

Project: Visualizing Movie Data

A new movie production company is looking to make a new movie and wants to make sure it's successful to help make a name for the new company. They want to understand the movie trends to help inform their decision making.

Data Cleanup and Attribute Selection

id	imdb_id	popularity	budget	revenue	original_title	cast	homepage	director	
296100	tt3530002	2.096273	25000000	52395996	The Night Before	Joseph Gordon-Levitt; Seth Rogen; Anthony Ma...	https://www.facebook.com/NightBeforeMovie	Jonathan Levine	Their past, p...
263472	tt3605418	2.080935	2500000	5214043	Knock Knock	Keanu Reeves; Lorenza Izzo; Ana de Armas; Aaro...	http://knockknockmovie.tumblr.com/	Eli Roth	One night ca...
228205	tt2726560	2.070708	34000000	63013281	The Longest Ride	Scott Eastwood; Britt Robertson; Lolita Davidov...	[null]	George Tillman, Jr.	Two couples
159824	tt2510894	2.060162	80000000	466183544	Hotel Transylvania 2	Adam Sandler; Andy Samberg; Selena Gomez; K...	[null]	Genndy Tartakovsky	They're back
333348	tt2279339	2.055572	17000000	24000000	Love the Coopers	Steve Martin; Diane Keaton; John Goodman; Ala...	http://www.zbsfilms.com/love-the-coopers/	Jessie Nelson	Christmas m...
285783	tt3488710	2.049886	35000000	61181942	The Walk	Joseph Gordon-Levitt; Ben Kingsley; Charlotte L...	http://www.sonypictures.com/movies/thewalk/	Robert Zemeckis	Dream High...
290764	tt2401097	2.026651	11000000	1863315	Tracers	Marie Avgeropoulos; Taylor Lautner; Adam Rayn...	http://www.unlimited-movie.com/	Daniel Benmayer	It's not a crin
296099	tt1524930	2.000338	31000000	104394188	Vacation	Ed Helms; Christina Applegate; Skyler Gisondo; S...	[null]	John Francis Daley; Jonathan M. Go...	What could g
243688	tt1029360	1.964529	35000000	95437994	Polytechnic	Sam Rockwell; Rosemarie DeWitt; Saxon Sharb...	[null]	Gil Kenan	They're Here
287424	tt1881002	1.980039	4000000	187112	Maggie	Arnold Schwarzenegger; Abigail Breslin; Joely R...	http://www.maggiethesfilm.com/	Henry Hobson	Don't Get Bi...
301875	tt3289728	1.959765	0	2000000	Equals	Kristen Stewart; Nicholas Hoult; Guy Pearce; Jack...	http://equals-the-movie.com/	Drake Doremus	Find your eq...
181283	tt1014763	1.95632	50000000	3324330	Child 44	Tom Hardy; Gary Oldman; Noomi Rapace; Joel Kl...	[null]	Daniel Espinosa	How do you
328425	tt4178092	1.926739	5000000	58978653	The Gift	Jason Bateman; Rebecca Hall; Joel Edgerton; Dav...	https://www.facebook.com/GIFTmovie	Joel Edgerton	Not every gi...
274167	tt1528854	1.904898	50000000	240360392	Daddy's Home	Will Ferrell; Mark Wahlberg; Linda Cardellini; Tho...	http://daddyshomemovie.tumblr.com/	Sean Anders	Choose your
370687	tt3608646	1.876037	0	0	Mythica: The Necromancer	Melanie Stone; Adam Johnson; Kevin Sorbo; Nic...	http://www.mythicamovie.com/#?blank;ysake	A. Todd Smith	[null]

Figure 1 The Movie Database (TMDb) dataset

Following are the steps taken to clean and format the dataset:

