Data Engineer Take Home Task

Assignment

1. **Construct a normalized data model for the dataset provided.**
2. **Load the data into the data model you constructed and into DB engines like** MySQL/SQLite/Postgres or some free online one using the ELT approach (Extra points if you use Python and SQL in combination to achieve this task). How would you make such loading repeatable on a regular basis?
3. **What control measures could be deployed here to ensure the correctness of data?** Do you notice something odd about the dataset provided?
4. **Write queries for answering the following questions:**

a. The most profitable country for Netflix.
b. The most popular packages per country.
c. Which country has the potential for improving earnings if Netflix starts charging subscribers an additional fee for sharing Netflix households outside of their own?
d. A report showing the popularity of Movies and Series in different customer segments and the device used to consume, across the different markets the company operates in.

Additionally, extra points are awarded if you bring out the reasons for your choices.

Glossary:

User ID: Customer ID of the person who made the initial contract
Subscription Type: Basic/Premium/Standard
Monthly Revenue: Fee receivable for the given subscription type
Join Date: The initial sign-up date for the customer
Last Payment Date: The date when the last payment was made
Country:
Age:

Gender:

Device: The device used to consume the content
Plan Duration:
Active Profiles: The number of profiles associated with one contract
Household Profile Ind: Shows if all the profiles in an account is from the same household or not (i.e. one house or not)
Movies Watched : The number of movies watched
Series Watched: The number of series/sitcoms watched