

Project: Visualizing Movie Data

Visualization Link: <https://public.tableau.com/profile/luyan.deng#!/>

Step 1: Data Cleanup and Attribute Selection

- Clean up any missing information and choose the most important attributes you will explore further in your visualizations.

Taking an overview of the attributes, 'homepage' has over 7930 null value, so the field is deleted. Other fields with null values are much less the entire pool of data, therefore only rows with null values are removed.

In our project, evaluating the movie sales performance will be a main focus. Thereby, we remove the rows that budget amount or revenue is zero.

After brief cleaning, we have 3680 valid rows from 10680 rows of the original dataset.

- List out the attributes (or variables) you plan to dive further with your visualizations. You should explore no more than 8 attributes.

The following attributes will be selected for further visualization

- original_title
- Popularity
- Budget_adj
- Revenue_adj
- Production company
- Keywords
- Genres
- Release_year

Popularity, budget and revenue represent the business performance. The movie title, genres and other variables will add more dimensions to find the hidden trend behind data.

Step 2: Tableau Visualizations

- Please make sure you follow the [rubric](#) and include Tableau Dashboards, Stories, and the appropriate visualizations (small multiples, scatter plot, bar chart, etc..) your reviewer expects your visualizations to contain. Remember: You need one Dashboard for every question (Q1-Q4) and in addition, you also need one Story, pertaining to a question of your choosing.

- Attach your visualizations as Tableau Workbooks in a zip file along with this report.

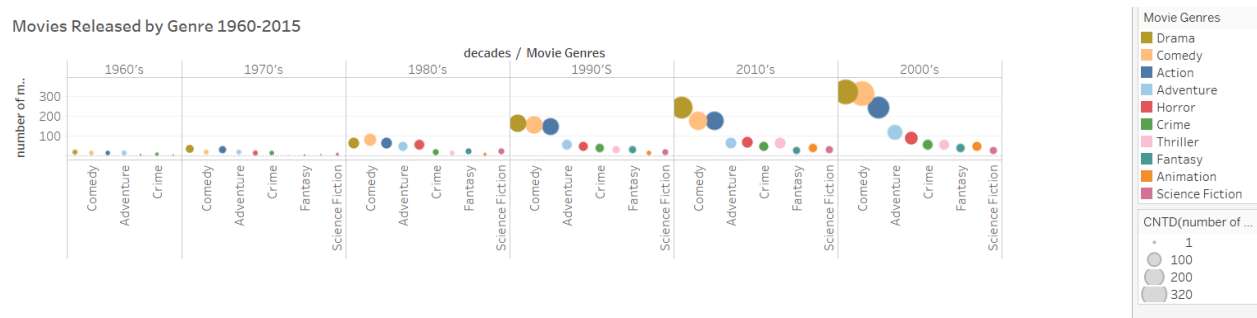
IMPORTANT: Please upload the workbooks to **Tableau Public** to allow reviewers to access your workbooks. Note that simply saving your file as a “.twbx” is not enough to allow all reviewers to access. [Instructions on how to do this.](#)

Question 1: How have movie genres changed over time?

From 1960 to 1980, the genres are limited mostly in adventure, drama. After 1990, both the numbers of movies and genres have increased dramatically.