- Convert data field type to the correct ones using 'Auto Field'.
- When the data are sorted by popularity column, majority of movies with 0 budget or revenue are poorly ranked thus in order not to skew the data will be removed. Using 'Filter' function to create a custom filter to remove movies with 0 budget or revenue. That leaves us with 3855 rows of data out of 10866 rows.
 - [budget] = 0 OR [revenue] = 0
- Using 'Select' function to select and rename the following fields:
 - Popularity
 - Original_title renamed to title
 - Keywords
 - Genres
 - Production_companies
 - Release_year
 - Budget_adj renamed to budget
 - Revenue_adj renamed to revenue
- Use formula tool to create 3 new columns:

Column Name	Reason	Formula
Is_Novel	To create a column to indicate whether the movie is based on a novel or not.	IF Contains([keywords], "based on novel") THEN "Yes" ELSE "No" ENDIF

Profit	To create a new column by calculating the profit using the revenue and budget	[revenue] - [budget]
Production_companies	To create a new column to indicate whether the production companies is either Universal Picture or Paramount Pictures.	IF Contains([production_companies],"Universal Pictures") THEN "Universal Pictures" ELSEIF Contains([production_companies],"Universal Studios") THEN "Universal Pictures" ELSEIF Contains([production_companies],"Paramount Pictures") THEN "Paramount Pictures" ELSE [production_companies] ENDIF

5. Split genre column into rows using the 'Text to Column' tool.
6. Using 'Select' function to select the following attributes to be explore:
 - o Popularity
 - o Title
 - o Genres
 - o Release_year
 - o Revenue
 - o Budget
 - o Is_Novel
 - o Production_companies
7. Exported the cleaned movie data using the 'Output Data' function.

popularity	title	genres	release_year	budget	revenue	Is_Novel	production_companies
32.985763	Jurassic World	Action	2015	137999939.280027	1392445892.52381	No	Universal Pictures
32.985763	Jurassic World	Adventure	2015	137999939.280027	1392445892.52381	No	Universal Pictures
32.985763	Jurassic World	Science Fiction	2015	137999939.280027	1392445892.52381	No	Universal Pictures
32.985763	Jurassic World	Thriller	2015	137999939.280027	1392445892.52381	No	Universal Pictures
28.419936	Mad Max: Fury Road	Action	2015	137999939.280027	348161292.489031	No	Village Roadshow Pictures Kennedy Miller Pro...
28.419936	Mad Max: Fury Road	Adventure	2015	137999939.280027	348161292.489031	No	Village Roadshow Pictures Kennedy Miller Pro...
28.419936	Mad Max: Fury Road	Science Fiction	2015	137999939.280027	348161292.489031	No	Village Roadshow Pictures Kennedy Miller Pro...
28.419936	Mad Max: Fury Road	Thriller	2015	137999939.280027	348161292.489031	No	Village Roadshow Pictures Kennedy Miller Pro...
13.112507	Insurgent	Adventure	2015	101199955.47202	271619025.407629	Yes	Summit Entertainment Mandeville Films Red W...
13.112507	Insurgent	Science Fiction	2015	101199955.47202	271619025.407629	Yes	Summit Entertainment Mandeville Films Red W...
13.112507	Insurgent	Thriller	2015	101199955.47202	271619025.407629	Yes	Summit Entertainment Mandeville Films Red W...

