**Comparison of Performance between Streaming Platforms**

Data Visualisation and Dashboarding

Table of Contents

[1 Research Question 2](#_Toc199674802)

[1.2 Introduction 3](#_Toc199674803)

[2 Data Acquisition 3](#_Toc199674804)

[3 Data Preparation 3](#_Toc199674805)

[4 Exploratory Data Analysis 4](#_Toc199674806)

[4.1 Amount of Content 5](#_Toc199674807)

[4.2 Highly Rated Content (IMDb ratings) 6](#_Toc199674808)

[4.3 Exclusive Content 9](#_Toc199674809)

[5 Conclusion 11](#_Toc199674810)

# Research Question

How do the performance metrics of streaming platforms compare in terms of amount of content, popularity and highly-rated movies, and exclusive streaming content?

## Introduction

The video streaming services are everywhere, and it’s one that people use daily and got habituated since the pandemic. What would we do if we couldn't binge-watch films, TV series, and fresh stuff every other day? With an ever-expanding array of streaming options, users are confronted with an abundance of choices, determining which sites are best for personal enjoyment and amusement can be tough. Delving into this inquiry provides valuable insights into user-centric content curation and platform distinctiveness, influencing the choices of discerning viewers seeking a tailored and worthwhile streaming experience (e.g., if they value a wide variety of content, popular or highly recommended or rated movies or shows).

# Data Acquisition

The data is taken from Kaggle, which is great for valuable resources, providing a diverse and vast collection of datasets to support a wide range of research, analysis, and machine learning projects.

The datasets, which come from a variety of fields and businesses, give users the tools they need to start exciting initiatives like exploratory data analysis that I have done for this report.

Below are the links of the data for the 4 streaming platforms, Netflix, Amazon Prime, Disney Plus and Hulu.

#Dataset1 – [Netflix Movies and TV Shows](https://www.kaggle.com/datasets/shivamb/netflix-shows)

#Dataset2 – [Amazon Prime Movies and TV Shows](https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows)

#Dataset3 – [Disney Plus Movies and TV Shows](https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows)

#Dataset4 – [Hulu Movies and TV Shows](https://www.kaggle.com/datasets/shivamb/hulu-movies-and-tv-shows)

As the datasets don’t have IMDb ratings and number of people rated to analyse data, I created my own data by extracting the open source IMDb data on their website.

Link – [IMDb Dataset details](https://developer.imdb.com/non-commercial-datasets/)

I chose these datasets as they have the data that I wanted to analyse, I did find other datasets in Kaggle, but they did not have few attributes such as proper age rating and genre.

There were, however, a couple more measures that I would’ve liked to include in my analysis, and could not source. These measures were, user engagement (viewership/times watched) and original or licenced content, these would have facilitated more in depth analysis that could have been factors to provide recommendation to individuals.

# Data Preparation

The purpose of preparing the dataset for exploratory data analysis is to remove any columns or data that is not useful. The removed columns are:

Show\_id, director, cast, description, date\_added (column had many null values) in all the four datasets.

Renamed a column named listed\_in column to genre. I wanted to show visualizations comparing these four datasets so I merged all four into one, in the previous step of merging we should be able to identify the rows according to the name of streaming service, therefore added a column named Platform and entered their respective names in their datasets.

For a metric, there was a need to add IMDb ratings and number of votes to all these datasets, using R programming I merged the tables of the IMDb datasets that were online due to their massive size, the columns that were taken are: primaryTitle, year, numVotes and averageRating and named the dataset as Ratings.

Later created duplicates of the 4 streaming service datasets, removed columns: country, rating (age rating) and duration, and then merged them with the Ratings dataset. Then created a measure called Rank - Rank = RANKX(ALL(amazon\_prime\_imdb), [Rating] \* [No\_votes],,DESC,Dense) for all the 4 newly created datasets so that there would be no bias while comparing the 4 datasets due to the amount of data.

# Exploratory Data Analysis

Understanding the level of variety each streaming service offers is one of the most crucial considerations to make when deciding which service to subscribe to in order to make the best economical decision. A consumer might not quickly run out of films to watch if they had access to a greater variety of content, which would be advantageous for the streaming service. More variety would, from the perspective of the client, eliminate the need for numerous subscriptions to various services. Why spend more money than necessary?