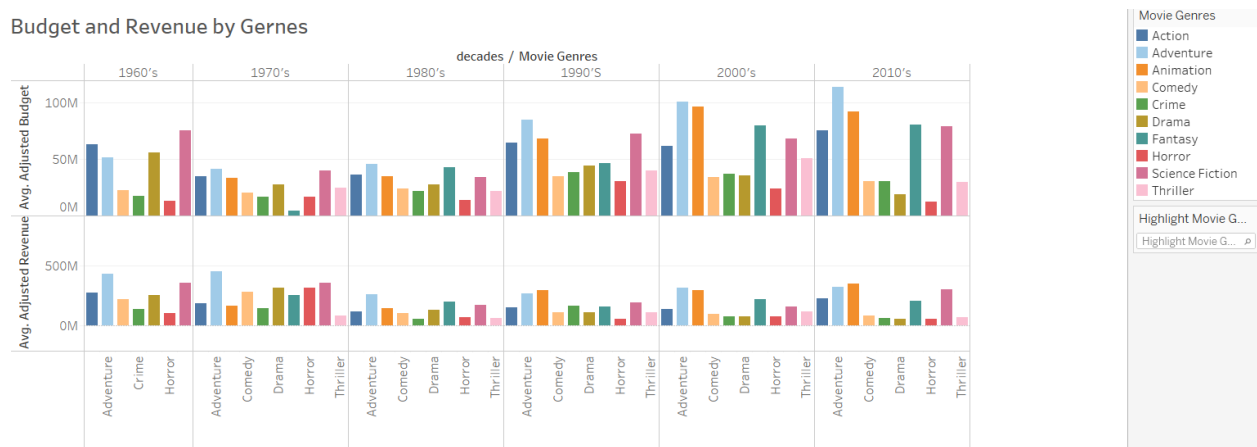
Over the years, the top 5 most popular and fastest growing genres are drama, comedy, action, adventure and horror.

Movies Released by Genre 1960-2015

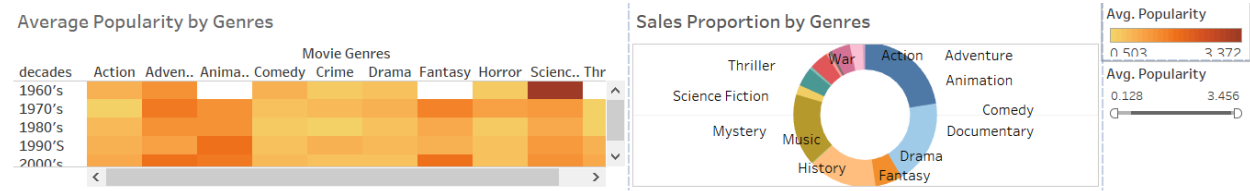


As to budgets, adventure movies attract higher average budget over the year followed by action movies. After 1990's, animation movies drew a higher average budget followed by science fiction and fantasy. Despite the high number of drama movies in recent years, the average budge of drama genre drop after 1990's.

Budget and Revenue by Gernes



The average budget of movies highly related to the average of revenue. Movie genres that have higher budgets often generate higher revenue. Adventure movies are one of the best-selling genres. But we can see that after 1990, new genres like animation and fantasy became new generators of revenue.



If we take a look at the popularity changes over the years, we find that the most popular genres attract more budgets and more successful in sales. In all, adventure, action, comedy and drama make up the main part of revenues. Although drama has lower revenue in average but the high number of release ensure an important proportion in the market.

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

Data Preparation

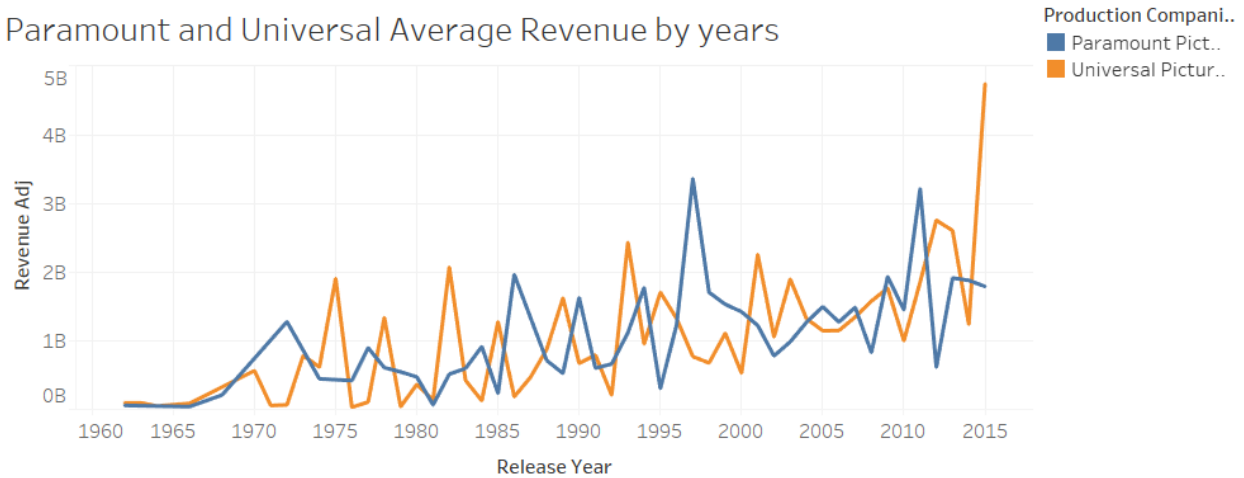
Reviewing the 'production_companies' field, we find that there are more than one string value indicate "Universal Pictures" or "Paramount Pictures".

