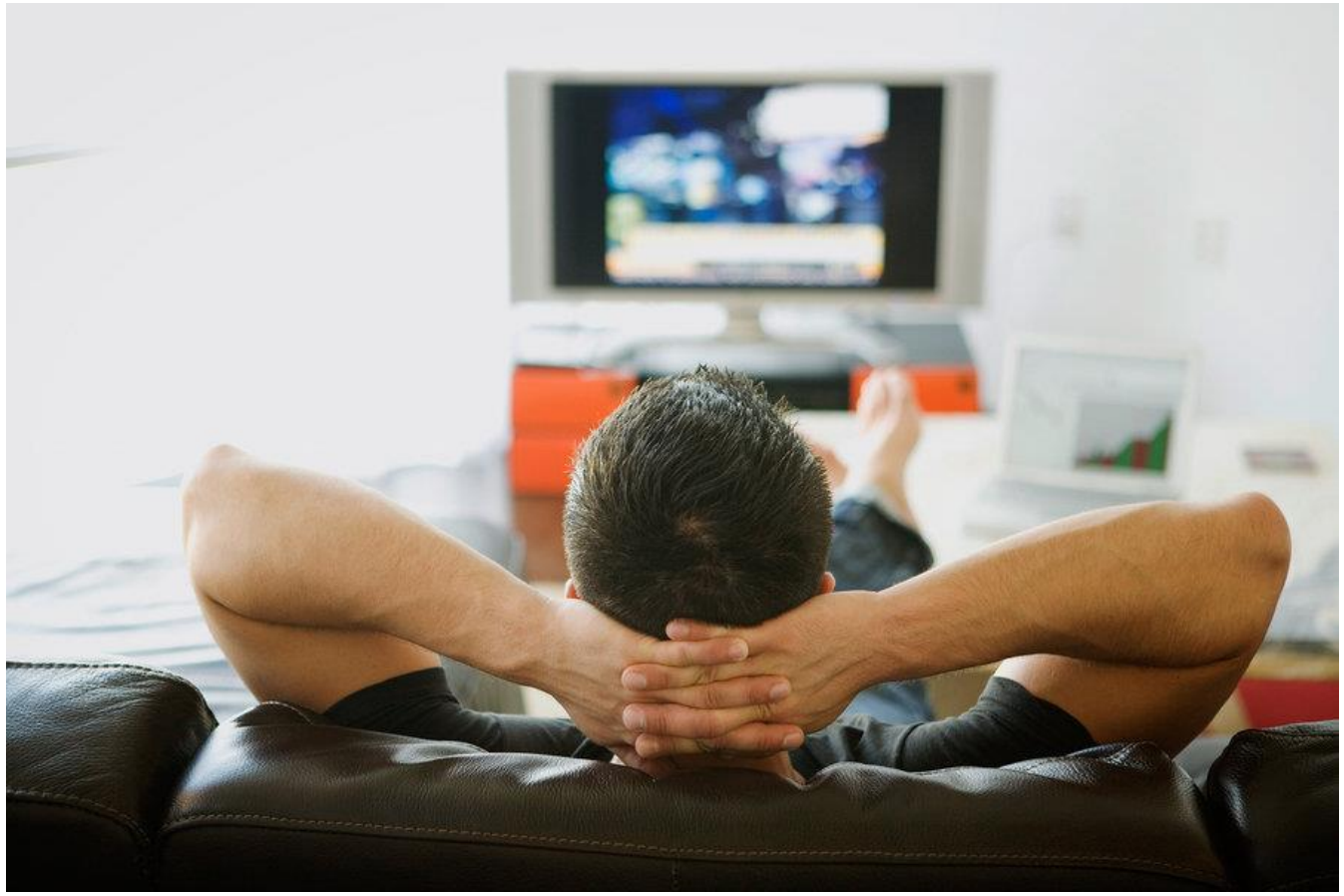
FINAL PROJECT 1

JANELLE CORDERO

DETERMINE A PROGRAM EPISODE'S LUCRATIVE SUCCESS BASED ON PROGRAM RATINGS AND SPECIALS



www.npr.org

- Typically, media companies measure the success of original programs on their ratings.
- Ratings correspond to the average number out of people watching out of all that.
- Media companies sell to advertisers on the ratings of commercials.
- A portion of the revenue comes from fees from cable providers.

