Data Specialist

Netflix Exercise

Please write up your responses according to this template.

To download the template, click File → Download

Note - Be sure to exclude the week of the outage

when answering questions 1, 2, and 3

Part 1

1. In the 'Films (English)' category, which film has the most appearances in our data set (NFLX Top 10 tab of the Sheet)? What were its average weekly hours viewed?

Title	Average Weekly Hours Viewed
Sonic the Hedgehog	Average 7.481.250 weekly hours viewed

2. In the 'Films (English)' category, which film has the lowest IMDb rating? What were its average weekly hours viewed?

Title	Average Weekly Hours Viewed
365 Days: This Day	Average 29.022.500 weekly hours viewed

3. In the 'Films (Non-English)' category, which film has spent the most weeks in the top 10? To estimate the number of users who watched this film, what assumptions would you make and what risks are involved?

Please limit your response to 150 words or less.

The film *Through My Window* has spent the most weeks in the top 10. Based on calculations from the runtime dataset, this film has been watched 775,862 times. Assuming it was watched once by every single user, it had 775,862 users.

4. What are the risks of ignoring the data from the week of May 22nd when calculating the metrics from the previous questions?

Please limit your response to 150 words or less.

The week of May 22nd was removed due to the issue with the weekly_hours_viewed field. This removal could affect the analysis if a film during this week was on the list of most-watched films, or if a film from this week had more weeks in the top 10.

5. While we've indicated that the 'weekly_hours_viewed' data for the week of May 22nd cannot be used in our estimates, we may want to fill in this missing information for other analyses. As a Data Specialist, what methodology would you propose to estimate the 'weekly_hours_viewed' for this missing week?

Please limit your response to 150 words or less.

One of the first approaches would be to examine the behavior of the weekly_hours_viewed data before and after the missing week. Using the mean or median of adjacent weeks, I could generate an estimate for the week of May 22nd. If there is a linear behavior, it could be filled with a linear regression prediction.

Part 2

6. Do you think the information presented in this report is positive or negative for NFLX as a company? Why?

Please limit your response to 300 words or less.

Positive, as NFLX has the most loyal subscribers, who have subscription priority with its platform. Furthermore, the data shows that there is historically a good acquisition of former subscribers who were churners in the past. Recently, the rate of NFLX churners joining competing services reached its lowest level, which means that it is still the absolute leader in the competition.

7. List 3 questions our clients may have after reading the report.

What is the average time a subscriber stays on NFLX before churning?
What is the churn rate from NFLX to Paramount+, considering the recent subscriber growth experienced by Paramount+?.
What is the correlation between title launches and the churn rate? Which titles have the highest churn rate after launch?

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 – 10/10
SQL – 9/10