"Universal", "Universal Pictures Corporation", "Universal Studios", "Universal TV" are all brands of Universal. Thus they are replaced by "Universal Pictures". And "Paramount Classics", "Paramount Vantage" are replaced by Paramount Pictures".

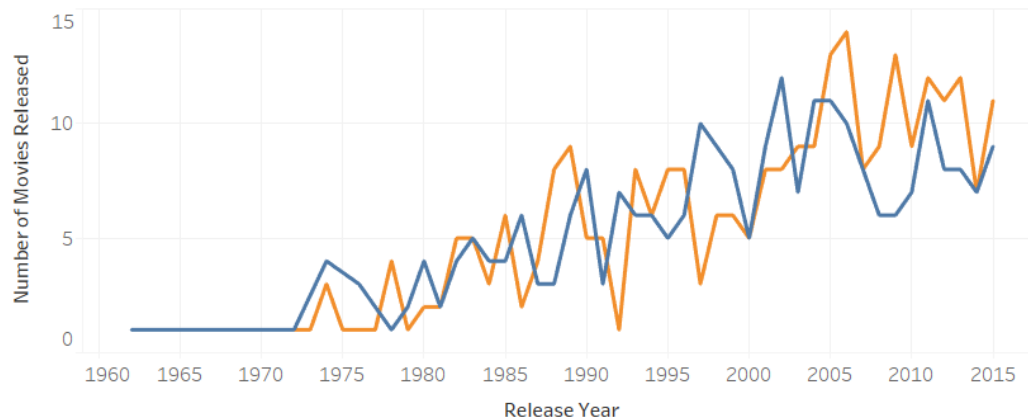
Data Visualization

Over the years, Universal Pictures(UP) generated higher revenue than Paramount Pictures(PP). Especially in 2015, UP reached an high of 4.8 billion revenue, 2 times higher than that of PP. UP also produced more movies overall, despite the fact that Paramount produced and earned more from 1995 to 2000.

Paramount and Universal Average Revenue by years



Number of Movies Released by Paramount and Universal



Take a look at both companies' budget plans, they have similar focus on genres. Adventure, action, animation, fantasy, and Science Fiction allocated higher budgets.

