

### Deliverable 3 - Backup Datasets

<https://www.kaggle.com/datasets/joy11117/customer-churn-dataset>

The Customer log dataset is a 12.5 GB JSON file and it contains 18 columns and 26,259,199 records. There are 12 string columns and 6 numeric columns, which may also contain null or NaN values. The columns include userId, artist, auth, firstName, gender, itemInSession, lastName, length, level, location, method, page, registration, sessionId, song, status, ts and userAgent. As evident from the column names, the dataset contains various user-related information, such as user identifiers, demographic details (firstName, lastName, gender), interaction details (artist, song, length, itemInSession, sessionId, registration, lastinteraction) and technical details (userAgent, method, page, location, status, level, auth).

The data would be a good backup for customer churn analysis.

[https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data?select=movie\\_titles.csv](https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data?select=movie_titles.csv)

This dataset is as a result of the Netflix competition. The goal was to create a model that was effective at predicting customer ratings. Included is a testing data set and a training dataset that Netflix has provided. This could be good data to test different recommendation models and pivot away from customer churn.