

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	97185000.0

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
Nobody Sleeps in the Woods Tonight 2	4610000.0

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 – The movie with the most cumulative weeks in the top 10

Assumptions:

1. Let us assume at least 30% are duplicate views watched multiple times
2. All of the users watched the movie completely
3. Only one user is using a Netflix user account

Risks:

4. The number of users who watched the movie again could be different
5. users who watched the movie partly could be missed view hours could be lower
6. If multiple people are using the same user account then view hours could be higher

Total Weekly hours viewed = 449900000 view hrs

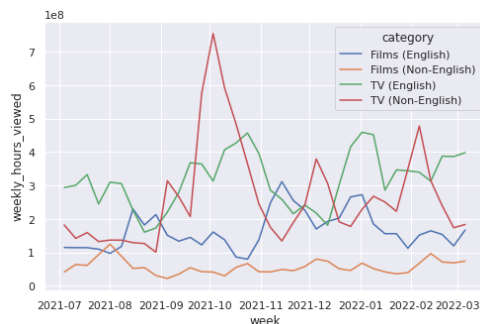
Total runtime of the movie = 1.967hrs

Total views = Total Weekly hours viewed / 1.967 = 228m views (include duplicate views)

Non-Duplicate Views = 160m ~ Users watched the movie red notice

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.



Trends evident from the plot

1. Non-English Films seem to have the lowest weekly hours viewed compared to the other three categories. Lower budget investments can be made in this category
2. Non-English TV series vary drastically compared to other categories. So It might be risky to invest in the same. Also sudden and peculiar rise during the week of 2021-10
3. English TV series have higher weekly hours views than other categories. The higher viewership is more or less constant with very less variation
4. Though the Popularity of English films was higher than Non-English. But it peaked soon after 2021-11.

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.

Monthly Payments dataset for a Year (could be confidential but accurate)

The payments data could have a subscriber Id, payment date, geography, Order Id

From this data, I would create a quarter column to identify the quarter each payment belongs to.

If we have monthly payment data, we can filter the US subscribers for the month and count the distinct subscribers for a month to provide the exact number of US subscribers available.

If Information for each quarter is required, First we need to filter US geography then we need to group data by quarter and count the distinct subscribers in each quarter.

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

1. The Web scraping frequency of the IMDB rating is unknown, the rating must be scraped together or immediately after Netflix data every week
2. The Geography is not specified. Though it's known that the data is global but it better specify that within the dataset
3. Most data seem to add 3 days after Netflix publishes its list and is not updated again. There are chances Netflix updates the data later on. A checking process could be relevant here to make sure the data scrapped is always updated.
4. Most Data from 7/4/2021 to 1/2/2022 has been scrapped on 1/4/2022 together in bulk which is okay but if there are updations to the Netflix website those updations might be slipped by the team.

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

Overall GPA	Major GPA
7.78/10	7.78/10

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 that is acceptable but please note that in your response to this question.

GRE Total 306, 154 Quantitative, 152 Verbal

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")

Python 9/10
R 7/10
Scala 6/10
Shell Scripting 9/10

10. When is the earliest you can start a new position? If you do not have a date in mind, please share with us what kind of notice you would need to give your current employer. (Note: your desired start date will not affect your candidacy)

Feb 13 2023