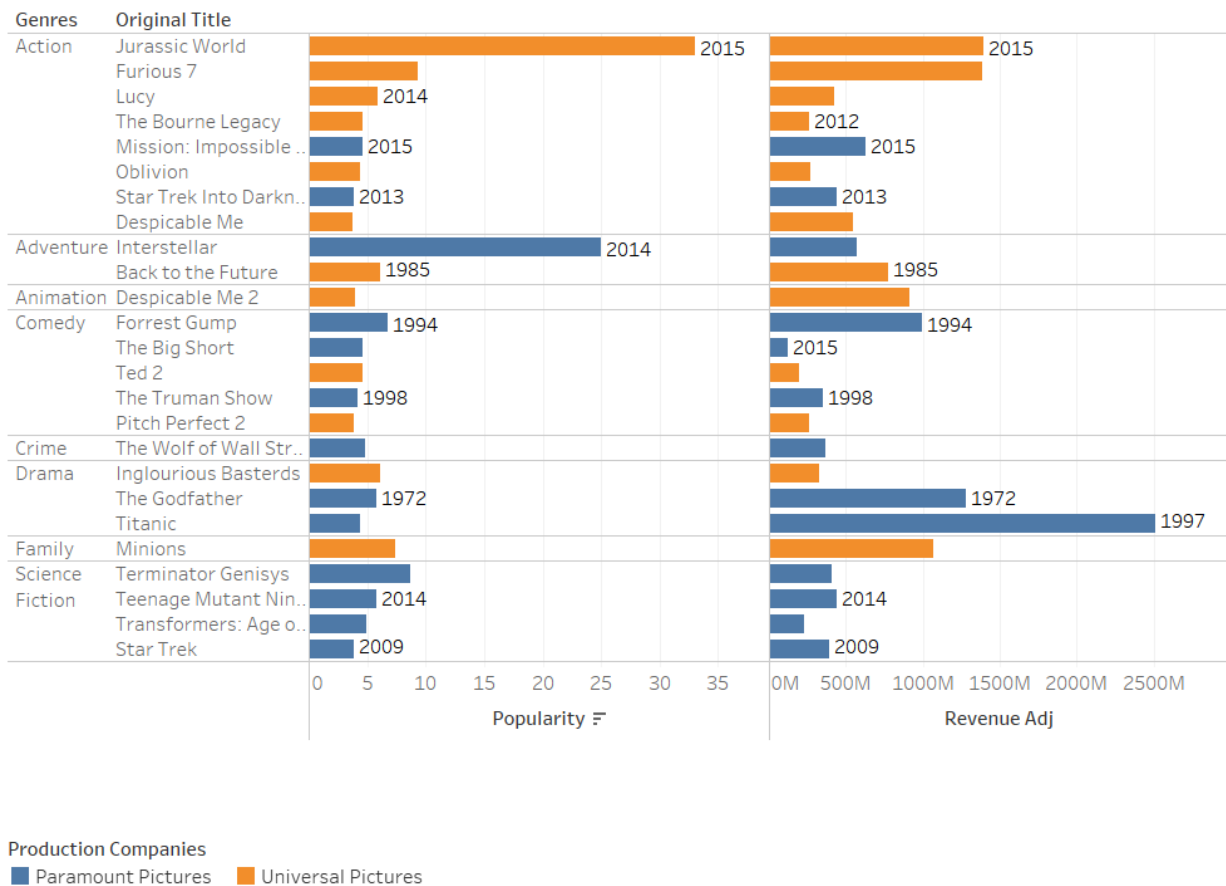
When comes to average revenue comparison, we find something interesting. UP generates most revenue on average in science fiction and animation, though these two budgets are less than PP.

Budget and Revenue Comparison



We can find out that a movie's popularity is highly related to the revenue. But a high popularity does not necessarily indicate a higher revenue. The most popular movies produced by Universal Pictures and Paramount Pictures are Jurassic World and Interstellar, respectively. In 2015, Universal Pictures' Jurassic World and Furious 7 became its best selling movies, generating almost 1500million dollars each. The movie that generated most revenue between these two companies are Paramount's Titanic in 1997, which partly explains its higher revenue than Universal Pictures in 1997.

Most Popular and Profitable Movies Produced by Paramount and Universal



In sum, Universal and Paramount are the top 2 production companies in producing movies and generating revenues. They both place high budgets on action, adventure, science fictions, animation, fantasy and so on. From there performance, we can see that action and adventure genre are more likely to build a brand name and in recent years, action genre are best selling genres.