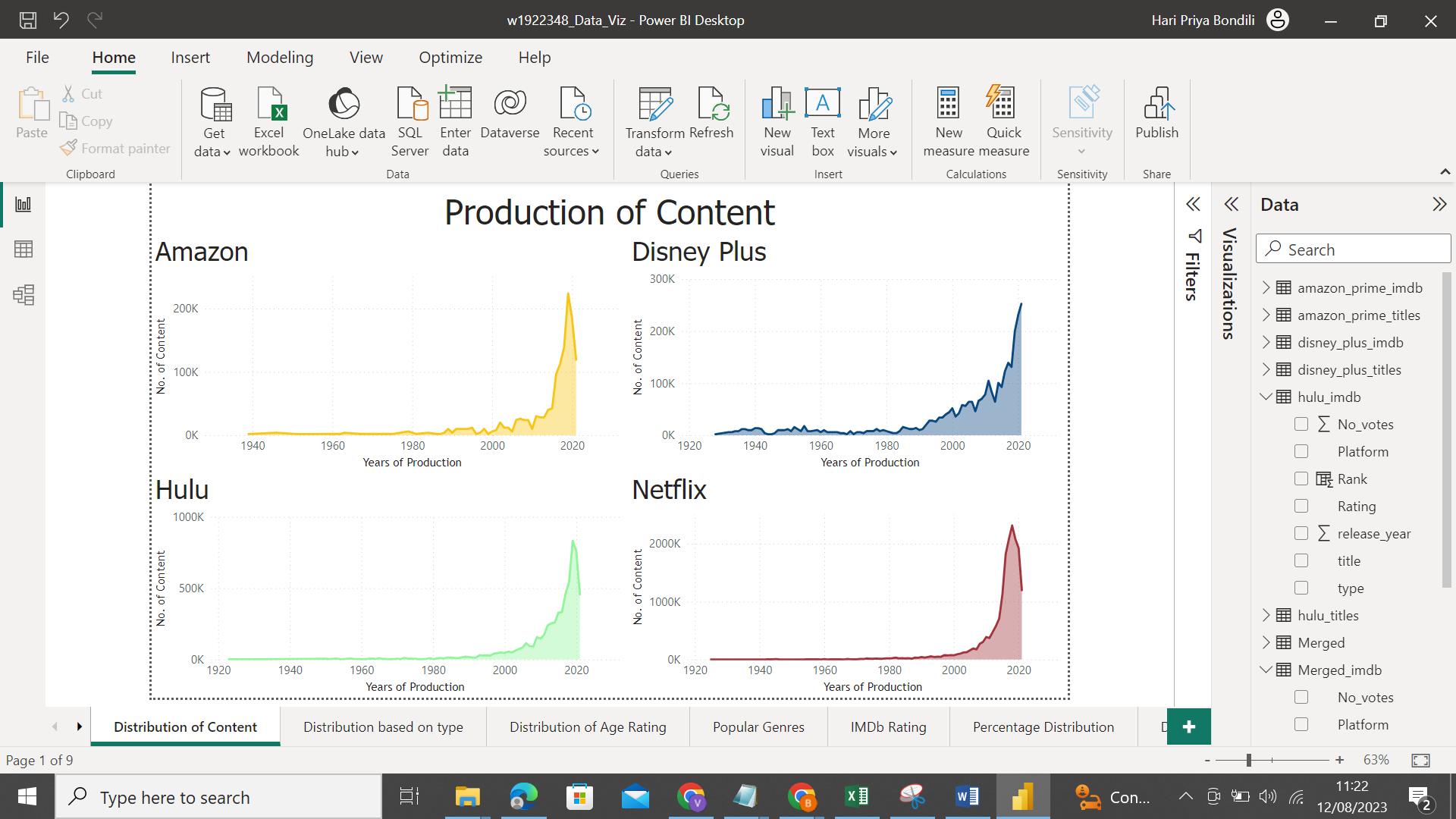


Figure. 1

There are a few peculiarities in the statistics, including how the distribution of films on Hulu and Amazon Prime is very similar, making it unclear whether streaming service actually offers the most diversified collection of films.

## Amount of Content

The variety of films and television series offered is a crucial consideration when selecting any streaming service. The amount of information accessible, especially for die-hard movie fans, is essential for discovering new entertainment and moving outside of one's familiar territory.

It was possible to compare the overall number of films and shows accessible on each streaming platform by using our datasets to distinguish between the films and shows available on each platform and gather the amount of content.

