NETFLIX

Netflix: Exploring Movie Popularity and Revenue Trends

Cari Bean
Sarah Suo
Susie Chen

NETFLIX

Table of Contents

Dataset Description: Columns and Rows

The Movies Dataset on Kaggle contains multiple .csv files, with a main file containing metadata on over 45,000 movies released on or before July 2020. The files also contain 26 million ratings from over 270,000 users.

This dataset consists of multiple .csv files but we will be using the following files:

- Movies_metadata.csv
- Keywords.csv
- Credits.csv
- Links.csv
- Links_small.csv
- ratings_small.csv

The main data file we will be working with is the movies_metadata.csv which will be referred to as the main Movies Metadata file. The file contains information on 45,000 movies featured in the Full MovieLens dataset. The columns are ID (integer), posters (string), budget (integer), revenue (integer), release dates (date), languages (string), production countries (string), genres (string), and companies (string).

The dataset consists of 45,466 rows of data. Each row contains details of a specific movie that has been released. The link to the entire data set can be found in Appendix A and a snapshot of the Movies Metadata file can be found in Appendix B. The main data file used for this project is movies_metadata.csv and is attached to the assignment submission

NETFLIX

Project Goals and Purpose

Netflix, as a leading streaming platform, needs to make data-driven decisions regarding content prioritization, production, and marketing strategies. My team is tasked with building a dashboard that enables marketing managers to make these decisions. After interviewing with the VP of Marketing at Netflix, the business needs include:

- Be able to view revenue trends across different content genres from the years 2014 to 2017, informing Netflix's content strategy by identifying growth areas.
- Be able to view popularity trends of movie content over time, facilitating insights into viewer preferences and engagement levels, guiding content release and promotion strategies.
- Be able to view the performance of movie content based on budget, revenue, and popularity rating, aiding in resource allocation and prioritization for Netflix Originals.

These visualizations collectively empower Netflix to maintain its competitive edge and drive continued success in the streaming industry.

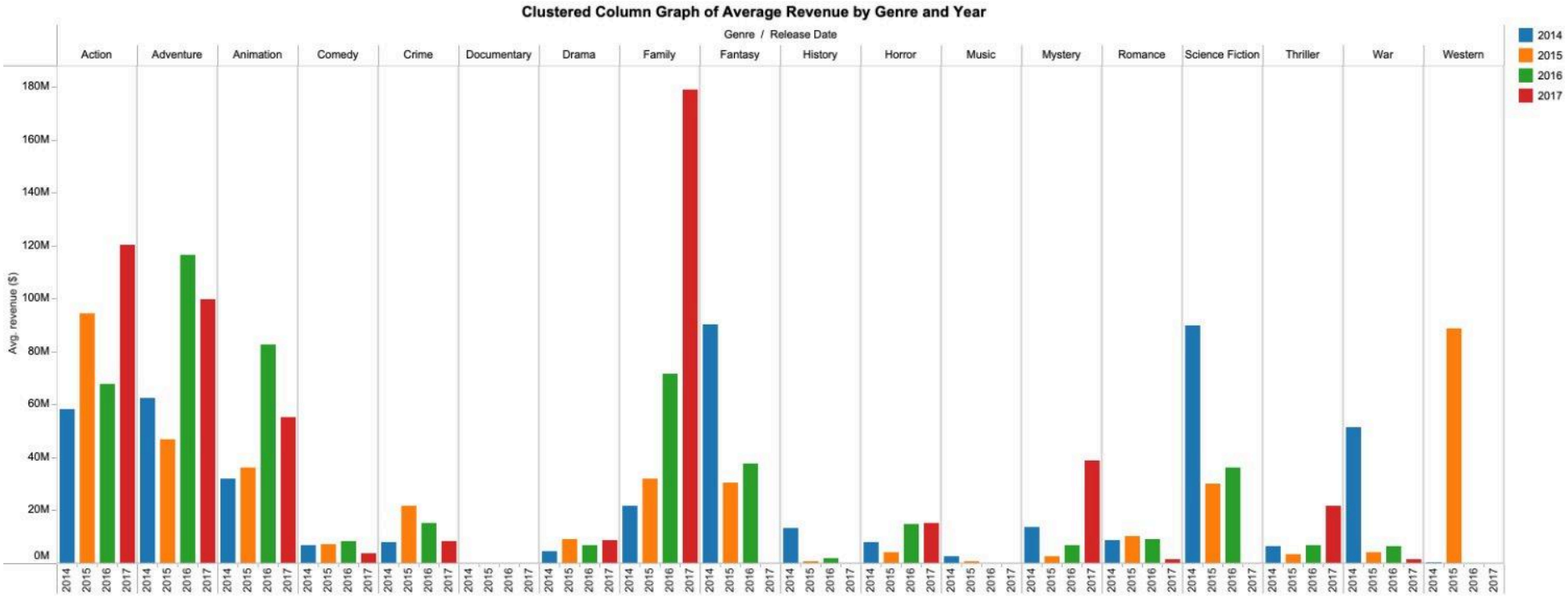
Related Works

Others have used the data set to observe whether there is a relationship between revenue and genre, but have failed to utilize other details such as keywords. We aim to take a look at more aspects of considerations for Netflix in making original content. Along with an expanded list of considerations, our goal is to create user-friendly visualizations with interactive aspects that many others have not utilized.