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

Data Preparation

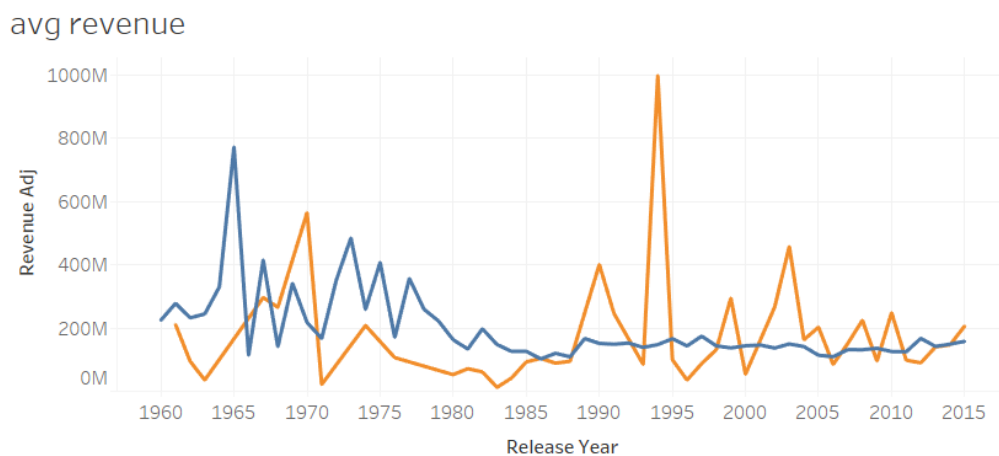
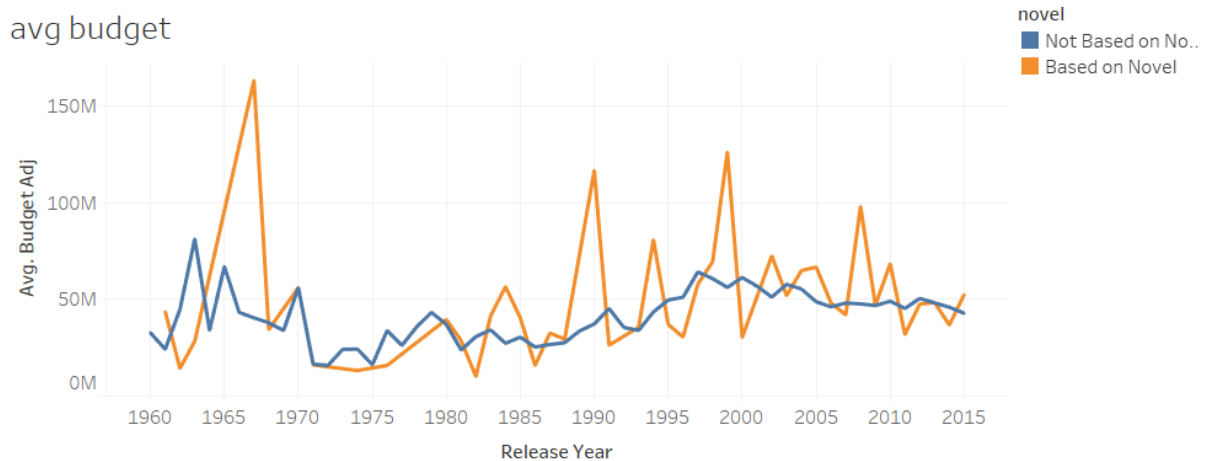
To divide the movies in to novel or not novel based, we check if the “keywords” field has the word “novel”. If “novel” is not exist in the keywords field, we label the movie as “Not Based on Novel”.

However, this group method is biased as some of the movies that based on novel have no “based on novel” strings in the “keywords” field. For instance, Jurassic World, Harry Potter Series are produced based on novels. As their keywords fields contains no relevant information, the program labels these movies as “Not Based on Novel”, which is not true.

To decrease the bias of the data, we use average variables to compare these two groups of movies. However, without additional data, the result is off the actual records.

