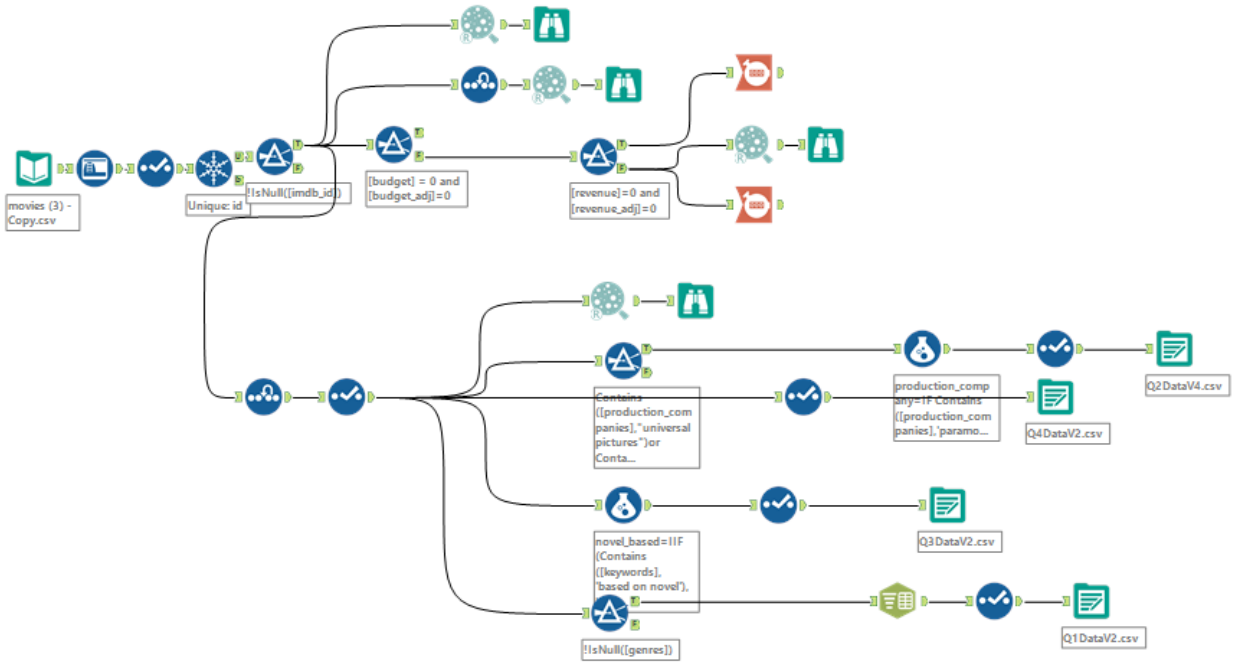
Project: Visualizing Movie Data

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.



Cleaned the data by removing any duplicate values then chose to impute the values instead of deleting them based on the scatterplots I made and compared. I ended up producing 4 outputs, one for each question as there are some questions that would require additional cleaning and splitting of the data that would not be of use to the other questions.

I selected the following attributes:

1. Budget Adj
2. Revenue Adj
3. Runtime
4. Vote Average
5. Vote Count
6. Popularity
7. Release Date
8. Number of Records

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.

LINK TO WORK BOOK :
https://public.tableau.com/views/Project3v9/Q4?:embed=y&:display_count=yes

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?

The visualization used for this question is found in story 'Q1 Story'. The first storypoint shows the various genres over time in terms of vote average, population and runtime. The rating shows that documentaries are the highest rated genre. While a decline can be seen in horror and science fiction. Some genres have an incline in popularity while others maintain a low popularity. Runtimes as of 2015 are on average 98.3 minutes and having a movie too long or too short appears to affect the popularity. The second storypoint shows that there is an increase in the number of movie releases for most genres excluding Foreign Movie and Western. The pattern in second visualization shows that the bigger the budget, the more revenue it the movie collects. The last chart shows that Drama, Comedy and Action are the most profitable yet popular genres.

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

The visualization used for this question is found in dashboard 'Q2' and the multiple bar chart shows that universal is leading in terms of the total number of movies produced and total revenue. However, Paramount earns more on average per movie. On average movies produced by both companies have better vote average and its movies are longer despite it not being as popular. In terms of popularity, vote average and runtime, paramount does not differ much to that of universal. Now moving on to the scatter plot, one can see from the trend that universal earns as a lot more than it budgets as compared to paramount.

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

The visualization used for this question is found in dashboard 'Q3'. From the multiple horizontal bar graph, one can see that on average, novel based movies do better in every category than non-novel based movies. On the second graph, one can see that there are not that many novel-based movies that outperform the non-novel movies released on the same year. In conclusion, it still safer to make a novel-based movie.

Question 4: Where do the 10 most popular movies position lie in contrast to the other attributes? The analyzing of the data in tableau sparked an interest in me wanting to see how the 'most talked about movies' actually do financially/rating-wise.

The visualization used for this question is found in dashboard 'Q4'. The most popular movies have a significant difference to the other attributes. Jurassic World has a big difference when it comes to vote average/rating as it is ranked 3,068 but it makes up on revenue as its ranked 13th and its budget was not even that big with it being ranked 213th. What is interesting to see is that John Wick being one of the most popular and with above average rating did not do as well financially compared to the other popular movies. Sorting the revenue, one can see that Avatar is the movie with the highest revenue and was the 12th most popular movie. Now looking at the scatter plots, one can see that Star Wars has the biggest revenue with a decent rating and least amount of budget. My conclusion is that even if a movie is popular it does not automatically mean they are the best rated or most budgeted movie.