Project: Visualizing Movie Data

Business problem: The client is a movie production company, who wants to understand movie trends to help inform their decision making. In particular the client is interested in answering the following questions:

- **Question 1:** How have movie genres changed over time?
- **Question 2:** How do the attributes differ between Universal Pictures and Paramount Pictures?
- **Question 3:** How have movies based on novels performed relative to movies not based on novels?
- **Question 4:** Investigate a trend of interest.

In order to complete this report, data from “The Movie Database” which is an open source database for movie meta data.

Step 1: Data Cleanup and Attribute Selection

Data was first imported in Alteryx Designer 2018.3, where null values were removed, and movies reporting zero budget or revenue were removed from the analysis. The justification for the latter decision is that these entries represent either input errors or conversely are amateur produced movies, neither of which are relevant to this analysis. The keywords, production companies and genre columns were parsed into their individual components and used to answer questions posed by the client. The eight attributes explored in this analysis are adjusted budget, adjusted revenue, estimated profit (adj. budget – adj. revenue), number of movies, movie runtime, movie vote count and movie vote average.

Step 2: Tableau Visualizations

Question 1:

https://public.tableau.com/views/Movie_Visualisation/MoviesGenreAnalysis?:embed=y&:display_count=yes

Question 2:

https://public.tableau.com/views/Movie_Visualisation/ProductionCompanyStory?:embed=y&:display_count=yes&publish=yes

Question 3:

https://public.tableau.com/views/Movie_Visualisation/MoviesbasedonNovelAnalysis?:embed=y&:display_count=yes

Question 4:

https://public.tableau.com/views/Movie_Visualisation/WhatContiniousvariablesDriveProfit?:embed=y&:display_count=yes

Step 3: Questions

Question 1:

Data was aggregated across genre and year. As can be seen from the dashboard “Movie Genre Analysis” (Percentage of All Movies), Drama, Comedy, Thriller and Action are the most common movie genres from 1990. Interestingly, by viewing the rolling average as a percentage of total movies, a drop off in the number of comedies can be observed from 1989. This is not captured when viewing the absolute figures. It is also worth noting, that the mean number of genres per movie has reached an all-time low since 2012, suggesting that producers are being more selective with their genre labels (Number of Movies and Genres per movie).

Question 2:

This tableau story focused on the profitability of movies for each of the production companies.

Across all years there has been little difference in profits of either production company [overlapping interquartile ranges](Sheet 1). With the expectation of highly profitable movies in 1972 to 1977, average profitability has not changed much over time. However total profits have demonstrated an upward trend (Sheet 2). The trend for increased total profits, appears to be generally driven by an increase in the number of movies, despite increasing movie budgets (Sheet 3). In the early 2000s, paramount demonstrated a drop off in total profitability (~1.5 billion in 1999 to ~0.5 billion in 2004). In response the production company changed their strategy, by almost halving the number of movies produced per annum between 2004 and 2010 and increasing the average movie budget from ~50 million to ~100 million per movie in the same time period. This appears to have succeeded with paramount stemming lost profits and for the first time since 1999 surpassing 1.5 billon and profits reported by universal. In contrast, universal have demonstrated much more stable profits and average movie budget. It appears this production company continues to favour a high movie production volume to generate its revenue (Sheet 4).

Question 3:

This analysis explored both the profitability and budget of movies based on a novel relative to those that are not. A boxplot of the analysis of these features demonstrate that there is no discernible difference in terms of profit or budget when comparing movies based on a novel relative to those that are not. The time series analysis of these factors demonstrate, that movies not based on a novel tend to have a more stable profit and budget, while movies based on a novel appears to have peaks and troughs.

Question 4:

This analysis explored the relationship between continuous metrics and movie profit. In general, all metrics explored (runtime, release year, adjusted budget, vote average, vote count) had little bivariate explanatory power (R^2:0.02-0.38). Of the two performance metrics (vote average and count), vote count appears to be more closely related to the profitability of the movie yet could only explain 38% of profit variance. Budget had the second highest association with profit (R^2 0.16), which may explain why there appears to be a trend of greater movie budgets in recent years.