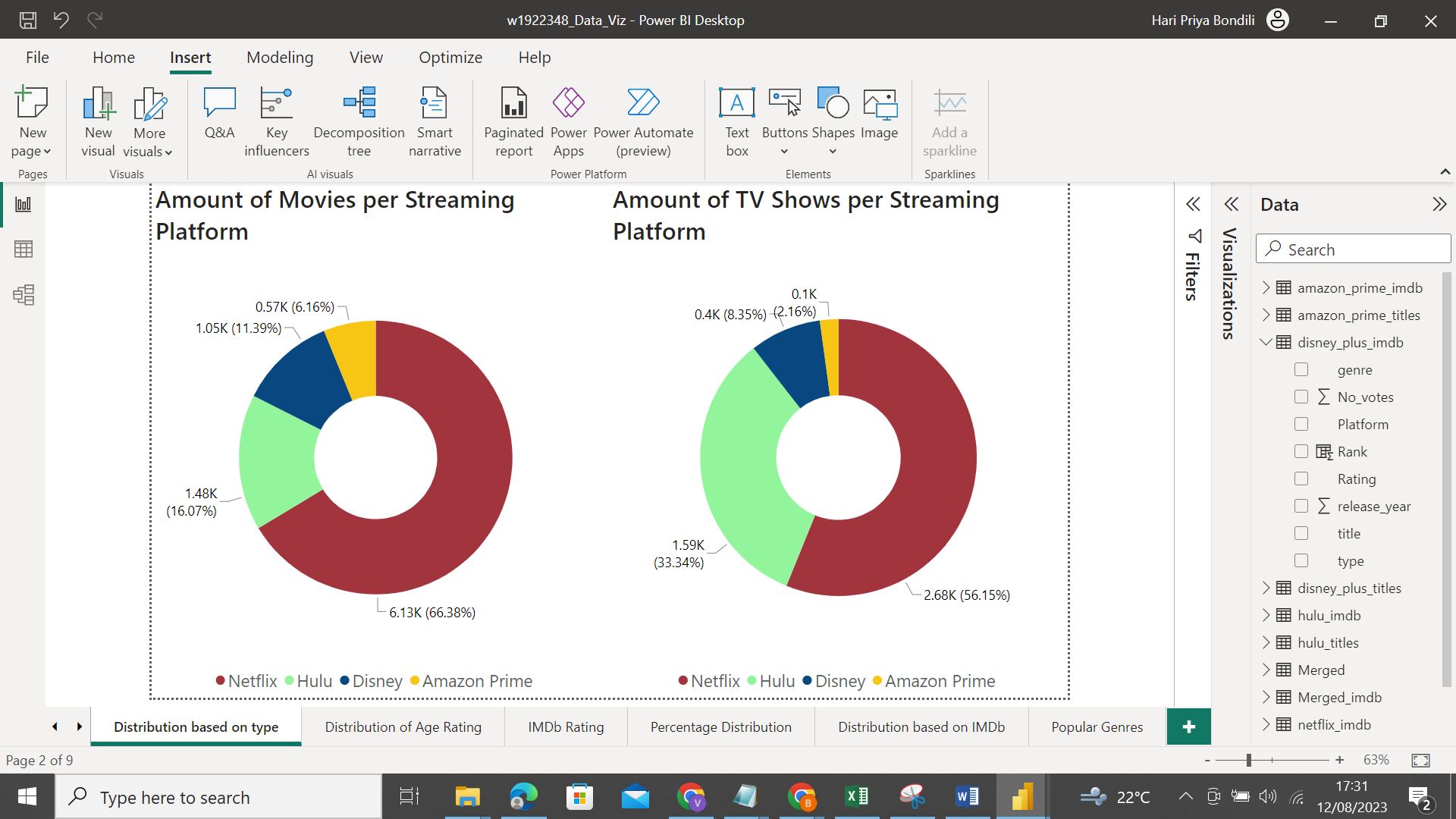


Figure. 2

From the graphs shown above, it is clear that Netflix contains an overwhelming amount of content from a movie selection of over 6,000 titles and 2,000 shows.

Netflix, Hulu, Disney Plus, and Amazon Prime are in order of having the most content to having the least, which reflects the patterns across the four services for both series and movies. The fact that Netflix has a large advantage in terms of the number of movie titles it offers is also significant to note.

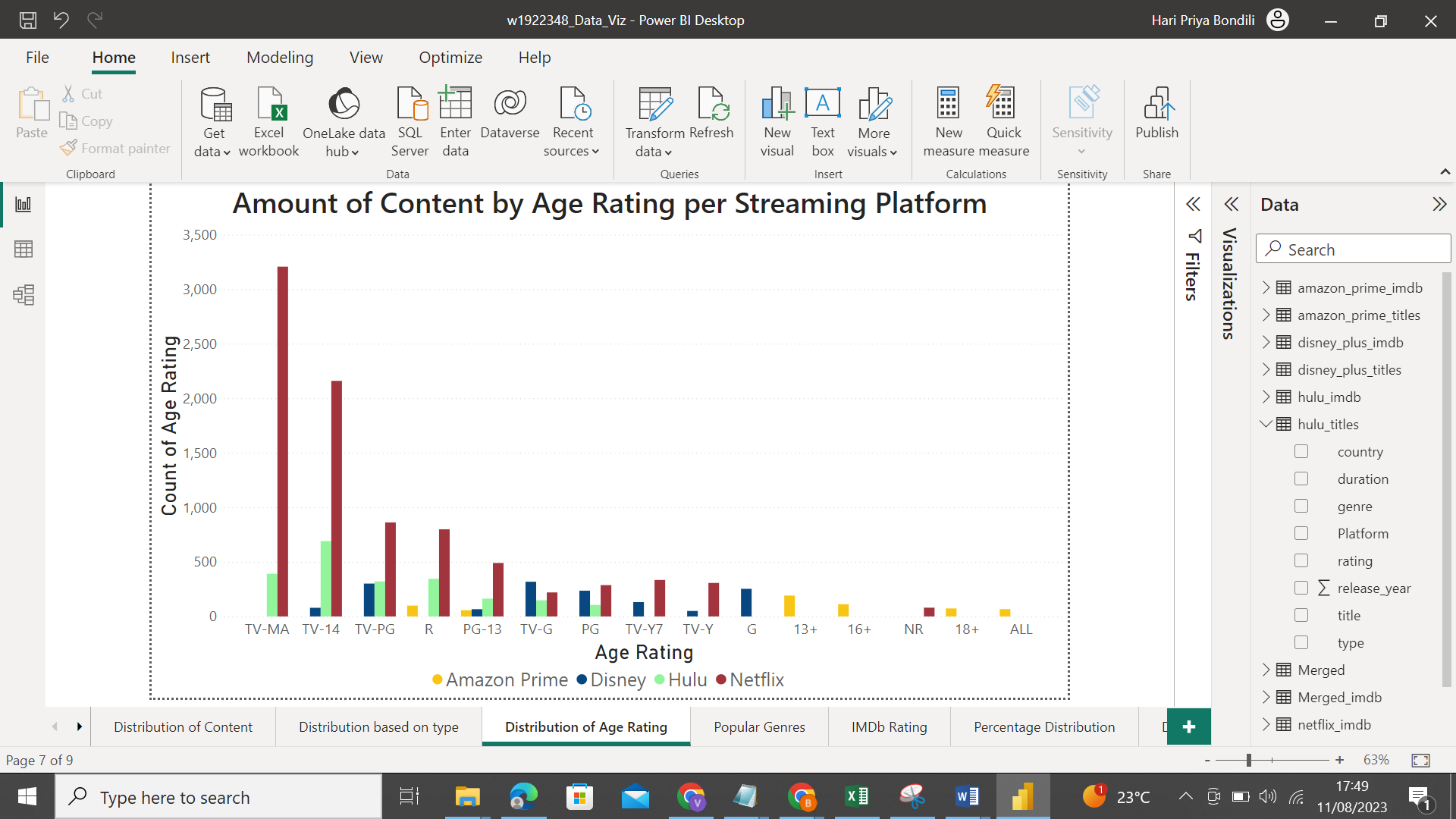


Figure. 3

From the Clustered chart above, we can see that each streaming service has a different emphasis on different age groups. Netflix has the most Rated 18+ movies overall as well as a significant advantage in 7+ and 13+ movies. However, Disney+ has approximately an equal amount of content as Netflix for movies recommended for all ages despite it having less content, which emphasizes Disney Plus kid-friendly focus. Amazon Prime has the least amount of content distributed between the ages, it could be possible that the data is not in quite handful amount as compared to the other three datasets that we have. In common:

* The largest count of movies is of 'TV-MA' rating. "TV-MA" is a rating assigned by the TV Parental Guidelines to a television program that was designed for mature audiences only.
* Second largest is the 'TV-14' stands for content that may be inappropriate for children younger than 14 years of age.
* 'TV-PG' and 'R' has almost equal amount of movies.

