

Project 3: Visualizing Movie Data

By: Huong Ivy Nguyen

Step 1: Data Cleanup and Attribute Selection

- After plotting the popularity by genres, I noticed there is a 'null' category. So I created a new set of genres which exclude the 'null' category out of my bar plot.
- The attributes this report used to make the analysis include: Budget Adj, Revenue Adj, Popularity, Vote Average, Runtime, and a new calculated field Profit which is determined by the difference between Budget Adj and Revenue Adj.

Step 2: Tableau Visualizations

Question 1:
https://public.tableau.com/profile/ivy.nguyen#!/vizhome/Project3_Business_Analyst_Udacity/Q1

Question 2:
https://public.tableau.com/profile/ivy.nguyen#!/vizhome/Project3_Business_Analyst_Udacity/Q2

Question 3:
https://public.tableau.com/profile/ivy.nguyen#!/vizhome/Project3_Business_Analyst_Udacity/Q3

Question 4: each category of the top 5 most popular genres, which is the most profitable movie?
https://public.tableau.com/profile/ivy.nguyen#!/vizhome/Project3_Business_Analyst_Udacity/Q4

Step 3: Questions

- Answer the following questions. Refer to your online visualizations to back up your answers:
 - **Question 1:** How have movie genres changed over time?
In order to investigate this question, I first plotted the popularity of all genres (without null) to see what are the most popular categories over the years. I noticed Action, Adventure, Drama, Comedy and Horror made the top 5 most popular among all the genres. I then shifted my focus to these 5 categories by plotting the number of releases for each over the years. I also plotted their popularity trend over the years. It appears that Drama has the highest number of releases while Action gains the most popularity in 2015.

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

I chose to compare the following attributes: the number of releases, budget adj, revenue adj, and popularity between Universal Pictures (UP) and Paramount Pictures (PP). I plotted these attributes over the release year to make the comparison. There is no set trend for which company is more dominant over the years since the data seems to fluctuate based on how company invested in their movie production. However, as we look at an individual year, it becomes very interesting how these attributes link to each other. For example, in 2015, UP spent a lot more money in their movie production than PP did, which could potentially help them gain more popularity and also make a higher revenue than PP did.

Surprisingly, when I plotted the popularity among movies that are made by these two companies. It appears that PP's movies are more popular than UP's movies. Specifically, Interstellar made the top of the list and was produced by PP.

- **Question 3:** How have movies based on novels performed relative to movies not based on novels?
 - In terms of popularity, novel-based movies are not as popular as non-novel-based movies.
 - Among the top 10 most popular movies, only one novel-based movie made the list.
 - Novel-based movies seem to have higher average adjusted budget and also higher average adjusted revenue than non-novel-based movies do.

- What is your additional question that you proposed? What is the answer? How did you come up with this question?

Question 4a: How is runtime related to profit and vote average? Do these attributes have a linear relationship?

Vote average and runtime appear to have a very strong linear relationship. Profit and runtime also has a positive linear relationship but not as strong of a relationship as vote average and runtime have.

Question 4b: In addition, for each category of the top 5 most popular genres, which is the most profitable movie?

The most profitable movie for each category of the top 5 most popular genres are:

Action: Avatar

Adventure: Star Wars

Comedy: Home Alone

Drama: Titanic

Horror: Jaws

I came up with this question just based on my curiosity. I was wondering if a production company makes a movie with a longer runtime, could that potentially help increasing the chance for that movie to get a higher vote average and the company to earn a higher profit for their movie production.

The second part of this question is additional to answer which movie makes the highest profit for each category in the top 5 most popular genres. It would be interesting to further extract the runtime, average vote, and profit of these movies to determine whether they indeed have a high profit with a long runtime or high average vote with a long runtime. This fact could potentially help validating the stated linear relationships in part a.