

Data Specialist

Netflix Exercise

*Please write up your responses according to this template.
To download the template, click File → Download*

1. Identify the English TV show with the most appearances in the top 10 list (you can treat each row in the data as a separate appearance). What were the average weekly viewed hours for that show across all appearances?

Title	Average Weekly Hours Viewed
You	43193333.33

2. For the "Films (Non-English)" category, identify the film with lowest IMDb rating. What were the average weekly hours viewed for that film?

Title	Average Weekly Hours Viewed
'Words Bubble Up Like Soda Pop'	'2180000.00'

3. Identify the film in the "Films (English)" category with the most cumulative weeks in the top 10. How could you approximate how many users watched this film? What assumptions would you make? What risks are there to your approach?

Please limit your response to 150 words or less.

Red Notice has the most cumulative weeks in the top ten for English Films, with 449.9 million of total weekly viewed hours.

Assuming that all users watched the film in its entirety only once by themselves, we can utilize the total weekly viewed hours across the weeks and divide the number by the film's total runtime, which is 118 minutes, or 1.97 hours. Approximating 228.4 million users!

The risk of this approach:

- We assume that each hour view count is generated by one user, however some people watch films with other people. Our calculation highly inflates the number of user counts

- We assume that each “user” watched the film in its entirety only once. Some users may not have finished the entire runtime of the film, or some users may have watched the movie several times, thus double counting the users.

4. If you plot weekly hours viewed over time (as an aggregate and for each of the four categories), what trends do you notice?

Please limit your response to 150 words or less.

After plotting the plot, we can see that non-English films has the lowest weekly hours viewed among all the category. On average, English TV shows has the highest weekly hours viewed, with the occasional dip. In the week of October 3rd, 2021, non-English TV shows peaked in weekly hours viewed, well above the usual average weekly hours viewed. Upon further investigation, we found out that the jump in viewership for non-English TV was due to the sudden popularity of the Korean show *Squid Game*.

You can reference the visuals made for this question in the below link:

[Line Graph](#)

[Bar Graph](#)

5. Another key investor question is how many US subscribers Netflix has each quarter. Name one type of dataset you could use to answer this question. How would this data source help you estimate Netflix's US subscribers?

Please limit your response to 150 words or less.

A dataset I could use to answer the above is using the NFLX Top 10 dataset. We can estimate Netflix's US subscribers by focusing on categories that are more relevant and have the most impact on US subscribers (English films and English TV shows). With the categories we'll aggregate it with weekly hours viewed. By summing these hours for each quarter, we can estimate total hours watched by US subscribers. We then convert this data into subscriber estimates based on average viewing habits. Joining the data with IMDB ratings can also help provide additional context for these estimates, such as indicating the perceived quality or popularity of titles among viewers, helping us refine our estimation of US subscribers by considering viewer preferences and behavior.

6. List three reasons why our web-scraping methodology may be inaccurate.

Reasons why the web-scraping methodology may be inaccurate:

- 1) We assume that Netflix will continue to publish the top 10 list each week. However if there are delays between the time when Netflix publishes its data and when the web-scraping occurs, this can cause discrepancies between the actual data and the data collected

- 2) As we utilize datasets scrapped from different sources and websites, we may run into format matching problems. Show/films titles may be labeled or formatted differently across sources, which leads to inaccurate ratings and runtime data being attributed to the wrong titles or not being able to capture at all
- 3) Even though Netflix is a global company with subscribers from all over the world, there are certain contents that are region-locked or available to specific user groups only due to specific licensing terms. Given this, data may not accurately represent the broader viewership trends if we are taking a global glance at the data.

7. What is your undergraduate GPA, both overall and for your major?

If you do not have easy access to this, that is acceptable, but please note that in your response to this question!

Overall GPA	Major GPA
3.6	N/A (Major GPA was not provided by my university)

8. What are the scores for your undergraduate standardized tests (eg. SAT) and graduate standardized tests (eg. GRE)? Please break out the scores into individual sections (eg. Math, Reading, etc.).

If you did not take an undergraduate or graduate standardized test (or do not have easy access to your results), that is acceptable, but please note that in your response to this question!

I do not have easy access to the result

9. Please clarify the level of proficiency in any coding skills that you have. (Note: coding skills are not a prerequisite for the position. If you do not possess any coding skills, please write "NA")

SQL: intermediate
Python: intermediate
R: beginner
Tableau: intermediate