The results indicate that Netflix would be the greatest subscription for individuals who prefer discovering a range of entertainment based only on the amount of content. Hulu, Disney Plus, and Amazon Prime are the Netflix rivals in decreasing order. Customers may feel more at ease selecting from the offerings on Hulu and Disney Plus if they are feeling overloaded by the amount of content accessible on Netflix.

## Highly Rated Content (IMDb ratings)

We wouldn't want to pay for a streaming service that provides films that are primarily of little value watching, it makes sense that one of our primary concerns is the average movie quality provided by the streaming service we subscribe to. Let's look at how IMDb ratings are distributed among the four streaming platforms.

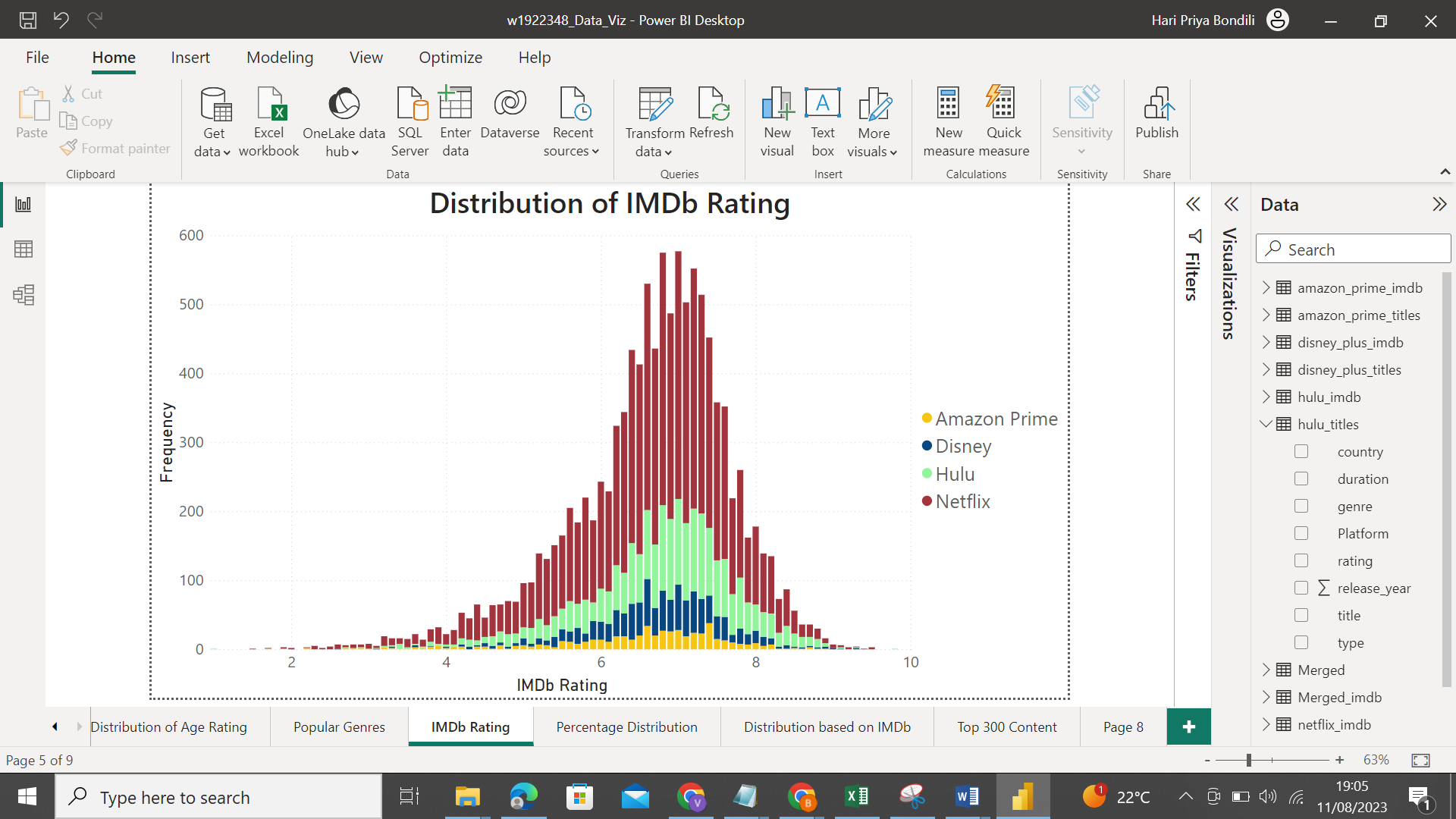


Figure. 4

The histogram above shows a slightly skewed distribution in terms of IMDb and it seems that each streaming service has roughly the same distribution.

