Project: Visualizing Movie Data

Shahrooz Govahi

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.
  Answer: Total number of records in the data set is 10866. There are many data points in "budget" or "revenue" fields with 0 value. I think these two fields are very important because base on these two fields we can calculate the profit and base on the profit we can have a real understanding about a movie success. So I decided to clean up the records with 0 value in at least one of these two fields. There are other fields with NULL value records like "homepage" variable. These are not very important fields in my analysis. So it's not necessary to eliminate these records from the data set.
  For making sure that cleaning up the records with 0 value in "budget" or "revenue" doesn't affect strongly our analysis, here I will look at two different measures statistics that shows another aspect of a movie success. The first one is "popularity" and the other is "vote_count".

Filed	Min	Max	Median	Mean	Std
popularity_Cleaned_Data	0.001	32.99	0.80	1.19	1.48
popularity_Deleted_Records	0	8.41	0.28	0.35	0.32
vote_count_ Cleaned_Data	10	9767	204	528	880
vote_count_ Deleted_Records	10	1329	22	47	77

  Based on these statistics, it seems the records with 0 value in "budget" or "revenue" are not so popular compare to movies that have a non-zero value for these two fields. And popularity is a very important aspect in analyzing movies. So I will work with 3855 cleaned records.

- List out the attributes (or variables) you plan to dive further with your visualizations. You should explore no more than 8 attributes.
  Answer: Between 21 fields available in the data set, I will focus on these attributes: keywords, genres, production_companies, vote_average, release_year, budget_adj, revenue_adj, original_title.

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.

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?
      Answer: Number of movies in different genres are changed over time. After year 2000 we can see these genres are the most produced genres:
      Drama, Comedy, Thriller, Action, Adventure.
      (see Q1 Dashboard, Chart: Number of Movies in each Genre)

      For answering this question I consider all genres for each movie and when I mention those top five genres, it means that these genres are most repeated among all genres.
      However if we look at profit for each genre the top five genres will change to: **Action, Adventure, Comedy, Science Fiction, Drama**;
      (see Q1 Dashboard, Chart: Total Profit for all Movies in a Genre($))

      On the other hand, genres best votes are ranked in this order for 2016:
      Animation, Romance, Drama, Adventure, Action.
      (see Q1 Dashboard, Chart: Average Vote for all Movies in a Genre)

      Note: In my profit calculations, if animation was a genre among other genres, I put animation share twice as others. For other cases I consider all contributed genres in a movie with equal shares in profit.

    - **Question 2:** How do the attributes differ between Universal Pictures and Paramount Pictures?
      Answer: In general number of movies in each genre that "Universal Pictures" produced is more than movies "Paramount Pictures" produced in the same genre. There is only one exception and that is "Adventure" genre that Paramount pictures produced more.
      (See Q2 Dashboard, Chart: Number of Genres Production)

      After 2004 number of movies Universal pictures produced are more than movies Paramount pictures produced. Before that there is not a clear trend. In some

years Paramount produced more and in other years Universal produced more movies.
(See Q2 Dashboard, Chart: Number of Movies each Company Produced)

Profit wise, after 2000 the profit of movies that Universal Pictures is part of producers is more than movies that Paramount Pictures contributed as a producer. The only exception in 2011 for this general rule.
(See Q2 Dashboard, Chart: Shared Profit for each Company)

- **Question 3:** How have movies based on novels performed relative to movies not based on novels?
  Answer: For answering this question first I created ScreenPlay category with two values: “Based On Novel” and “Not Based On Novel”. I calculated the average profit for each category in different years. In majority of year “Based On Novel” movies produce more profit on average rather than “Not Based On Novel”. But there are some exceptions in some years.
  One important year for movies “Based On Novel” is 2003. In that year “Lord of the Rings” produced a huge profit and that’s the reason there is a peak in that year on Average profit.
  (See Q3 Dashboard, Chart: Avg. Profit regards to Screen Play Source)

  Vote average for these two categories shows an interesting point. Movies “Based On Novel” has very fluctuated graph. It seems like that these kind of movies are very polarized. The rage of average vote in this category is: 5.77 to 7.017
  On the other hand this range for “Not Based On Novel” is: 5.95 to 6.31; It shows more predictable results for this category.
  (See Q3 Dashboard, Chart: Avg. Vote regards Screen Play Source)

- What is your additional question that you proposed? What is the answer? How did you come up with this question?
  For a new questions I am thinking about attributes related to successful movies. If I am working as a Business Analyst for a Movie production company, one of the most important things for them should be understanding the variables behind a profitable and popular movie and the best way for considering this topic is looking back at best previous movies and analyze them base on the profit they produce and votes they got. Based on these factors we can find out which kind of movie we should produce for maximum profit and popularity.

  Here is the question:
  - **Question 4**: What are the well performed movies based on their profit and people’s vote? What genres are the most successful?
    Answer: Between 50 top profitable movies in different years, 28 of them has “Adventure” and “Action” main genre. By main genre I mean the genre that mentioned at first of genre attribute in the data set.
    Top five profitable movies are:

Star Wars(1977), Avatar(2009), Titanic(1997), Exorcist(1973), Jaws(1975);
It is interesting that 4 of these movies are produced before 2000.
There are very few movies in top profitable ones in “Crime” or “Horror” or “Mystery” genres.
(See Q4 Dashboard, Chart: Most Successful Movies)

Next I focused on people’s vote for each movie in different years. It shows that there is not a clear correlation between top profitable movies and top voted ones. In top 20 votes there are 12 Drama movies and only 2 movies in “Adventure” or “Action” genres. Number of movies with less than $500M profit is 15 out of 20 top voted ones. 9 of these top 20 voted movies are produced in 90’s; These top voted movies have this range of vote: 8.00 to 8.40;
(See Q4 Dashboard, Chart: High Vote Movies)

Link to the Tableau Public Workbooks:

https://public.tableau.com/profile/shahrooz.govahi#!/