NETFLIX

Results

Figure 1: Cluster Column Chart of Average Revenue by Genre and Year



Average of revenue (Sarah!movies!metadata) for each Release Date Year broken down by Genre. Color shows details about Release Date Year. The data is filtered on Release Date (Movies!Metadata) Year and Release Date (Movies!Metadata) Month. The Release Date (Movies!Metadata) Year filter has multiple members selected. The Release Date (Movies!Metadata) Month filter has multiple members selected. The view is filtered on Genre, which excludes Null, Foreign and TV Movie.

Figure 1 (zoomed-in version can also be found in Appendix C) features a column graph that shows the average revenue by the type of genre. Each genre features four years, 2014-2017, and you can filter to show specific years and months. The four years are colored in four different types of colors to assist in the readability of the chart. Years 2014, 2015, 2016, and 2017 are colored blue, orange, blue, and teal respectively. The taller the bar, the more average revenue was generated for the specific genre for that year. For example, in 2017, the 'Family' genre teal bar was the tallest so that means that the 'Family' genre generated the most revenue in not only 2017, but amongst all genres in the four years. More importantly, within the 'Family' genre cluster, the genre's revenue has increased exponentially since 2014. This signals Netflix to push out more 'Family' genres as they have generated more revenue throughout the years.

The graph assists Netflix in identifying genre trends that will yield a higher average revenue when considering which original movie pitches to invest in.

The columns utilized from the Movies Metadata file are Genres, Revenue, and Release Date.