Additionally, since we often do not watch content with poor ratings, the number and percentage of excellent quality content is the more important information we may look at. Two charts showing the frequency of content by service and rating are shown below. Three categories of ratings are used: above 8.0, between 6.0 and 8.0, and below 6.0. We can easily see that although Netflix has the most number of content scoring higher than 6.0, Netflix achieved this with a big base of content, as demonstrated in the previous section, therefore the proportion of content scoring higher than 6.0 is actually relatively low. Disney Plus, on the other hand, has the least content that score better than 6.0 but a disproportionately greater percentage of good content. Moreover, Amazon Prime has the lowest amount of content but relatively the same percentage of distribution of good content as Netflix and Hulu.

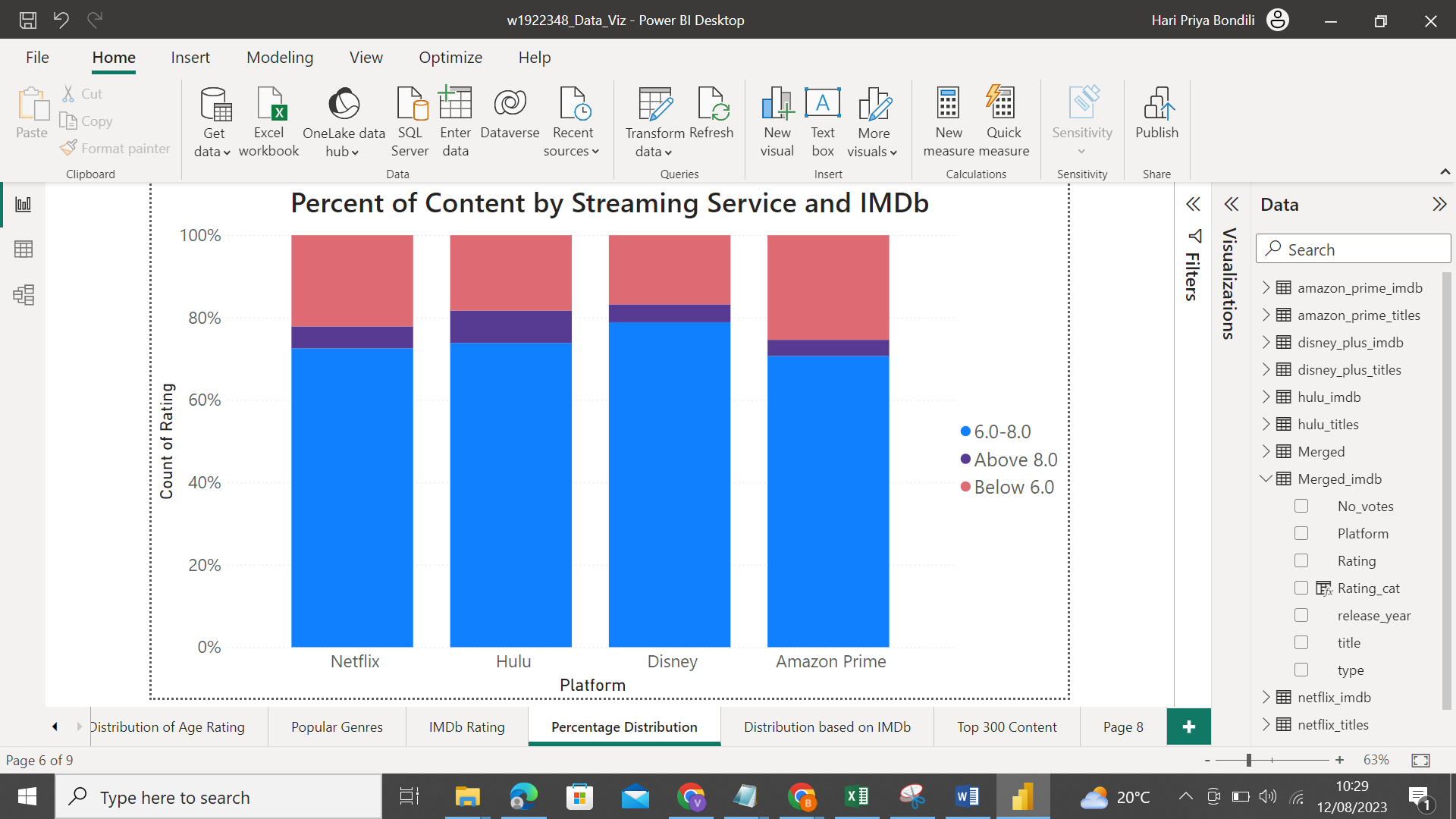


Figure. 5

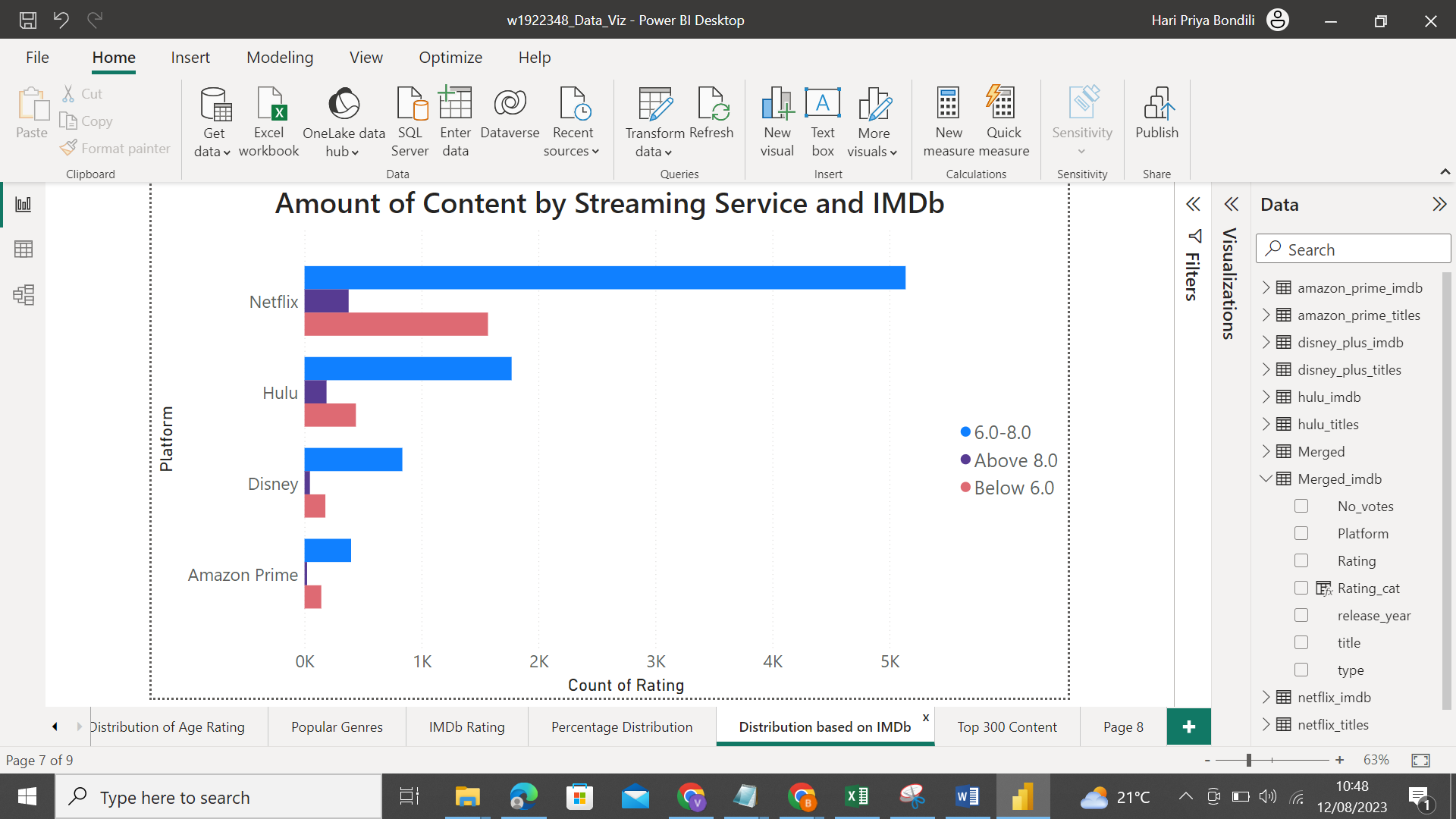


Figure. 6

As a result, Disney Plus, Hulu and Netflix are the top three services in terms of the percentage of content with IMDb ratings above 6.0. This score may be used as a reference to help users decide which streaming service they would prefer if they are unwilling to spend money on poor content.

