**Movies or Scams?**

**Data Visualisation 1**

**Name: Ethan Lee**

**ID: 29711444**

**Word Count: 901**

**URL:** <https://public.tableau.com/app/profile/ethan.lee5702/viz/MoviesorScams/MovieDashboard?publish=yes>

**Domain**

The data being presented pertains to movie franchises and their revenue and ratings, broken down into individual movies.

**Why**

This data is being presented to show the viewer if there is a correlation between rating and revenue for movies across franchises. The main question trying to be answered is: are viewers and critics enjoying new movies in franchises or are the film companies only creating more to piggyback off previous successes?

**Who**

The audience for this data is people who are interested in movies, or potentially people who are curious as to whether it is worth spending their money to see a movie.

**What**

The publicly available IMDB dataset (<https://imdb.com>) was not sufficient for this domain as it gives a random sample of movies/TV shows from the website. Any datasets that were sourced from IMDB on Kaggle (<https://www.kaggle.com>) also had this same issue. Because of this, I had to scrape data provided by The Numbers (Nash Information Services, 2023), which provided data for the revenue for franchises and their movies In conjunction with this, rating was data manually extracted from IMDB itself.

This data was copied into an Excel file, and then individual .csv files were created for each sheet, as it made updating data in Tableau more clean (Tableau wouldn’t always update the data source when the original file was update), since a single .csv file could be updated instead of having to re-add a whole Excel sheet and create the links between the tables again.

Python, with the Pandas library, was also used to generate the required .json files for the cluster diagram tool (Ladataviz, 2023). This tool helped build the framework for the cluster diagram (calculating x and y positions for nodes and specifying the lines that join nodes and their parent).

The attributes used for the whole dashboard are:

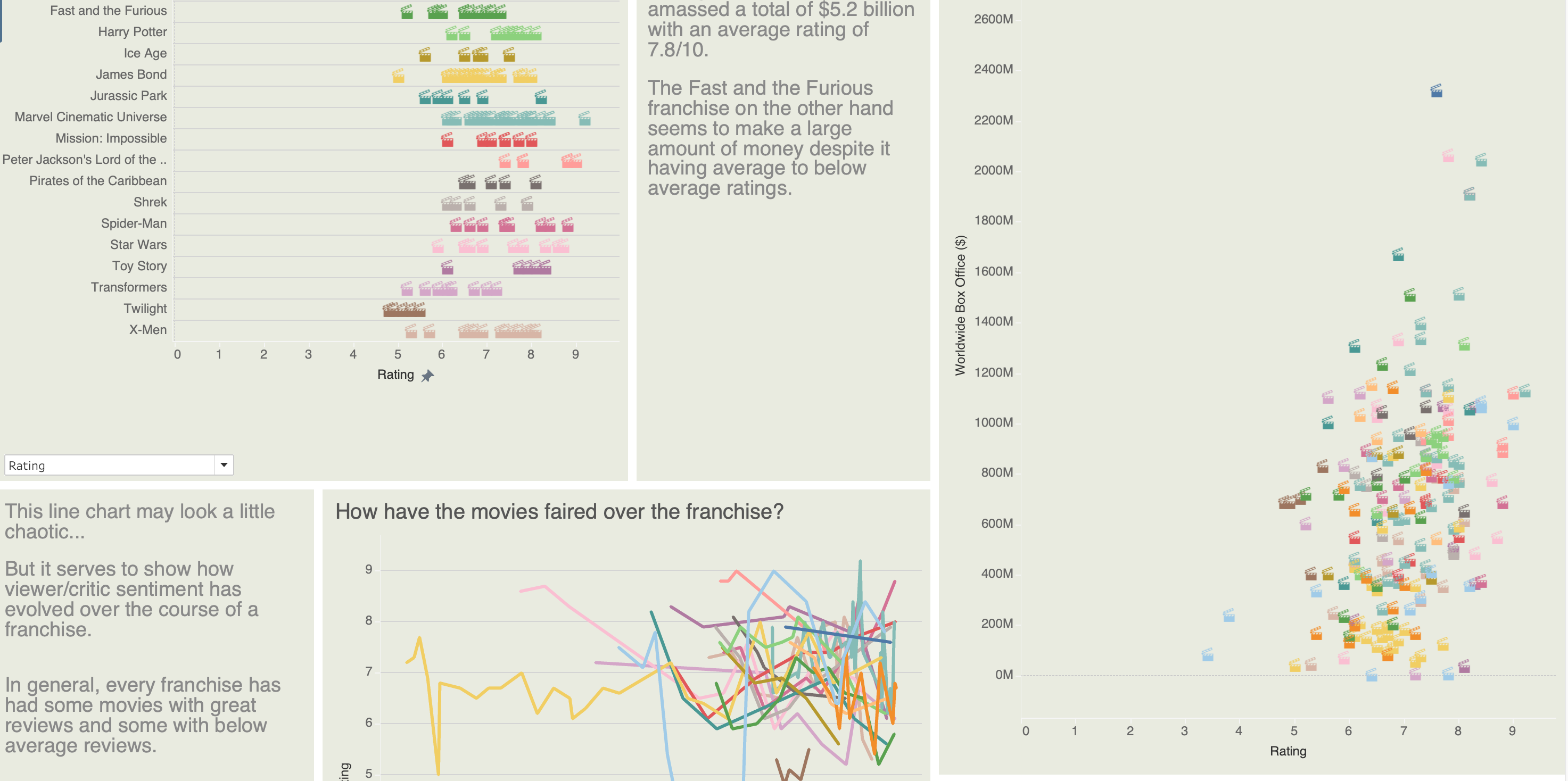
* Movie name and year (categorical)
* Movie franchise (categorical)
* Movie rating (quantitative)
* Global box office revenue (quantitative)
* Release date (ordinal)