NETFLIX

Figure 2: Line Graph of Average Popularity by Month and Genre

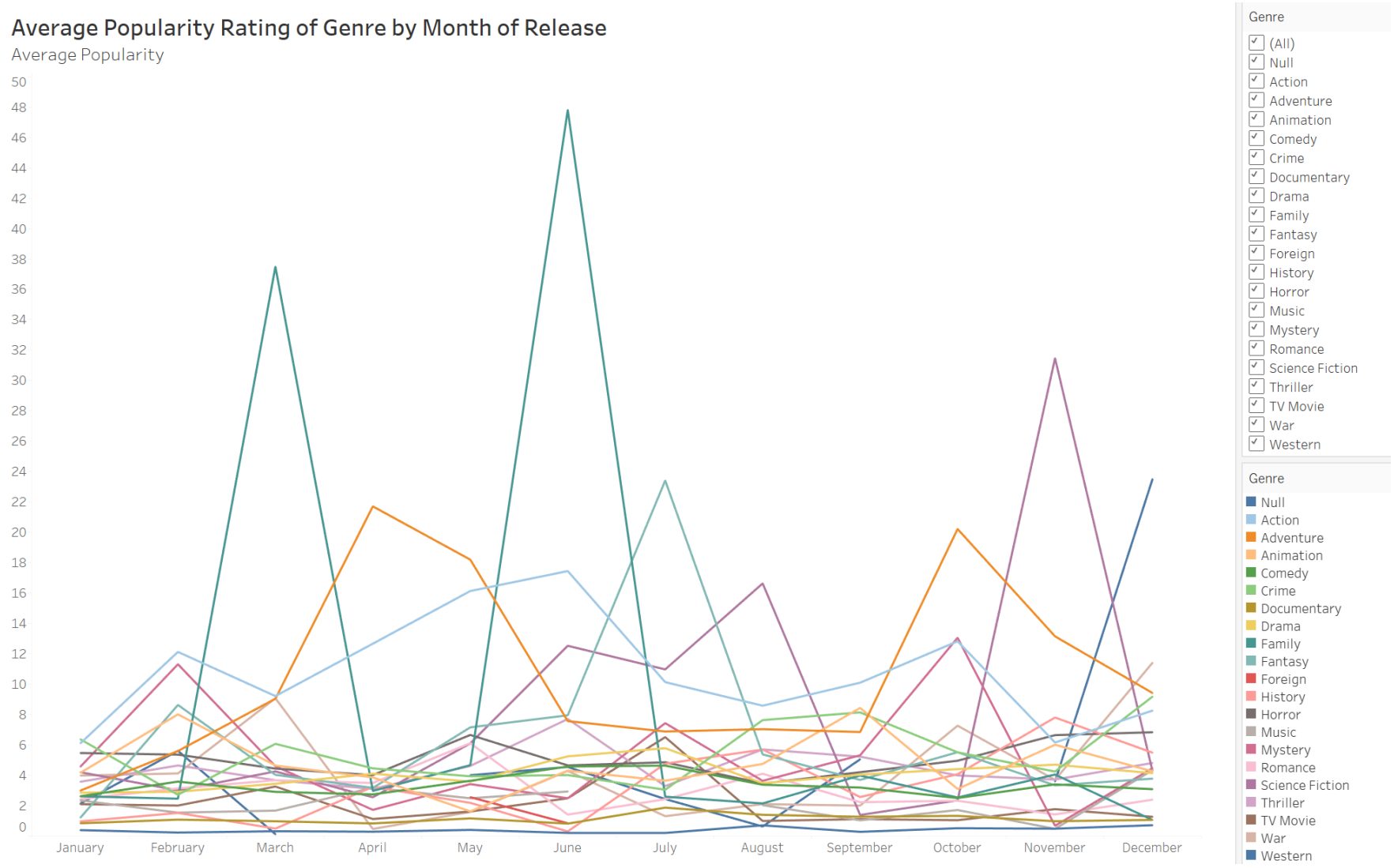


Figure 2 (zoomed-in version can also be found in Appendix D) is a line graph showing the average popularity of each genre of film over the months of the year. Each genre is a different colored line, and there is a filter for the genre, allowing the audience to pick which genres show on the graph. The taller the line for a genre is, the more popular that genre was during that particular month on the x-axis. For example, we can see that the 'Family' genre of movies had higher average popularity during March and June, but they had less popularity for the rest of the months of the year. We could infer from this that movies released in the genre of 'Family' have more popularity and more people watching during those two months. Similarly, movies in the genre of 'Science Fiction' have a higher average popularity during November. Using the filter, the audience can easily compare the average popularity of certain genres over the months of the year.

One of our project goals is to be able to view popularity trends of movie content over time, and the graph shown in Figure 2 accomplishes this. The graph can be used to show the popularity trends of movie genres over the months of the year. The graph can then be used to show viewer preferences for movie genres over a year. Netflix can use this information to determine what genre of movie they should release and when to generate the most user engagement and potential popularity, and from there they can more accurately determine their promotion schedules and strategies for their content.

The columns from the Movies Metadata file used in creating the graph in Figure 2 include Genres, Release Date, and Popularity.

