Data Science Project Proposal

Group V

Blake Lockley (44587643), Sam Hornsey (44851375)

Art vs Money.
Who's making our films?

Looking at the poster for almost any Hollywood blockbuster you will notice that the names of the starring actors never seem to line up with the actor or actress on the poster itself. One would assume that the poster designers would line the names of actors up with the order they are shown on the poster. However, it turns out there is a very specific reason for the position and order of names on a movie poster. In the film industry, there is a concept called '**Top Billing**' and it refers to the earnings and screen time of each actor in a film.

Top Billing will usually be allocated to the actors with the most significant roles and screen time for the particular film. In the past it was also common practise that top billing was determined by the actor's level of fame. The order of Top Billing is an important feature of a film that can suggest the significance of an actors role and potentially the success of the movie. A great explanation can be found in the video: Why Don't Movie Poster Names EVER Line Up?! by Austin Mcconnell, which served as inspiration for this project.

Delivery

Our analysis will be presented in the form of an 'interactive visual essay' web app. As the user scrolls the page narrative or our analysis will develop. The user will be able to select particular areas of interest such as their favourite actors and films for interactive graphs. A back end will be implemented to allow for the dynamic creation of predictive models and an API to fetch new data (as new movies are released) has been already developed. During the development of our project we will create a collection of Jupyter notebooks to experiment and illustrate our analysis.

Our analysis will explore two primary ideas related to the concept of Top Billing.

1. An actor's progression over the period of their career. This involves observing an actor's position in the billing order over time, the other actors they co star with, and success of the films they are in. This is a focus on the individual actor.
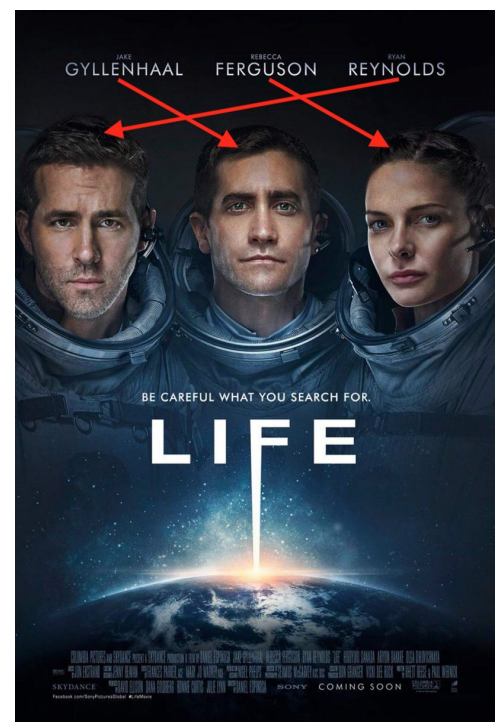
2. The success factors of a film based on the composition of the cast. By exploring the actors and the order in which they appear to make predictions about a films success. This is a focus on the film and its features being the starring actors.

A dataset is being created indexing the entirety of films listed on Wikipedia. This is being accomplished using scraping techniques to generate reusable csv files. Wikipedia will consistently list the starring actors of a film in the correct billing order (this is predominantly true for all major films). The headings within the data set are currently as follows.

```
["title", "director", "starring_0", "starring_1", "starring_2", ...,
"starring_8", "starring_9", "released", "budget", "box_office",
"url"]
```

The data set provides columns for 10 starring actors. As mentioned the order is essential to the concept of Top BIlling, the actual of position of the actor in this order is considered a feature of the film. Preliminary datasets of movies from 1990-2005 (2366 samples) have been created and are being explored, due to formatting inconsistencies some further processing is required.

Analysis Techniques to be Implemented
Regression (**Simple Linear, Random Forest**): Predicting film's success factors.
Time Series Analysis (**ARIMA**): Observing actors career progression over time.
Clustering (**K-Means**): Discovering similar movies due to their casting.
Classification (undecided): Interactive module to predict the success of a "Dream Cast".

Project Milestones
Milestone #1 - Project Base - Week 8:
Establish the base of the project this includes, source control (GitHub), project management (Trello), web framework (Flask), datasource (Wikipedia - scraping), **all tasks are complete.**

Milestone #2 - Model Identification and Analysis - Week 10:
Create completed dataset of movies from 1960 - 2020.
Complete jupyter notebooks with the intended analysis to be displayed on the web app.

Milestone #3 - Web app - Week 12:
Decide and implement all interactive features of the web.
Complete web app and deploy on a platform such as Heroku.