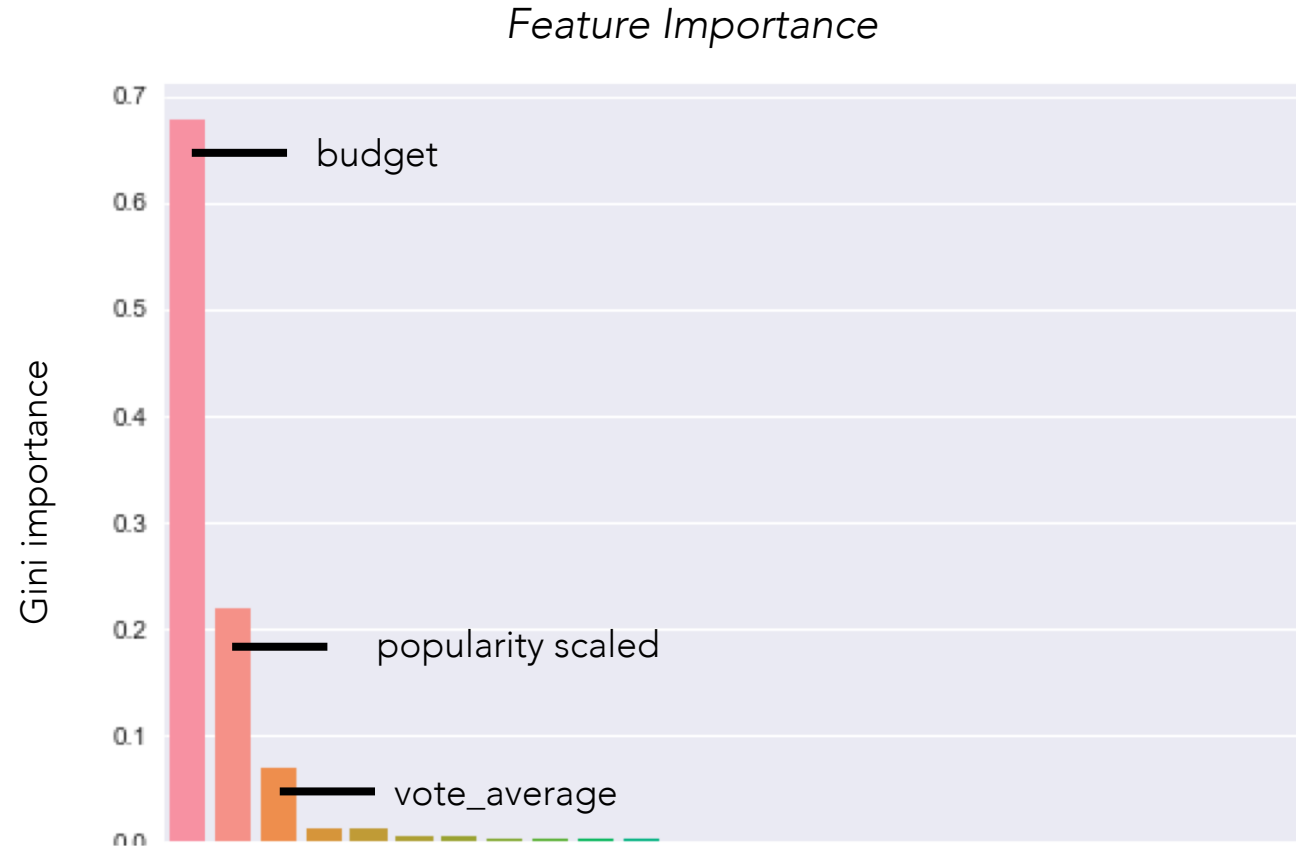
Movie revenue and budget have a strong correlation of .73

## INSIGHTS



# INSIGHTS

Movie budget and public perception of the movie are most influential in determining a movie's financial success



# THANK YOU

GitHub  
email

@ vhsieh920  
veronicahsieh2018@u.northwestern.edu