Figure 3: Bubble Chart of Movie Budget, Revenue, and Popularity

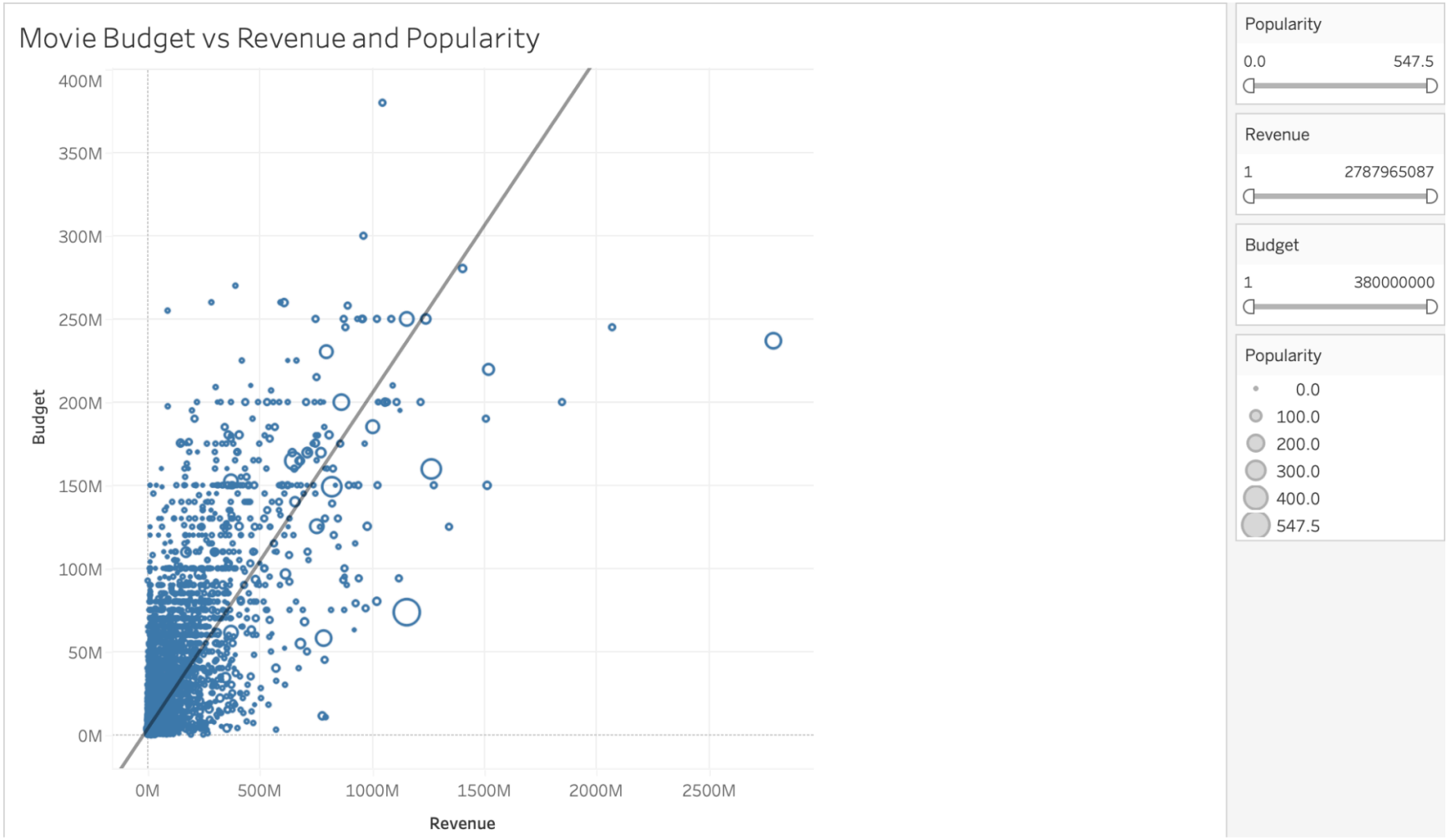
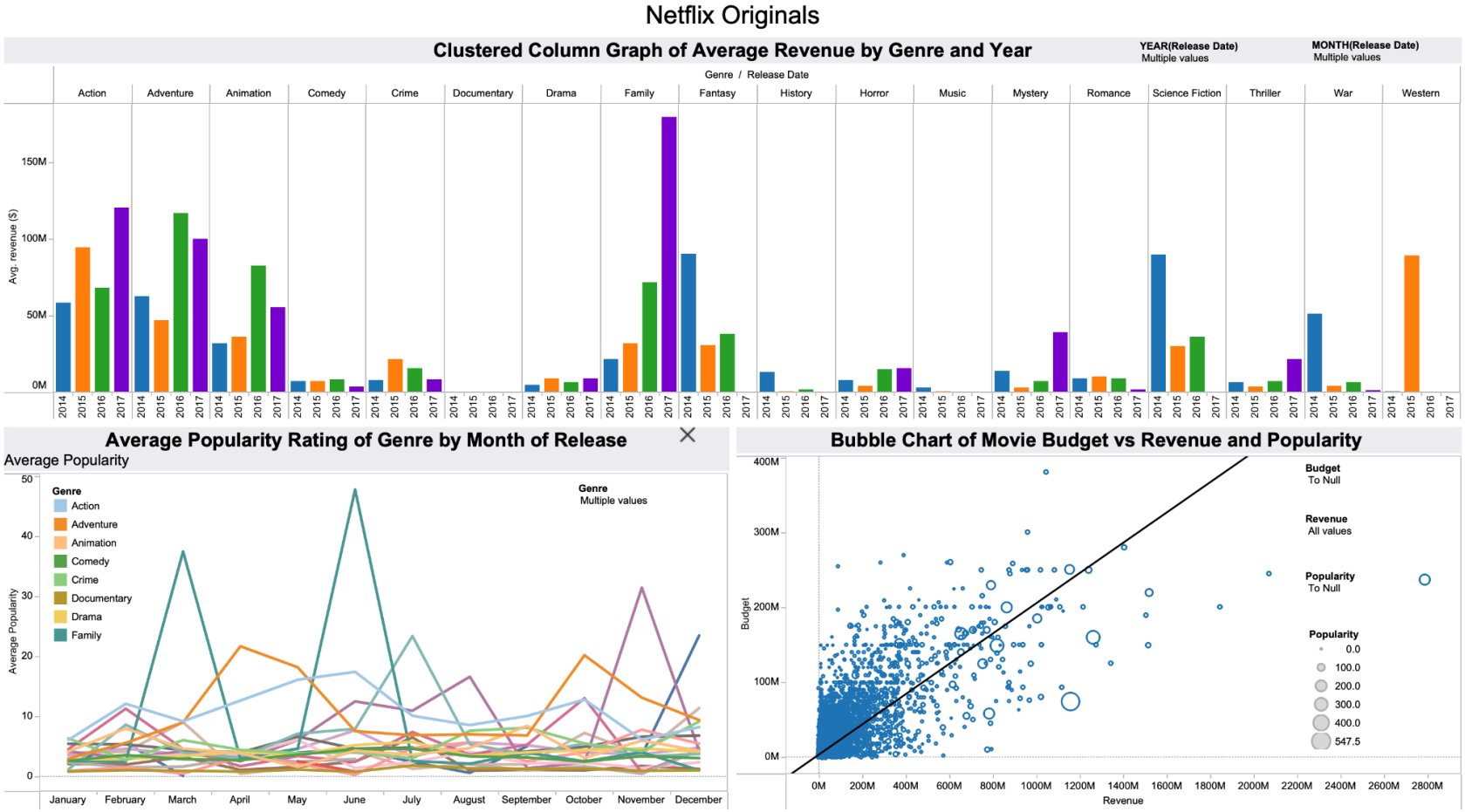


