**Project Title**: Film Budgets and Box Office Revenue

**Team Members**: Danielle White, Al Misa, Susan Skinner, Alex Goldstein, Ciara Spencer, Jeff Johnson

**Project Description / Outline**:

Our project aims to find correlative trends among films produced from 2006 to 2018 that correspond with budget and box office revenue. We’ll examine relationships between genre, date of release, and MPAA ratings year to year.

**Research Questions**:

* Correlations between production budgets and box office revenue based on genre
  + year to year
* Correlations between production budget and box office revenue based on MPAA rating
  + year to year
* Release Data: If films a certain month performed better based on ROI then others [i.e. movies released Christmas vs. Summer?]
  + Did certain genres perform better in summer vs Christmas?
  + Were certain genres released more in winter vs summer
* Production Companies: Will examine budget and box office data to compare and determine which production companies performed better when comparing ROI.
* Taking the most profitable
* International vs Domestic Box Office: comparing genres performance domestic vs international.

Breaking down

By budget (range bins)

By rating

By genre

By budget

Then rating

By genre

Percent Successful

Films with over certain percentage ROI

Taking top performers

Highest roi films

By genre

Rating

Percent successful

Films with over certain percentage ROI

**Datasets**:

1. ***The Numbers: from Opus Data (DataSet .csv file)***

About the dataset: Film Data for Movies produced between 2006 -2018, figures for domestic and international box office with productions over $10 Million.

1,900 films with data values for movie name, production year, movie id, production budget, domestic box office, international box office, rating, creative type, source, production method, genre, sequel, running time.

* Consistent numbers are provided within the data set with minimal NaN values for: movie name, production year, movie id, production budget, domestic box office, international box office, rating, creative type, source, production method, genre, and sequel.
* Column for running time is insufficient for data values for majority of the films listed.

Well Documented:

“OpusData is backed by a rigorous classification methodology that makes it possible for independent producers, investors and industry researchers to build accurate financial models and to measure the performance of individual movies, production companies, studios, market segments, and the industry as a whole.”

1. ***OMDB: (API)***

About the database: An API to obtain relevant movie information for title, release year, MPAA rating, date released, runtime, genre, director, writer, plot synopsis, languages, country released, nominations, ratings (IMDB, Rotten Tomatoes, Metacritic, Metascore), box office revenue, and production company.

* Need to confirm that it contains relevant data for majority of films within the csv file from Opus Data.

**Rough Breakdown of Tasks**:

Danielle White - Data Clean-up & Analysis

Al Misa – Data Clean-up & Analysis

Susan Skinner – Matplotlib

Alex Goldstein – Data Clean-up & Analysis

Ciara Spencer – Wrangling & API’s, Presentation

Jeff Johnson

**Relevant Data Comparisons**:

The Numbers Data

* Movie Title, Production Year, Production budget, Domestic box office, International box Office, Rating, Genre

OMDB

-Country of Release, Date Released (MM/YY), Critic Ratings (IMDB, Rotten Tomatoes, MetaCritic, MetaScore), Production Company

Extra: can look at data of all four metrics of critic ratings and determine if higher ratings correspond with ROI and/or Total Box Office Revenue.

Extra: Can cite inflation calculator that adjusts budget and Box Office Revenue Totals year to year from 2006 – 2018 to compare models for accuracy of assumptions.

Notes: ROI is determined by taking the Total Box Office Revenue, dividing by the production budget and multiplying by 100 for percentage value.