Figure 2 Cleaned Movie Data Dataset

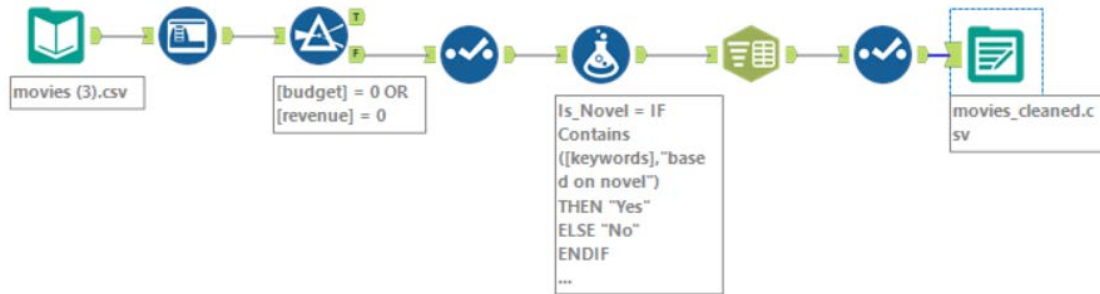


Figure 3 Alteryx Workflow

Tableau Visualizations and Questions

Public Profile Link -

https://public.tableau.com/profile/emily6902#!/vizhome/BAND_P5_Visualizing_Movie_Data/Q4?publish=yes

- **Question 1:** How have movie genres changed over time?

Q1 Movies genres changed over time

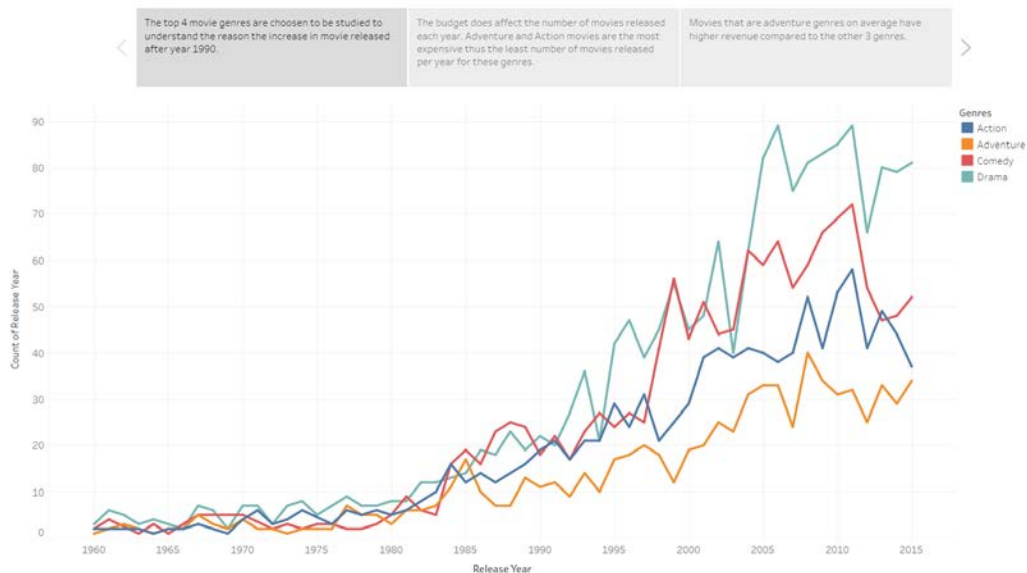


Figure 4 Movies genres changed over time by count

- The top 4 movies genres - Action, Adventure, Comedy and Drama are filtered out in order to study and understand the movie released. There is an increase in movies released after year 1990. The top goes to the genre Drama followed by Comedy, Action and Adventure.