Figure 3 (zoomed-in version can also be found in Appendix E) is a Bubble Chart that examines the relationship between 3 quantitative variables: Movie budget, revenue, and popularity. Budget lies along the y-axis, revenue lies along the x-axis, and popularity is illustrated by the size of the "bubbles" on the chart. A smaller bubble equates to a lower popularity rating while a larger bubble equates to a higher popularity rating. Based on the Bubble Chart, we can see that generally as budget size increases, so does revenue. Since over 40,000 data records are being plotted on this graph, it is difficult to distinguish the bubble sizes and popularity ratings. Upon closer examination of popularity using the filter, we see that there is a general trend that as budget size increases, revenue increases and popularity ratings also increase. However, some data points in the dataset deviate from this trend. For example, the biggest bubble on the bubble chart is associated with the Minions movie released in June 2015. This movie has a more conservative movie budget in the dataset, but made much more in revenue compared to movies with similar budgets, and was widely more popular than any movie in the dataset. Movies that exhibit these characteristics should be closely monitored by Netflix managers when it comes to developing movie content strategy.

This bubble chart helps enable Netflix managers to view the performance of movie content based on budget, revenue, and popularity rating. Managers can use this chart to generally determine how much money to spend on making an original and assess the ROI in terms of revenue and audience popularity aiding in resource allocation and prioritization for Netflix Originals.

The columns used are budget, revenue, and popularity which can be found in the Movies Metadata file.

Dashboard, Conclusions, & Actionable Recommendations



NETFLIX

The dashboard that our team has built contains 3 graphs and addresses the following business needs:

1. The Clustered Column Chart enables users to view revenue trends across different content genres from the years 2014 to 2017, informing Netflix's content strategy by identifying growth areas.
2. The Line Graph enables users to view popularity trends of movie content over time, facilitating insights into viewer preferences and engagement levels, guiding content release and promotion strategies.
3. The Bubble Chart enables users to view the performance of movie content based on budget, revenue, and popularity rating, aiding in resource allocation and prioritization for Netflix Originals.

Based on the dashboard above, we can draw the following insights from each graph:

1. Clustered Column Chart: The Genre Family resulted in the highest average revenue of any genre in the data set in 2017.
2. Line Graph: The Genre Family has a higher average popularity during March and June
3. Bubble Chart: The biggest bubble is the Minions movie released in June 2015 under the genre Family with a more conservative movie budget (approx. $74,000,000), but made more in revenue ($1,156,730,962) compared to other movies with similar budget size and was more popular than any movie in the dataset.

Based on these insights, Netflix should produce and release more original Family genre content during March and June when families are together for spring and summer break since this genre gives the highest ROI in terms of revenue and audience popularity.

