README

- Data is organized in a single csv-file
- Columns that were not deemed interesting by the data processing team were removed. That includes information from both Spotify and Reccobeats.
  - We kept the column 'is.local' because we thought it was about whether the song was from the country the account was created in. However it seems like it is about whether the song is saved locally on your device which didn't happen in any of the experiments. So you can disregard this column as well.
- All songs from a single experiment have the same ID in the column 'user'. That also means that the same song can be listed in multiple rows!
- The column 'position' contains the position the song was played in for the individual experiment. The position was obtained by sorting the songs by the column 'played_at'. Since some people reported some issues with the listening times in the column 'played_at', we cannot guarantee that the column 'position' contains the true order the songs were played in. The position of the first song is 0.
  - The songs from a single experiment might not be in the right order. Be sure to sort them according to their 'played_at' or their 'position'
- Since in some instances one or both of the start songs were missing, we did the following:
  - we removed the start songs from all experiments if they were present
  - then we added the respective start songs to all experiments
  - the playing times were set to October 1, 2025 1 AM and 2 AM to guarantee that their positions are 0 and 1
  - because of that the number of songs per experiment varies between 50-52
- Besides a few sanity checks (50-52 songs per user?, number of songs per age group/gender/genre about the same?) we did not run any comprehensive integrity checks
  - as we wrote this we noticed that for some users they have apparently listened to the same song twice. We don't know how this could be...