Q1 Movies genres changed over time

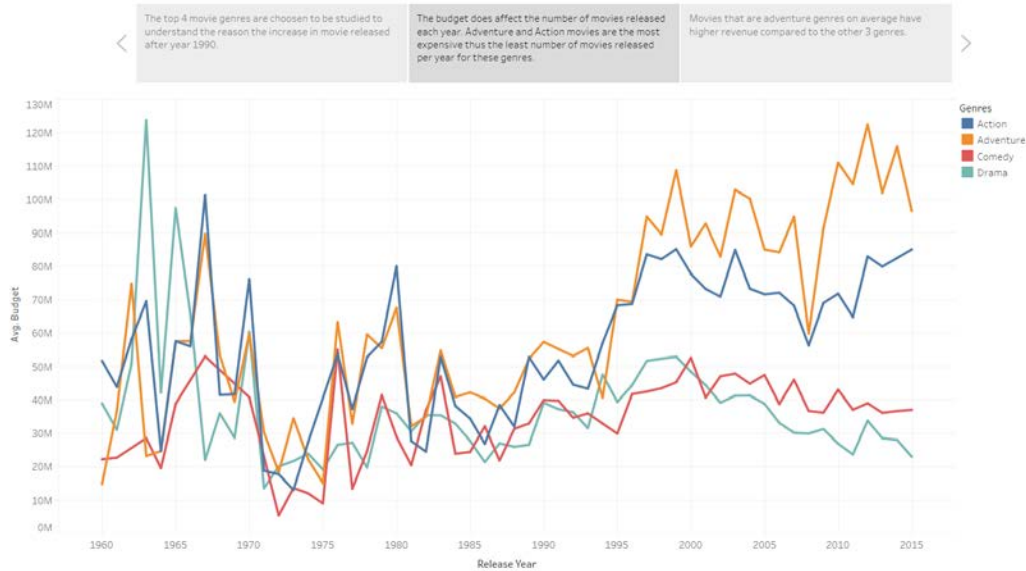


Figure 5 Movies genres changed over time by average budget

- The budget does affect the number of movies released each year as we can see that the genre Adventure and Action movies have higher budget which could be the reason why least number of movies released for that genre.

Q1 Movies genres changed over time

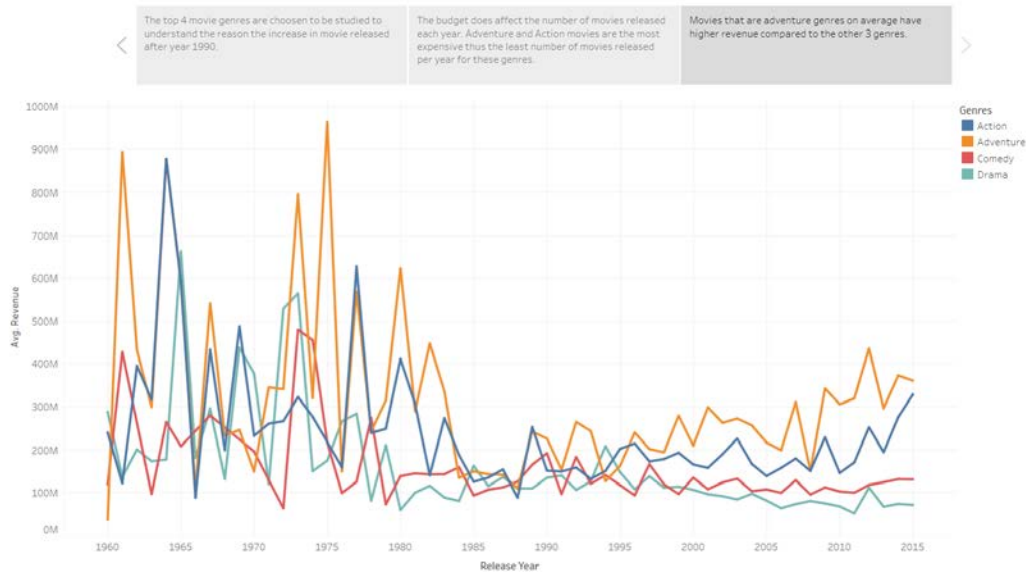


Figure 6 Movies genres changed over time by revenue

- Revenue of the movies are affected by the budget of the movies. Higher budget movies typically also enjoy higher revenue.

- **Question 2:** How do the attributes differ between Universal Pictures and Paramount Pictures?

