Project: Visualizing Movie Data

Step 1: Data Cleanup and Attribute Selection

Movies.csv contains 21 variables. I chose following variables to explore.
1. original_title
2. keywords
3. genres
4. production_companies
5. release_date
6. bote_average
7. budget_adj
8. revenue_adj

The following data are missing
1. 9% of production_companies
2. 0.2 % of genres
3. 14 % of keywords
4. 52 % of budget_adj value are zero.
5. 55 % of revenue_adj value are zero.

I decided to drop all missing values from first three variables. For budget_adj and revenue_adj, I checked the distribution of data. The distribution of both budget_adj and revenue_adj was right skewed and both variables had too many negative outliers. So, I decided to impute 0s with median values.

Step 2: Tableau Visualizations

Link to My visualization on tableau public

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?
  - Within the span of 55 years, from 1960 to 2015, movies genres had changed a lot. Action and adventures movies positioned mostly in top 5 of ranking. It looks like these two genres are all time favorite. The rank of profit made by science fiction is 3 in 2015, but back in 1960, its rank is 10. For history movies, the rank is 3 in 1960, but its rank declined constantly to 18 in 2015. Total numbers of movies made is increased greatly in these 55 years.

- ○ **Question 2:** How do the attributes differ between Universal Pictures and Paramount Pictures?
- ○ Paramount Pictures made more movies, more gross profit, and more average profit than Universal Studios. And movies made by Paramount Pictures mostly got more average vote than that of Universal Studios. But in 2015, the average profit of Universal Studios is about 5 times higher than that of Paramount Pictures.
- ○ **Question 3:** How have movies based on novels performed relative to movies not based on novels?
- ○ Movies based on novels usually make more profit and usually get higher vote than movies that do not based on novels.

- What is your additional question that you proposed? What is the answer? How did you come up with this question?
  Q. How have movie genres that based on novels changed over time?

  - The numbers of movies based on novels in each genre grow obviously after 2004.
  - Drama, thriller, science fiction, comedy and action movies, among movies based on novels, became more and more popular and made more profit over time.
  Drama, Thriller, and Romance movies mostly based on novels.

  - From 1960 to 2015, Drama movies made most gross profit than other genres. Movies based on novels got more revenues than those that didn't. And they also got higher average vote. Some questions came into my mind. How have genres of movies based on novels change over time? Did they grow or shrink, and to what percentage? Which genres of movies usually based on novels? What is the average profit of each movie genres based on novels?