Movie Success Data Analysis

Parmvir Johal, Raymond Eng

August 3, 2018

# Problem

Producers, investors and organisations are highly interested in determining what makes a successful movie. What makes a successful movie? How should we determine success?

# Data

We used multiple sources of data, the first being data to describe movies from WikiData. The wiki data was downloaded as a compressed JSON and was filtered using spark to get the meaningful data. The code was provided by Greg Baker but we made changes to include more data such as box office prices, movie cost and actor labels. Greg baker used a scraping algorithm to get data from rotten tomatoes. The rotten tomatoes data include audience and critics data, and could easily be joined with wikidata using the uniques imdb ID. Lastly we used the OMDb API to get movie plots based on the imdb ID. Most of the data was clean but the occasional empty values were filtered out using pandas in our notebooks.

# Analysis & Techniques

## Success correlation

## Genre profitability

## Predict genre

## Genre Popularity based by year

## Audience Rating Prediction

# Results

## Visuals

# Limitations