Q2 Attribute differs between Paramount and Universal Pictures

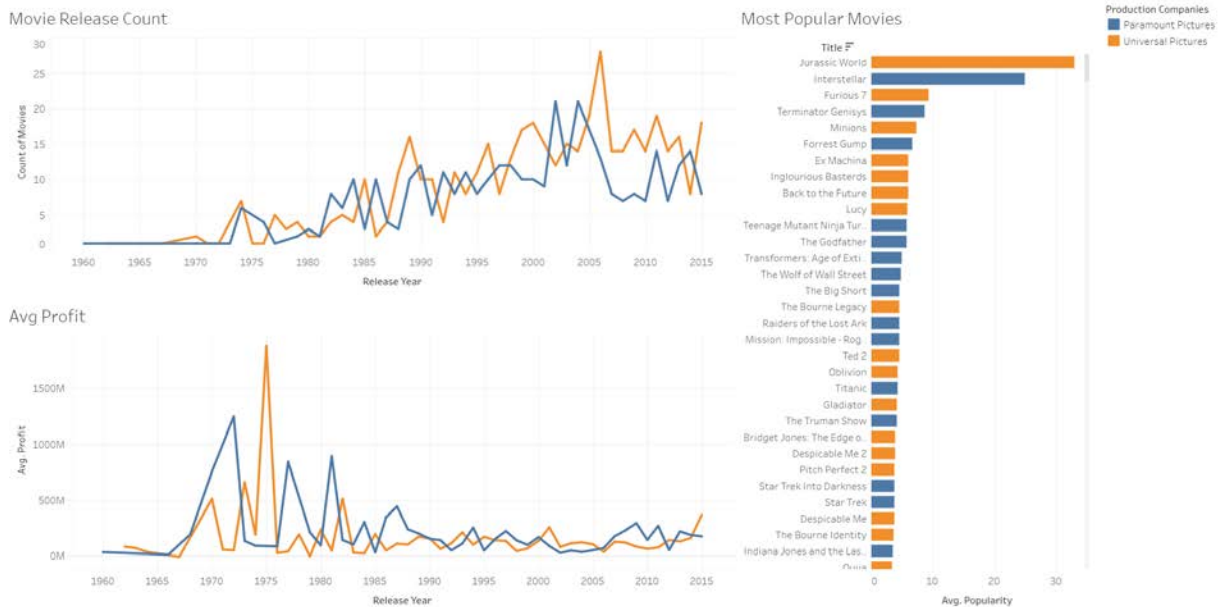


Figure 7 Attribute differs between Paramount and Universal Pictures

- Through the years, Universal Pictures have released more movies compared to Paramount Pictures. The highest was more than 25 movies.
- In Year 1975, Universal Pictures has a higher average profit compared to Paramount Picture. However, through the years both actually have quite similar average profit.
- Looking at the list of most popular movies, most of the top movies are from Universal Pictures. It is not known whether Universal Pictures's high popularity and revenue are due to its higher number of movie release, Universal Pictures is definitely performing better than Paramount Pictures.

- **Question 3:** How have movies based on novels performed relative to movies not based on novels?

Q3 Movies based on novels performed relative to movies not based on novels

Movies Performance (Novel vs Non-Novel)