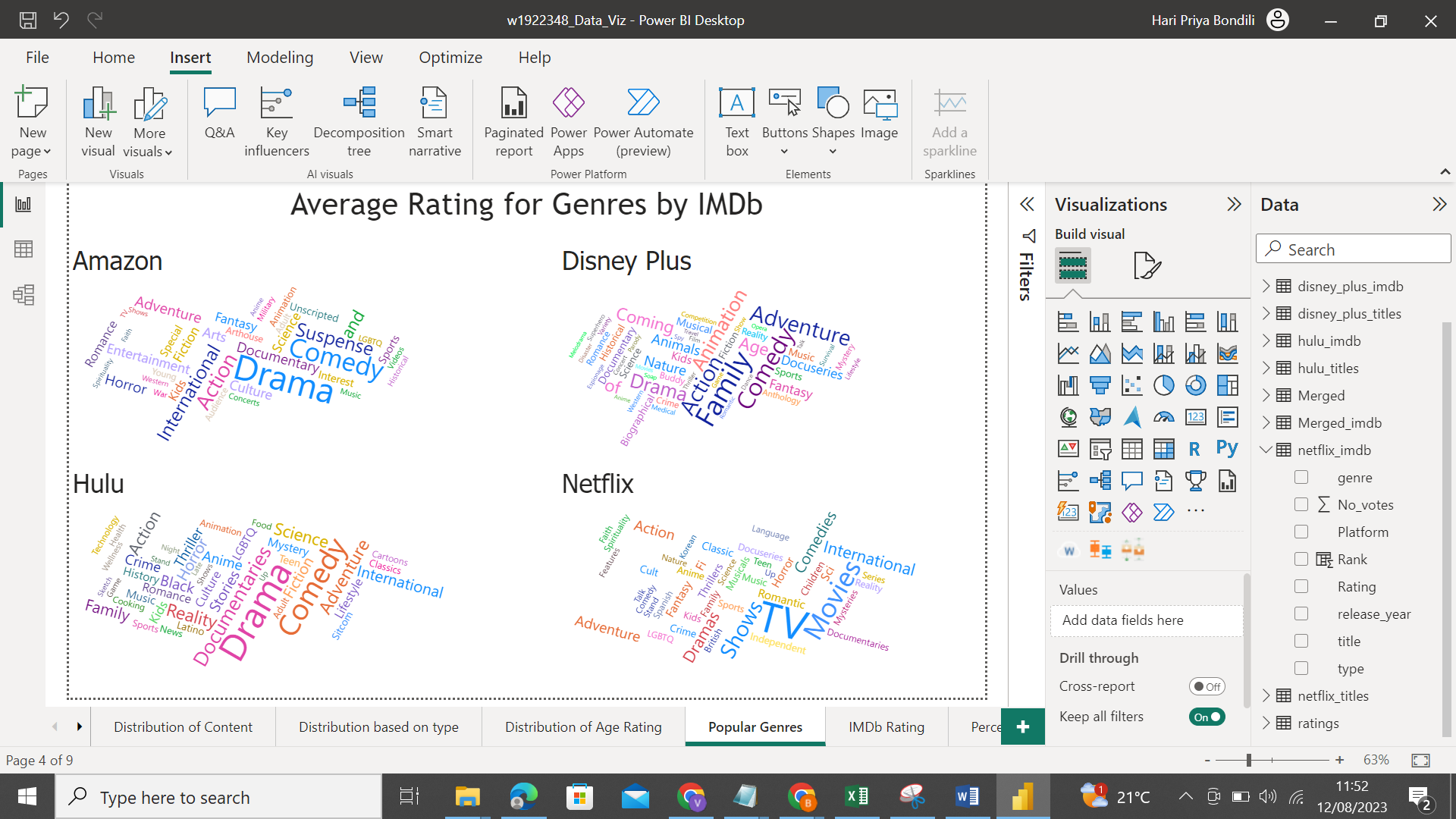


Figure. 7

Here are some word clouds representing various genres in order of the average rating for each streaming provider. Netflix, for instance, has greater ratings for comedies and dramas in films and on television. Family, comedic, action, and adventure films and TV series on Disney Plus have received favourable reviews in comparison. The drama, suspense, and documentaries on Hulu are ranked higher; the drama, comedy, and suspense on Amazon Prime are somewhat better than other genres. As a result, if you have a strong taste for the kind of films you like, the movie genres that each streaming service excels at may also be something you want to take into account.

## Exclusive Content

Due to the fact that we think what's accessible is what draws in the typical movie viewer, I chose to examine the data and scores across each streaming service based on the top 100 movies and TV series that each streaming service offers.

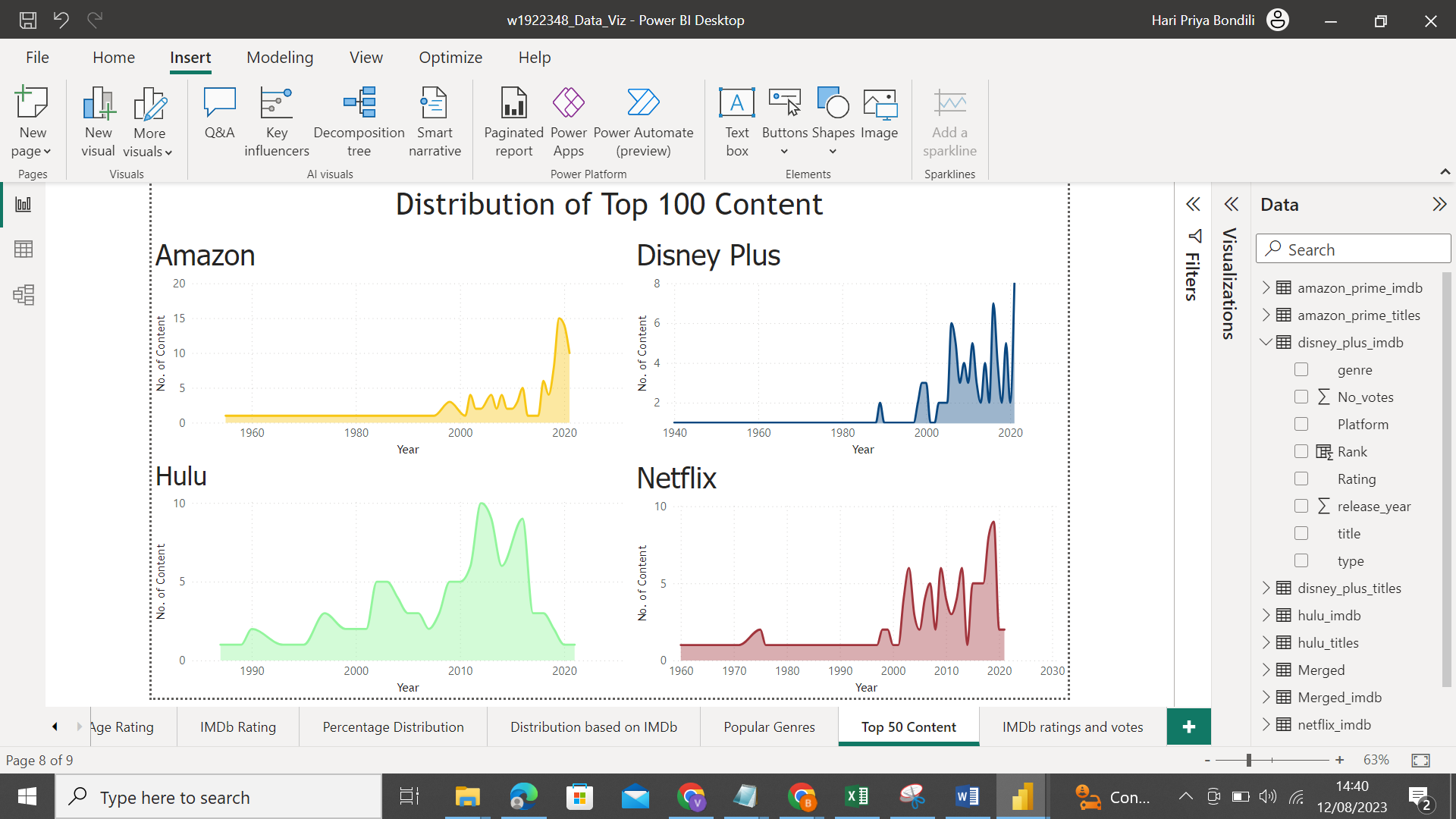


Figure. 8

We also want to look at the median ratings, which might be a better indicator of the usual or middle ratings of the movies than the actual mean as the histogram above depicts a slightly skewed distribution in terms of IMDb in Figure. 4.