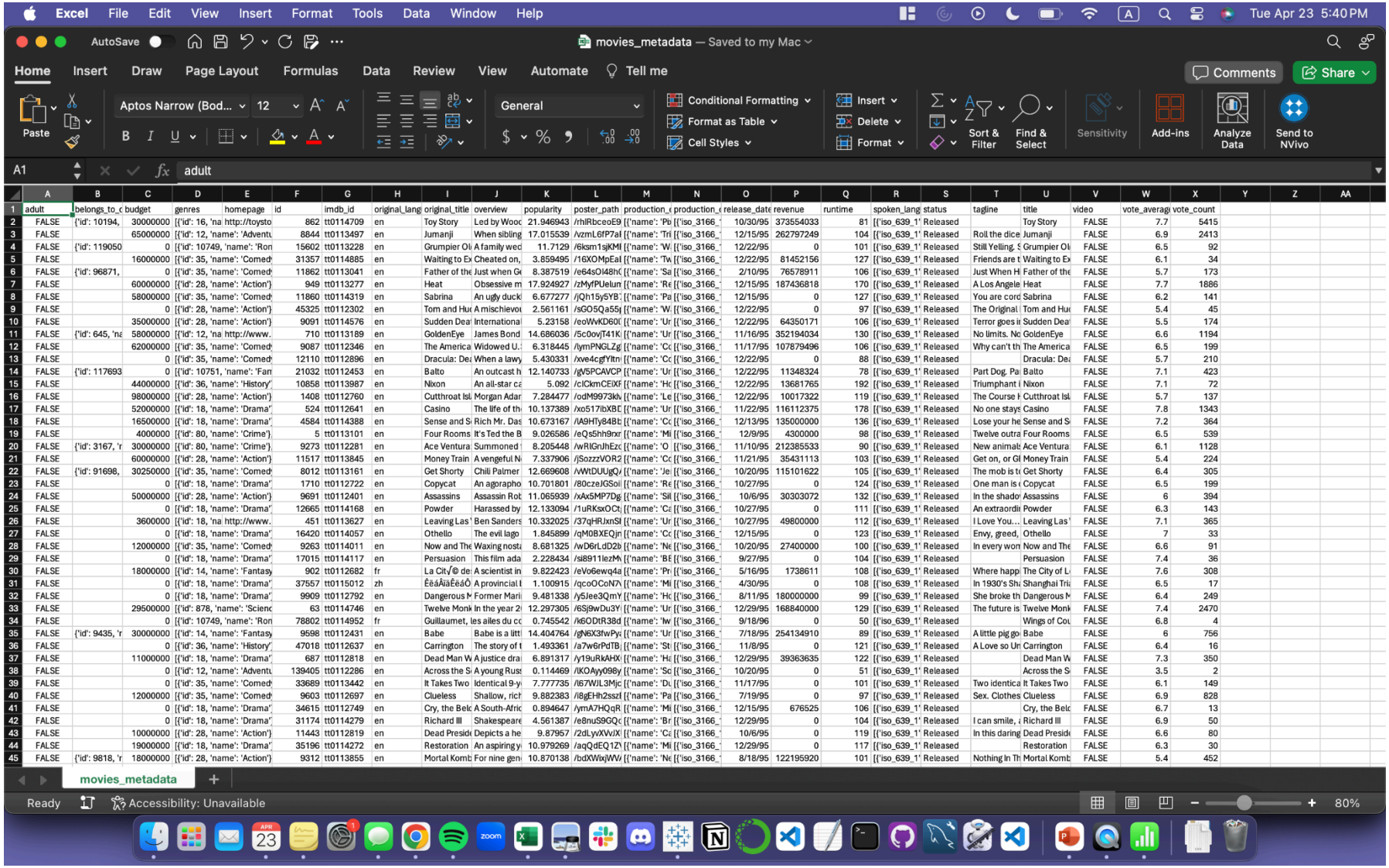
NETFLIX

Appendix

Appendix A: The link to the Data file can be found at https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset
The main data file used for this project is movies_metadata.csv and is attached to the assignment submission, but the entire data file can be found at the link above.