Data Visualization

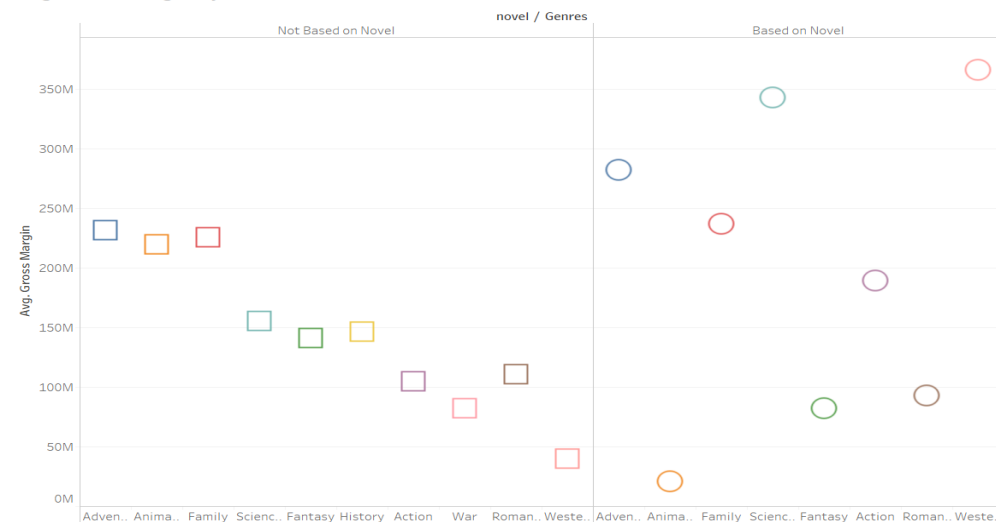
Over the years, novel-based movies seem to collect higher budget on average. Before 1990, not-novel-based movies generate higher revenues on average, but Novel-based movies create more revenues after 1990.



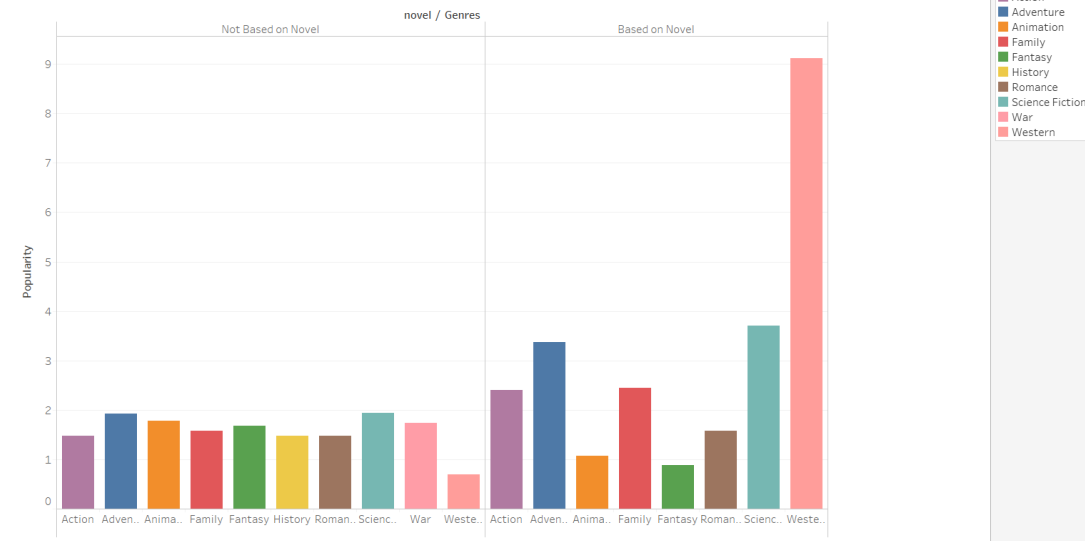
Amongst the top 10 most popular genres, novel-based movies are more in action, adventure, family, science fiction and western genres. Specially, in western genre, the novel-based movies are more popular than not-novel based.

