

# Cleaning

We remove songs with duplicate track\_id.

```
spotify_songs <- readr::read_csv("raw_spotify.csv", show_col_types = FALSE)
length(spotify_songs$track_id)
```

```
[1] 32833
```

```
spotify_songs = spotify_songs[!duplicated(spotify_songs$track_id),]
length(spotify_songs$track_id)
```

```
[1] 28356
```

```
write.csv(spotify_songs, "cleaned_spotify.csv", row.names = FALSE)
```