DATA DICTIONARY – EXAMPLE FEATURES

Field Name	Type	Description
Originator	String	Network where the program originated
Episode	String	Name and/or number
Special	String	(S/blank) binary that tells whether a program was a special
Telecast	Integer	The amount of airings
Total Duration	Continuous	The duration of an episode
MC US AA%	Continuous	Episode Rating
MC Weighted NCCM AA %	Continuous	Commercial Rating

- The data focuses on all episodes that aired during the week of 9/28/2016 for all persons at least two years of age in a television household within the U.S.

+ 13 More…

DESCRIPTION OF FEATURES

Stat	Telecast Count	Duration
Count	28,935	28,935
mean	2	100
std	3.993694	187.407599
min	1	7
25%	1	30
50%	1	60
75%	2	120
max	178	5,322

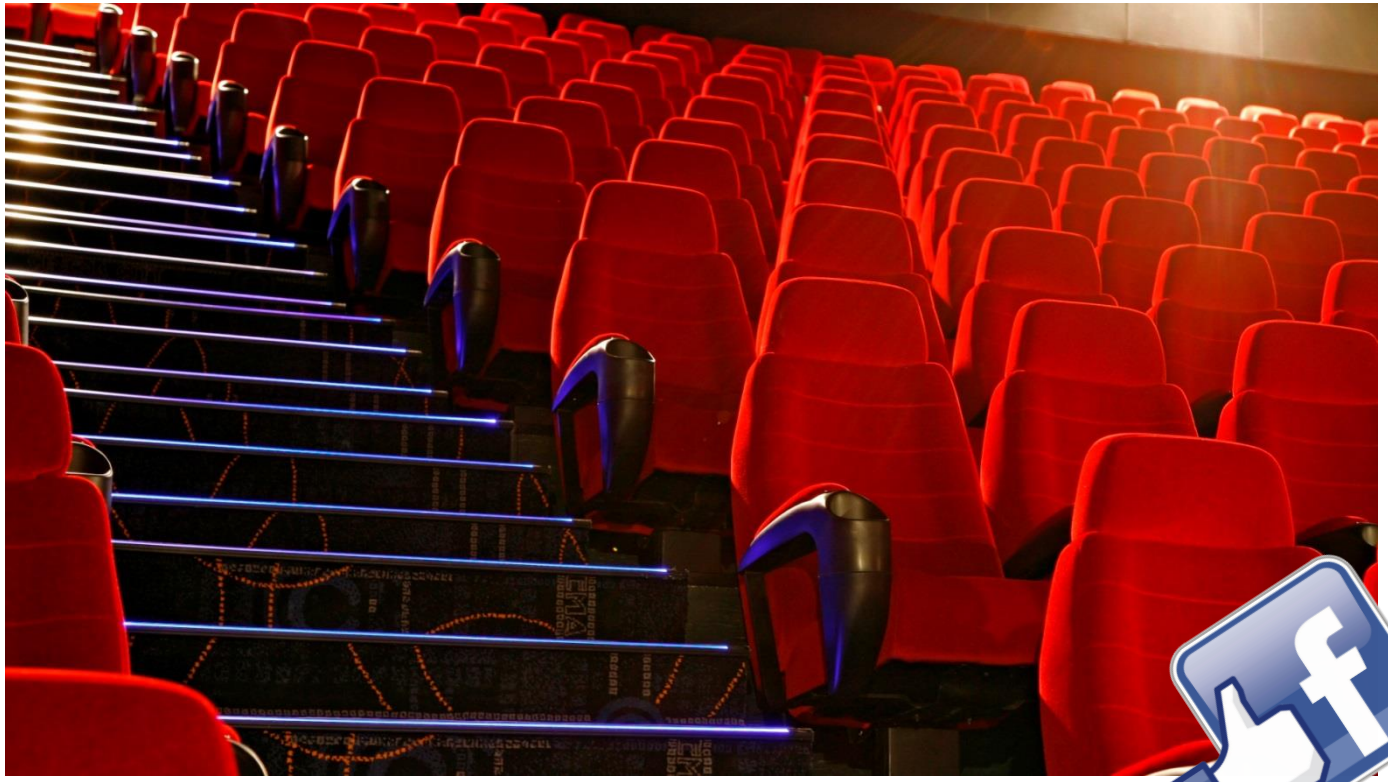
SOLUTION/HYPOTHESIS

In order to predict success of an episode, I will use the commercial rating as the outcome variable since it is vital in selling to advertisers and revenue.