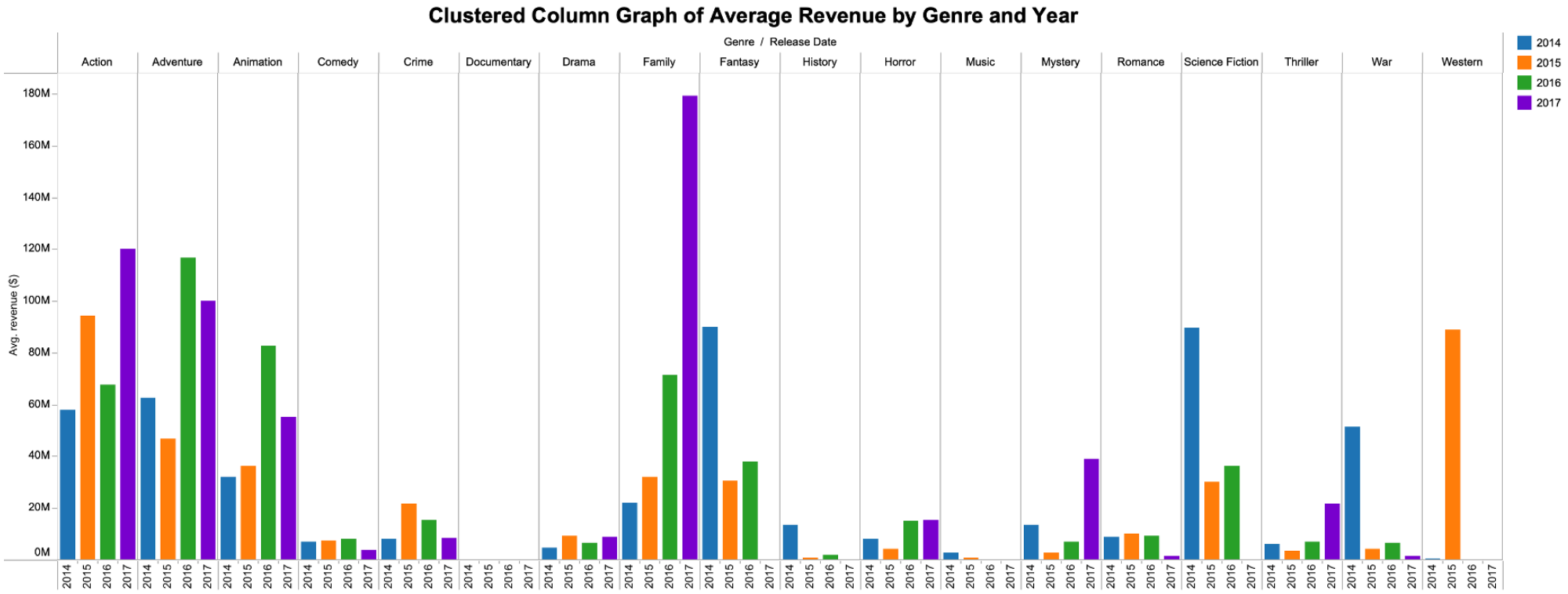
NETFLIX

Appendix B: Snapshot of the Movies Metadata file



NETFLIX

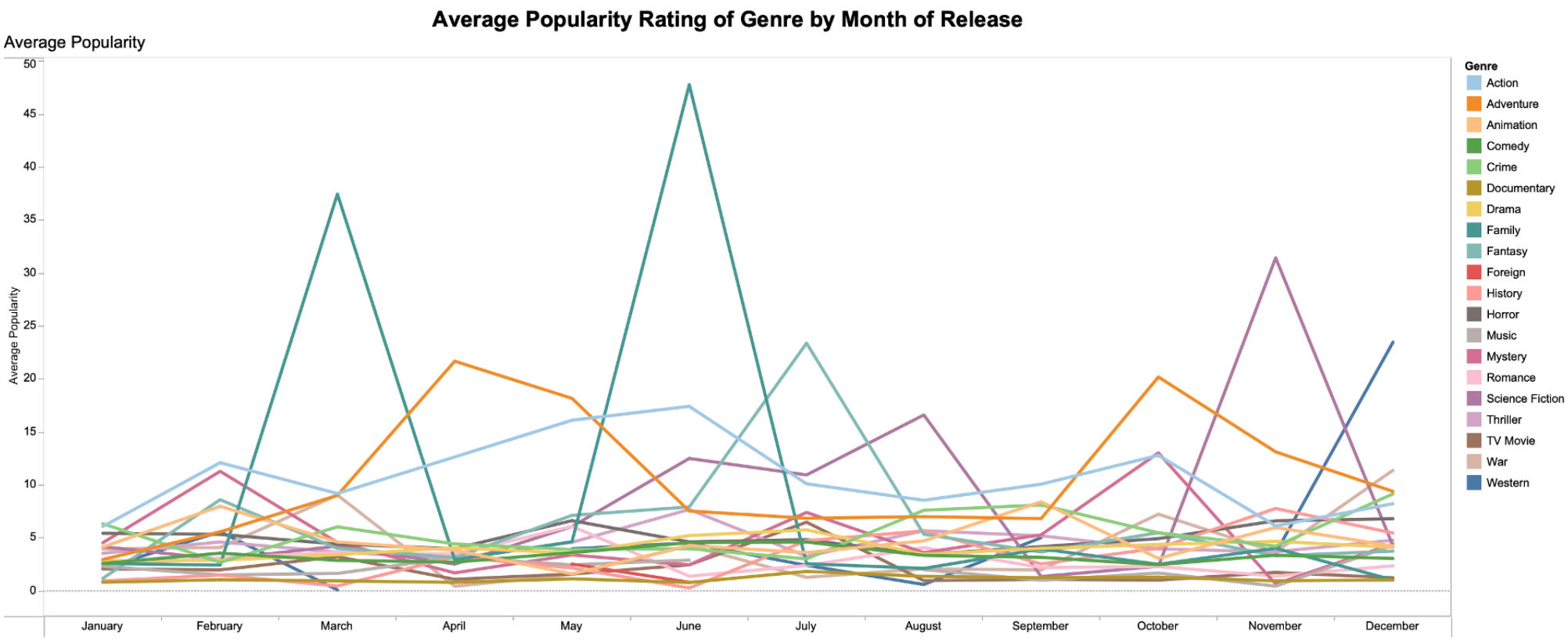
Appendix C: Figure 1



Average of revenue (Sarah!movies!metadata) for each Release Date Year broken down by Genre. Color shows details about Release Date Year. The data is filtered on Release Date (Movies!Metadata) Year and Release Date (Movies!Metadata) Month. The Release Date (Movies!Metadata) Year filter has multiple members selected. The Release Date (Movies!Metadata) Month filter has multiple members selected. The view is filtered on Genre, which excludes Null, Foreign and TV Movie.