The average popularity usually highly relates to profits. However, we find that western is not in the top 10 most profitable genres.

Avg Gross Margin by Genres



Avg Popularity by Genres

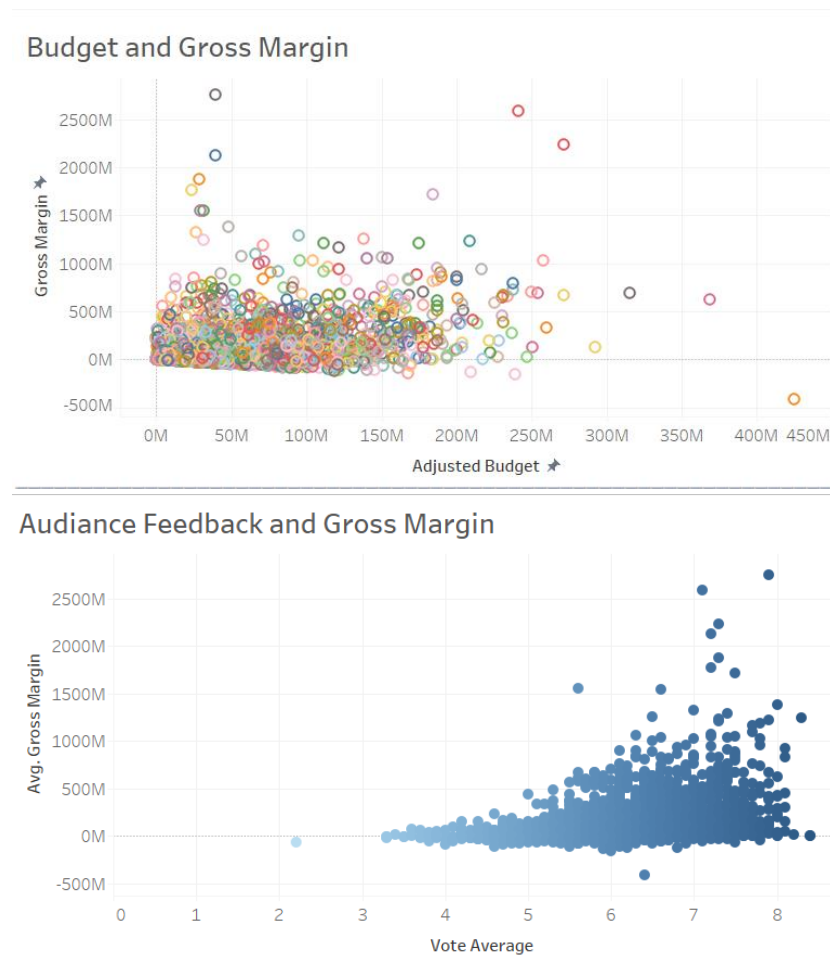


In short, we can find out that in the best selling genres: adventure, science fiction, action, if a movie is produced based on a novel, it tends to have a higher popularity and revenue.

Q4: What makes a profitable and popular movie.

Since we are a new movie production company looking to make a new movie. So it is important to understand the trend, return on investment and the attributes of a best selling movie.

We can see that the higher budget highly relates to a higher gross margin. Adventure and action movies allocate more budgets but generate higher gross margin. Most successful adventure and action movies are based on novels or comic books. Also, the quality of a movie is highly related to its gross margin. The averaged vote represents the quality of a movie. The higher vote indicates higher gross margin.



