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	43,193,333

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
Der Fall Collini	2,410,000

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?

The English Film with the max cumulative weeks in the top 10 is Red Notice. I could convert the average weekly hours the film was viewed, 34,607,692 to minutes, multiply the minutes by the cumulative weeks it was in the top 10, 13 and divide the product by the runtime, 118 minutes to approximate how many users watched the film, 228,762,712. This is assuming that every user only watched the film one time, every viewer completed the film in its entirety, and that every user watched the movie using their personal Netflix subscription. This approach is risky, because People tend to watch movies they enjoy more than once, and given that this movie was in the top10 for largest number of weeks, it was likely viewed more than once by millions of users. It’s risky to assume that every user watched this movie in its entirety. That is impossible, because the actual number of users I calculated came out to 228762711.8644068 which is not a clean whole number. Movies are often enjoyed with friends and family, so it is unlikely that every user watched the movie using their own personal account, but with a loved one’s.

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.

- **Non-English TV experienced the highest increase in hourly viewership From September 2021 to October 2021 ($\approx 800,000,000$ hours) This was the highest monthly increase in viewership recorded among all the categories from July 2021 to March 2022.**
- **Non-English TV experienced the highest decrease ($\approx 600,000,000$ hours) in hourly viewership between October 2021 and November 2021. This was the highest decrease among all the categories from July 2021 to March 2022**
- **From July 2021 to March 2022: Non-English Films experienced the lowest overall average viewership**
- **English TV experienced the second highest overall hourly viewership of all of the categories from July 2021 to March 2022.**
- **English films experienced its highest increase in hourly viewership ($\approx 300.000.000$ hours) from mid-October 2021 to mid-November 2022**

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.***

Quarterly, Netflix releases its *financial reports* which include subscriber metrics like the number of paid subscribers broken down by region. Based on this report and previous financial reports you could predict what the trend should be for the next quarter.

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

1.) **It does not account for users who watch shows and movies with other users' accounts and people who are not users who utilize another person's account to watch a production.**
2.) **The categories: TV (English), TV (Non-English), Films (English), Films (Non-English) are extremely broad and may lead someone to make incorrect assumptions about the popularity of English and Non-English productions especially since non-English encompasses every other language offered on Netflix.**
3.) **Films and TV shows have varying properties, so some column titles can be misleading like "show_title" which also includes film titles and particularly, "season_title" which creates a lot of null values for Films since they don't have seasons. Null values typically make analysis more difficult and less precise.**

6. 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
3.6/4.0	3.4/4.0

7. 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!

Overall Score: 26 (Don't have access to the individual scores)

8. 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 (Pandas, TensorFlow, MatPlotLib, PyMySQL) – Proficient, SQL – Proficient, Java (JUNIT, OCSF, JDBMS) – Proficient, C++ - Proficient Progress ABL – novice, JavaScript – Novice, HTML – Proficient, CSS - Novice