The dataset type is a table.

**Why and How**

A screen shot of a movie

Description automatically generated A screenshot of a computer

Description automatically generated A graph with a blue line

Description automatically generated

The main idiom that is used in this dashboard is a cluster diagram, it very clearly shows the relationship between movies and which franchise they belong to, by employing the principles of connection and proximity. It provides a very visually clean and appealing means to provide an introduction for the data being presented. The marks are points and lines. Hue is used to differentiate franchises and show the relationship between movies and their franchise, and the size channel is used to show the magnitude of revenue for a datapoint.

The remaining idioms are there to provided finer details and for the viewer to analyse trends/relationships, hence why a dot plot, line chart and scatter plot are used. The hue channel, again, is used to encode the franchise, and the marks are points (with movie clapper symbols) for the dot/scatter plots and the lines for the line charts.

**Design**

**Layout**

For the layout of this dashboard, I tried to stick to a grid, so there’s 5 columns and the main visualisation spans all of them but the charts and text further down are split, charts and legend are two columns wide, and the text is one column wide.

The way visual centre of the dashboard sits just below the cluster diagram, which initially draws the viewers eyes to the cluster diagram, which is the main visualisation on the page.

I didn’t want to clutter the cluster diagram with text, so only the top five movie franchises have labels.

**Colour**

I’ve tried to keep the colours consistent throughout the dashboard, varying shades of grey for text and headings (lighter being less important). A neutral green background makes the colours of the franchises pop more, along with the white borders (dark seemed to make the chart feel more claustrophobic). The colours for the franchises are consistent across the charts, so it’s easier for the viewer to quickly infer data for each franchise.

**Figure-ground**

The use of figure-ground is employed throughout the dashboard by having less important text in a lighter grey and more important text being bolded with a darker grey hue. Having the consistent green background as the ground in the dashboard helps provide a base for the viewer to be able to infer the importance of specific aspects of the dashboard.

**Typography**

The typeface used across the dashboard is the Roboto typeface, it is a thick sans-serif font, that comes across as modern, which these movies generally are. The text blocks are aligned with the headings of the charts, which makes it easier to connect the contents of the charts and the text and the text, and the text next to the cluster diagram is in the same sightline as the legend.

In general, the importance of text is dictated by the specific Roboto font being. From least important to most important: Roboto Light, Roboto Medium, Roboto Black

**Storytelling**

The viewer naturally goes through the visualisation from top to bottom and left to right, so the more abstract data is presented at the top, which gives a high-level view of what’s being presented. After that, the viewer will go through a chart then some text, chart then some text then a chart, where each chart and text provides some finer detail and the text provides insight or background. The final visualisation of rating vs revenue is intended to answer the overall question of the dashboard.

**References**

Ladataviz BV. (2023). Network. Retrieved from

https://www.ladataviz.com/tools/advviz/network

Nash Information Services, LLC. (2023). Movie Franchises. Retrieved from

<https://www.the-numbers.com/movies/franchises>

**Appendix A – 5 Design Sheets**