Project: Visualizing Movie Data

Complete each section. When you are ready, save your file as a PDF document and submit it here.

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.
- List out the attributes (or variables) you plan to dive further with your visualizations. You should explore no more than 8 attributes.
- Please refer back to the Data Cleanup course to help you clean up your data.

ANS – After analysing the questions that we have to answer we have finalized the following attributes. Any cleanup, if required is also mentioned

Id field is used as an identifier (not a separate attribute)

1. Release_year – No clean up required, it remain as a numeric field and there are no blanks
2. Popularity - No clean up required, it remain as a numeric field and there are no blanks
3. Vote_average – No clean up required, it remain as a numeric field and there are no blanks
4. Revenue - Here we will leave out values which are under 10,000 because they are probably errors
5. Genres – This field will have to be split. There are 23 blanks which we will ignore because it is a very small number as compared to the overall population
6. Keywords – This field will have to be split. There are several blanks here , about 1496 of them. Since it is a significant gap we will filter these out in our calculation
7. Production_companies – This field will have to be split. There are 1030 blanks here which we will filter out.
8. Budget – Here we will leave out values which are under 10,000 because they are probably errors.

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?
Dashboard link - https://public.tableau.com/profile/vaibhav.rastogi#!/vizhome/Q1-Dashboard_2/Q1Dashboard?publish=yes

The dashboard show how movie genres have changed over time
In the top half section we see the growth of movies in particular genres by year. This shows a lot of growth in drama and comedy movies over the years, while somewhat lesser growth in action movies over the years. Many new genres have also been added over the years. Lastly in the last few years number of comedy movies have fallen a lot relatively and number of thriller movies so a considerable jump in the same period.

There are 3 bar charts in the bottom half section. These reflect the relative percentage shares of genres in three periods – 1961-65, 1986-90, 2011-15
In each of these graphs popularity has been used as a 3rd continuous dimension. The legend is placed on the rightmost column for each of the graphs.

- In 1961-65 period Drama movies had the highest share of 26.44% followed by Comedy and action genres at 19.54% and 12.07% respectively. Animation and adventure were the most popular genres
- In 1986-90 period Share of comedy and action movies jumps a lot to 24.85% and 17.58% respectively. Comedy movies have the highest share. Share of drama movies drops in this period to 18.64% . Animation and Science fiction were the most popular genres
- In 2011-15 period, share of drama movies is back at highest at 22.46% while share of comedy movies drops to16.21% and share of action movies drops to 11.07%. A lot more new genres also get introduced in this period such as Family movies, tv movies. Western and adventure were most popular genres

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

Note – this dashboard has the story as well

Universal pictures has produced 34 films more than paramount pictures has produced. Paramount pictures produces movies which are slightly more popular on an average than movies produced by universal pictures. This has been reflected in the legend.

Furthermore universal pictures has made more movies in genres such as comedy, action, adventure whereas paramount pictures has made more movies in genres such as science fiction, crime, thrillers.

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

Movies based on novels have higher popularity (approx. 0.5 points) and higher vote average (by approx. 0.32 points) than movies not based on novels. This can be seen in the dashboard as well . Movies based on novels on an average have earned greater revenue than movies which were not based on novels. This has been reflected by the legend shade as well.

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

The question is what sort of relation exists between budget of a film and the revenue it earns. I was intrigued to know if higher budget , on an average would lead to higher revenue.

It seems there is a strong positive linear correlation between budget and revenue. As can be seen in the dashboard the scatter plots clearly reflect this that as budget increases revenue increases somewhat proportionally. We have 2 scatter plots , one of which has additional dimension of genre and other has an additional dimension of production_company.

Both these plots have an additional continuous dimension of popularity also because in addition to this we also wanted to see how the popularity varies with budget and revenue for different genres and production companies. It seems that the distribution is fairly random, with some cases of low budget films being more popular than high budget films, while in some cases the opposite is true. This is reflected by the size legend on the left.