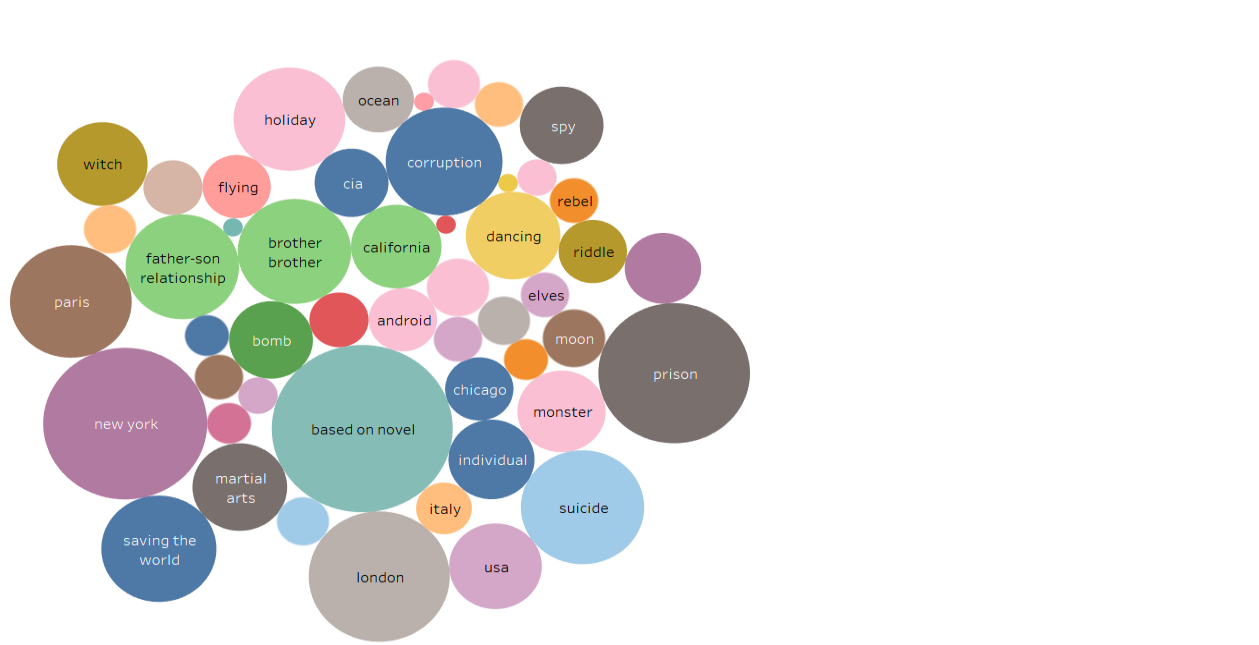
So how can we ensure the quality of a movie ? A reputable and excellent director can determine the success of a movie. From the director's portfolios ,we can find that a list of the profitable movies created.

Director's Most Profitable Movies

| Movie Director | Original Title |
|----------------------|-------------------------|
| William Friedkin | The Exorcist |
| James Cameron | Avatar |
| | Titanic |
| | Terminator 2: Judg.. |
| J.J. Abrams | Star Wars: The Forc.. |
| Francis Ford Coppola | The Godfather |
| George Lucas | Star Wars |
| | Star Wars: Episode I.. |
| | Star Wars: Episode I.. |
| | Star Wars: Episode I.. |
| | American Graffiti |
| Joss Whedon | The Avengers |
| | Avengers: Age of Ult.. |
| Steven Spielberg | Jaws |
| | E.T. the Extra-Terres.. |
| | Jurassic Park |
| | Close Encounters of .. |
| | Raiders of the Lost A.. |
| | Indiana Jones and th.. |
| | Indiana Jones and th.. |
| | Indiana Jones and th.. |
| | Saving Private Ryan |
| | War of the Worlds |
| | Schindler's List |

Over the years, the top 1 best selling movie is Avatar. If we take a look at the keywords of the movies, we can find that a novel-based movie is more profitable. The topics are about supper heros ('saving the world', 'martial arts'), magic ('witch', 'flying'), crime ('prison', 'corruption').

Keywords Cloud



Recommendation:

Based on the above analysis, if the budget is allowed, our best plan is to invest on a adventure or action movie based on novel. Supper heros, witches, monsters, robots are some of the most popular elements.