For the dataset, the feature engineering for the three last columns:

1. Commercial Account

We assigned the account to be commercial based on the duration that the user spent on the day. So, if it is higher or equal than 16 hours, the account is commercial, because the avg time of sleeping is 8 hours, if we took it out from 24 hours, we would have 16 hours.

And the reason behind this because the data did not include detailed information. And we thought about it thoroughly:

* If the user is subscribed to a package, that allow multiple users in the same session from different devices.
* If the session is cumulative, so if two users with the same user id, the session will be summed together, that is why there is one user with 103 hours.
* The data was manipulated by STC, and it is not correct.
* If the session starts, it will not count the next day start. So, if it started on Sunday, it would add the session duration even if Monday starts, and it will not count the next day.
* There is no IP address to find whether the user entered from another device.
* The session duration minimum value is 3 sec, and this means that the user is just going through channels.

1. Family Account

Is the same as the commercial column, however the duration is higher than 8 hours and less than or equal 15, because of the avg sleeping time which is 8 hours.

1. Score

The score is based on the duration in the genre, so we sum the whole hours in genre, and then sum the hours for each movie, then divide the hours in each movie by the total duration in the genre multiply by 100.