NETFLIX

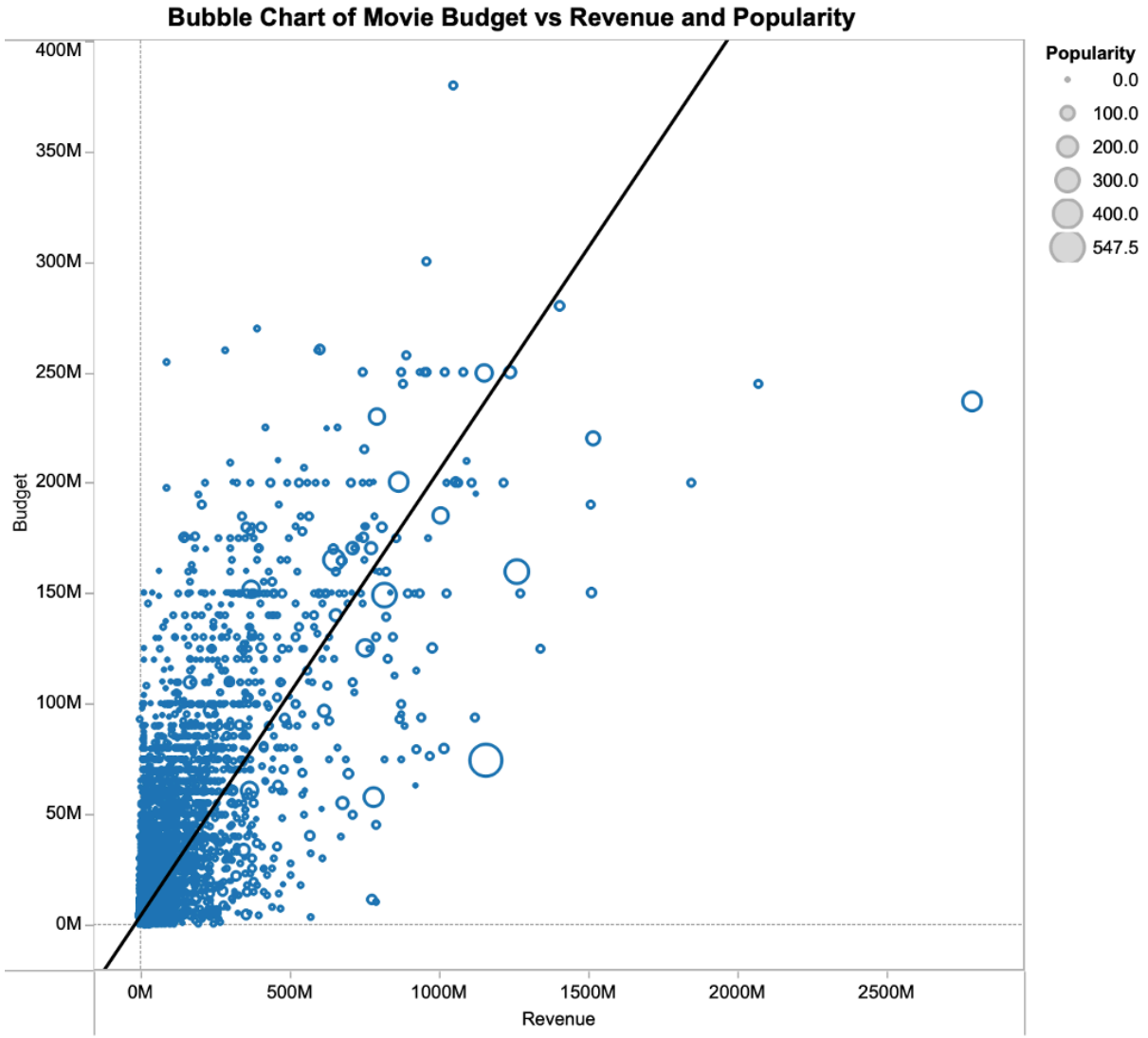
Appendix D: Figure 2



The trend of average of Popularity for Release Date Month. Color shows details about Genre. The view is filtered on Genre, which excludes Null.

NETFLIX

Appendix E: Figure 3



Revenue vs. Budget. Size shows Popularity. The data is filtered on Budget, Revenue and Popularity. The Budget filter includes everything. The Revenue filter keeps all values. The Popularity filter includes everything.