I will use a machine learning algorithm, likely logarithmic regression, to predict what changing each variable can do to increase/decrease the commercial rating.

To predict a higher commercial rating, several factors (i.e. rating, specials, number of telecasts (which predict the amount of repeats)), and originator may affect a program episode's commercial rating. The program episode rating and whether or not a program is a "special" will both predict a the commercial rating of an episode.

WHAT INTERACTION BETWEEN TRADITIONAL KPIS AND SOCIAL MEDIA FACTORS PREDICT THE SUCCESS OF MOVIES?



http://practicallyviral.com

- An IMDB rating, customer reviews, budget, actors, and directors are often good indicators of how great a movie will turn out to be.
- Social Media often gives us a glimpse into what is currently popular based on buzz around a topic.
- Social Media aggregators like Nielsen Social and Brandwatch are already mastering social media buzz in specific industries especially TV and movies.

DATA DICTIONARY – EXAMPLE FEATURES

Field Name	Type	Description
Movie_Title	String	Title of the Movie
budget	Continuous	Budget Allocated for movie
content_rating	String	A rating of age appropriateness of the material
imdb_score	Continuous	Official score given on IMDB
actor_1_facebook_likes	Continuous	Likes on actor's Facebook page
director_facebook_likes	Continuous	Likes on director's Facebook page
movie_facebook_likes	Continuous	Likes on movie's Facebook page
gross	Continuous	Amount of money made by movie

- The data looks at 5,00 movies titles over the course of 100 years along with traditional performance variables and Facebook likes of the actors, director, and movie as a whole.

+ 20 More…

DESCRIPTION OF FEATURES

Stat	gross	Movie_facebook_likes
Count	4,159	5,043
mean	48,468,407.53	7,525.96
std	68,452,990.44	19,320.45
min	162	0
25%	5,340,987.50	0
50%	25,517,500.00	166
75%	62,309,437.50	3,000
max	760,505,847	37,953,439

SOLUTION/HYPOTHESIS

The budget of a movie and IMDB score are solid indicators on the success of a movie, but do not always tell the complete story. The Blair Witch Project had a production budget of only $60,000, but grossed nearly $250 million.

Social Media buzz around movies cause “highly anticipated films” and expectations before movies are even released.

I will look at traditional KPIs (budget, IMDB score) alone, social media metrics (Facebook likes), and the interaction of the two to more accurately predict a movie’s success.

WHO WHERE AND WHAT ARE SIMPSONS FAVORITES?

- The success of its program is often based on its ratings which corresponds to the amount of views

http://simpsons.wikia.com

The Simpsons Dataset

DATA DICTIONARY – EXAMPLE FEATURES

Field Name	Type	Description
Simpsons_script_lines.csv		
simpsons_characters.csv		
Simpsons_locations.csv		
Simspons_episodes.csv		
Views	Continuous	Amount of views an episode received

- Three Datasets are used with unique IDs that are all included and linked in simpsons_episodes.csv to.

- Simpsons_script_lines.csv includes 158,271 instances of text that a specific character mentioned
- simpsons_characters.csv lists the characters in The Simpsons with a unique ID
- Simspons_locations.csv lists all location that appeared in The Simpsons with a unique ID

SOLUTION/HYPOTHESIS

Very detailed factors of TV sows are not often used to predict the success of a program because it is difficult to quantify

Determine the success of an Episode of the Simpsons based on scripts, characters present, and location.

The Simpsons is a classic and all memorable factors often include the main characters, a few secondary characters (i.e. Moe), and noteworthy lines (i.e. D’oh!). Using random forests, I can determine if these two factors will play the biggest roles in determining the success of an episode.