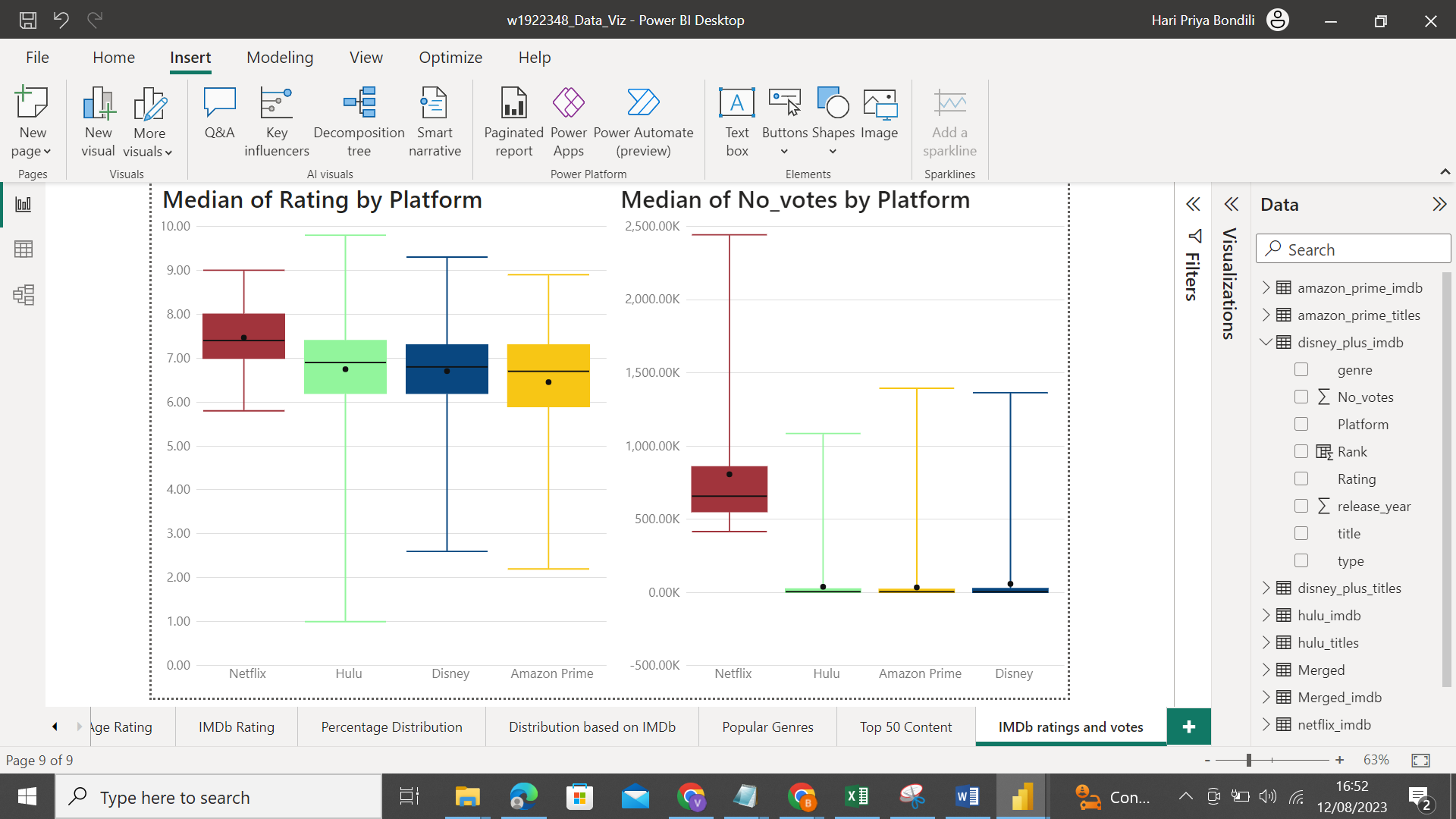


Figure. 9

When comparing each service, we can see that Netflix has the highest median number of IMDb reviews for its top 100 films, while the median for the other three services is zero due to their extreme negative skewness, which occurs when a small number of very large values cause the mean to skew to the right while the median remains close to the group of smaller values. Additionally, the top 50% of all user reviews for Netflix's most popular films are able to outperform the top 75% of user evaluations for each of the competing services.

# Conclusion

Now that you've gone through everything, you might be wondering what the final rating for each of the streaming services is, the total ranking across all three categories, including the quantity of material, IMDb ratings, and exclusive content is as follows:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Quantity of Content | Highly Rated Content | Exclusive Content |
| Rank 1 | Netflix | Disney Plus | Netflix |
| Rank 2 | Hulu | Hulu | Hulu |
| Rank 3 | Disney Plus | Netflix | Disney Plus and Amazon Prime |
| Rank 4 | Amazon Prime | Amazon Prime |  |

Table. 1

Netflix! According to our data analysis and visualisations, Netflix was the top streaming service overall, coming in first place for quantity of content, third for highly rated or popular content, and first for exclusive content. According to the distribution of the amount of material by age rating and to the percentage of content that has more high-quality films comes in as first, Disney Plus comes in second. Surprisingly, the two streaming services, Prime and Hulu, matched for third with the rankings. Therefore, if you're unclear about what to watch the next time, Netflix is a great option.

This analysis is only based on the data that I have taken. However, you can always choose the streaming service according to other metrics like number of devices, cost of subscription plan, device compatibility etc.