

# Data Loss - Songlyrics

Patrick Zivkovic

October 15, 2022

## Not all albums were considered

Due to the limit = 50 parameter of the get-artist-album function we did not extract all albums. With the offset parameter we were able to correct this and produce about 2500 more matches, thus increasing the size of our dataset from 34116 to 36932. (originally 57650 songs - so we are working with about 64% of the full dataset).

## Matching: Stringdist = 2

Had to make a correction on 5994 titles. Mostly they were just punctuation errors as the following example shows.

artist	original_songname	songname
ABBA	Ahe's My Kind Of Girl	She's My Kind Of Girl
ABBA	Gonna Sing You My Lovesong	Gonna Sing You My Love Song
ABBA	Hey Hey Helen	Hey, Hey Helen
ABBA	Knowing Me Knowing You	Knowing Me, Knowing You
ABBA	Money Money Money	Money, Money, Money

## Problems with appendages

Most songs were lost to us because the song was only available in Spotify with an appendage such as "Live", "Outtake", "Single" or something similar. For example, the song "Love Isn't Easy" by Abba appears in this form in the original record, but in Spotify only as follows:

```
## [1] "Love Isn't Easy (But It Sure Is Hard Enough)"
```

Another example is the Beatles' song "All I've Got To Do".

```
## [1] "All I've Got To Do - Remastered 2009"
```

## Availability

The rest of the songs were not matched because they are actually not available in the Spotify database. For example, some songs from ABBA ("Burning My Bridges" or "Dream World").