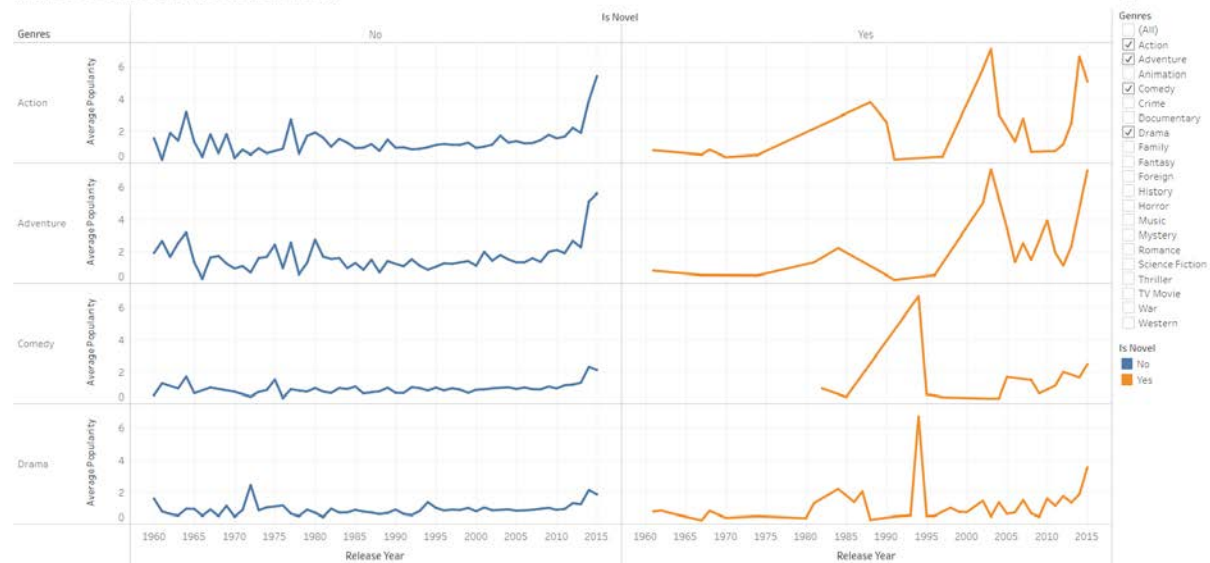


Figure 8 Novel based movie performance

- Through the years, movies based on novel are more popular compared to non-novel based movies.
- For the top 4 genre – Action, Adventure, Comedy and Drama, the highest average popularity for novel based movies is 7.122 and the highest for non-novel is 5.610.
- When using the filter option to view all genre, we can see that there are very few movies with the genre TV movie (non-novel based) and Foreign (novel based).

- **Question 4:** What is the relationship between profit and movie budget?

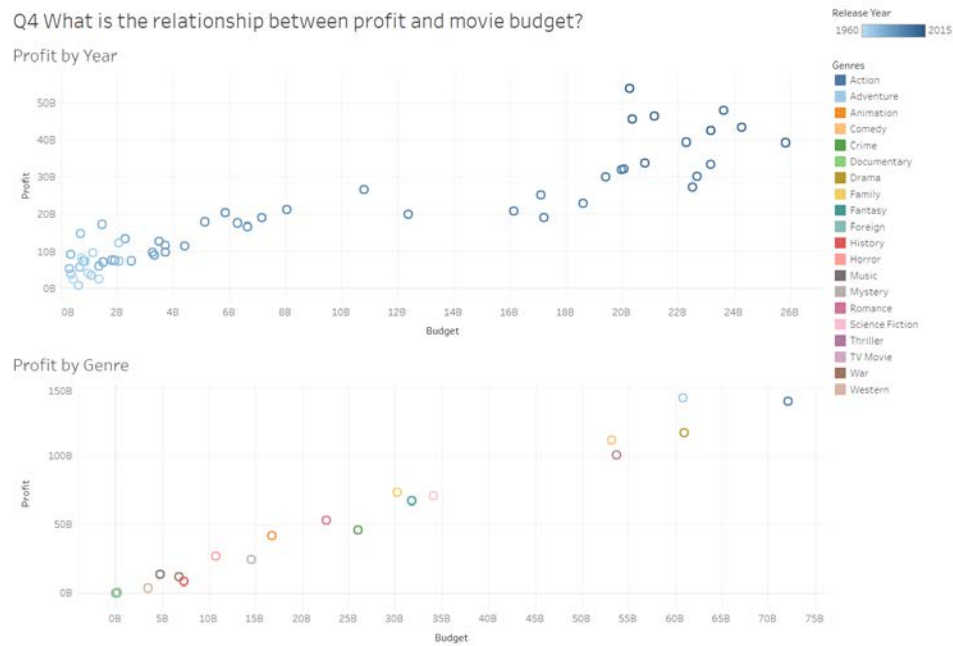


Figure 9 Relationship between movie profit and budget

- We can see that there is a relation between movie budget and profit. As the budget goes higher, the profit goes up as well.
- The top 5 movies genres with high profit and budget are Action, Adventure, Drama, Comedy and Thriller.