

## Initial Design Paper Landscape and Proposal

### Introduction

This landscape proposal includes background information about the dataset, the specific persona, and the 2 critical questions to be beneficial for the persona. The requirements will also be included on how I plan to approach and answers these 2 questions through data visualisation. Lastly, will provide a final drawn out paper landscape.

### Data

The movie industry dataset was extracted from Kaggle dataset (<https://www.kaggle.com/>). The dataset was created in 2017 by Daniel Grijalva, a software engineer (<https://www.kaggle.com/danielgrijalvas/movies>). The table below illustrates the metadata of the dataset, showing the different columns, with a description of and the data type of each variable. The variables highlighted in bold are the ones that I intend on using for my Infographic (Tableau Dashboard) as well as new variables I plan on adding to from existing variables.

Name	Description	Data type/Domain
name	Name of the movie	Categorical (nominal)
rating	Ratings of the movie	Categorical (nominal)
<b>genre</b>	<b>Main genre of the movie</b>	<b>Categorical (nominal)</b>
<b>year</b>	<b>Year of Release</b>	<b>Integer (continuous)</b>
released	release date (YYYY-MM-DD ((country of release))	Integer (continuous)
scores	IMDb user rating	Categorical (continuous)
votes	number of user votes	Integer (continuous)
director	the director	Categorical (nominal)
writer	writer of the movie	Categorical (nominal)
star	Main actor/actress of the movie	Categorical (nominal)
<b>country</b>	<b>Country of origin</b>	<b>Categorical (nominal)</b>
budget	the budget of a movie.	Integer (continuous)
gross	revenue of the movie	Integer (continuous)
company	the production company	Categorical (nominal)
runtime	duration of the movie	Integer (continuous)
<b>Date</b>	<b>Date of release</b>	<b>Integer (continuous)</b>
<b>Month</b>	<b>Month of release</b>	<b>Integer (continuous)</b>
<b>Year</b>	<b>Year of release</b>	<b>Integer (continuous)</b>
<b>Country of release</b>	<b>Country of release</b>	<b>Categorical (nominal)</b>
<b>Net Profit</b>	<b>Gross-Budget</b>	<b>Integer (continuous)</b>

### Persona and Questions

The persona: movie-producer

Q1. What is the net profit across the top 10 movie production companies and does the release date of the movie have an impact on this?

Does the release date influence the net profit of the top 10 movie production companies?

Q2. Given the score of each movie, is there a relationship between movie genre and the top 10 movie production across the US, UK and Europe?

### Requirements

1. To approach and answer question 1, the user should be able to see the trend of net profit for the different companies across each year
  - a. Will create a new attribute called net profit (Gross – Budget)
  - b. The net profit is the measure (in count)
  - c. The company is the dimension
  - d. Utilise line graph to show the trend in net profit across the years and month

2. Will add a radical bar plot to show an alternative view of the net profit across which movie production company in ranked order.
  - a. Will have global brushing and linking for the movie production
3. To approach and answer the first part of question 2 (relationship between movie genre and movie production), we need to show the user nominal and numerical comparison between movie genre and the top 10 companies (through the movie scores)
  - a. Will do this by using a treemap
  - b. Genre will be colour
  - c. Company will be size and label
  - d. Will add % of movies score as label as well
4. To answer the second part of question 2 (movie genre across the US, UK and Europe), will need to also show the user the relationship comparison
  - a. Will use a stack boxplot
  - b. Countries (nominal) against the movie scores (numerical) with filtered movie genre being the colour encode
  - c. With this box plot and the treemap, the genre will be the global filter for these two plots.

## Paper Landscape

### What Makes a Movie Producer Successful?

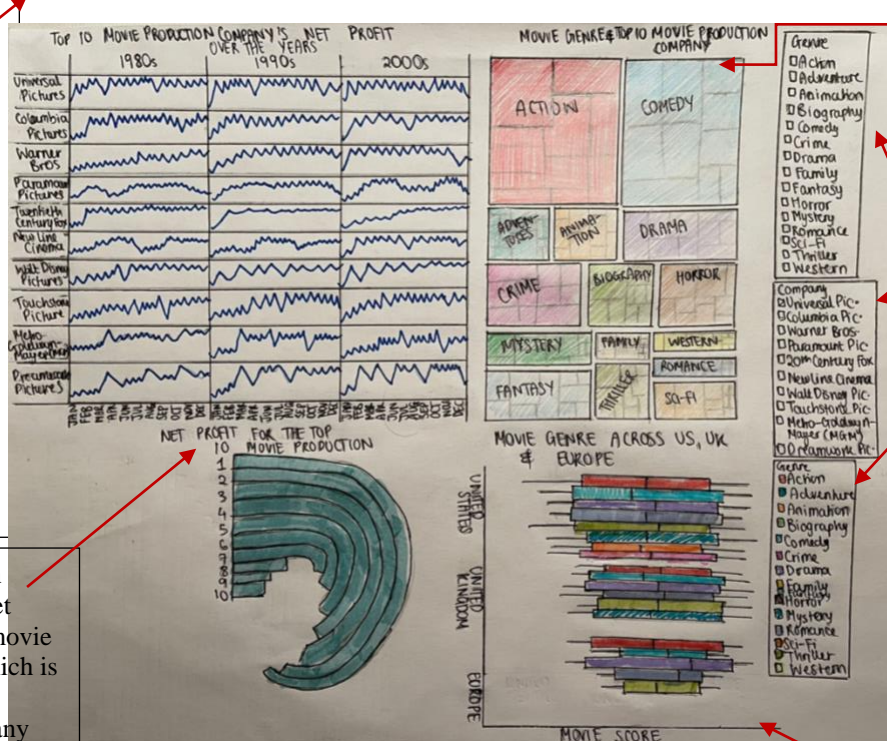
Goal: Explore what makes a movie producer successful through exploring different factors such as movie scores, movies genres, net profit and movie of release across countries.

Q1. What is the net profit across the top 10 movie production companies and does the release date of the movie have an impact on this?

Q2. Given the score of each movie, is there a relationship between movie genre and the top 10 movie production across the US, UK and Europe?

Shows the trend of net profit for the different companies across each year through line graph. The time-series is divided between 1990s, 1980s and 2000s movies, show the net-profit trend over the months over the each top 10 movie production. The time-series on the month would probably be average, e.g. average of all January from 1980 - 1989.

The second plot to get an alternative view of the net profit across the top 10 movie production company (which is shown in numbers), will replace it with the company names. This is ranked which from highest to lowest, which will be either in average or sum.



Treemap to show the relationship between movie genre and the top 10 movie production (1<sup>st</sup> part of Q2)

The global linking and brushing for Q1 and Q2.

boxplot to show the relationship between movie genre across US, UK and Europe (2<sup>nd</sup> part of Q2)  
The movie genre is colour encode according to the selected filtered and the colour will be the same as the TreeMap

## Results:

The results of this infographic should provide an insight of what makes a movie producer (or production) successful through different factors.

For Q1, seeing a trend of the net profit over time can give an insight over financial around movies have changed over time; is the same or is there a drastic increase from 1980 to 2000 in order to see where budget and gross can be improved, for example. For Q2, the relationships can give an insight on, for example, for movie producer who are looking to work for production movie company and want to work with the suitable company according their success in certain movie genre.

Alternative to the radical plot for the Q1, was thinking of doing a square plot showing different net profit interval across the movie companies over certain months such as summer month. Or making 4 grid plot that divide the net profit into high and low and doing a dot plot of the movie companies:

