



# Billboard Songs Dataset

MLESTONE 1 - DATASET PREPARATION

# Obtaining and preparing datasets

- ▶ Primarial data was obtained from the billboard dataset at data.world to songs\_1970\_2018.csv
- ▶ songs\_1970\_2018.csv file was then merged with data obtained by crawling the Genius API in order to obtain artist.csv, album.csv and lyrics.csv files, and indexed so to be easily merged with all files.
- ▶ Datasets are then stripped of null values, duplicates and column names were normalized.

# Characterizing the datasets

song\_1970\_2018.csv:

- Has all the songs from Billboard
- More than 20 000 songs
- Contains Song and Artist Names, Peak Position, N° Weeks, etc.

artists.csv:

- Has information about the artists
- Has around 6 700 artists
- Contains Artists Name and Aliases and a short description

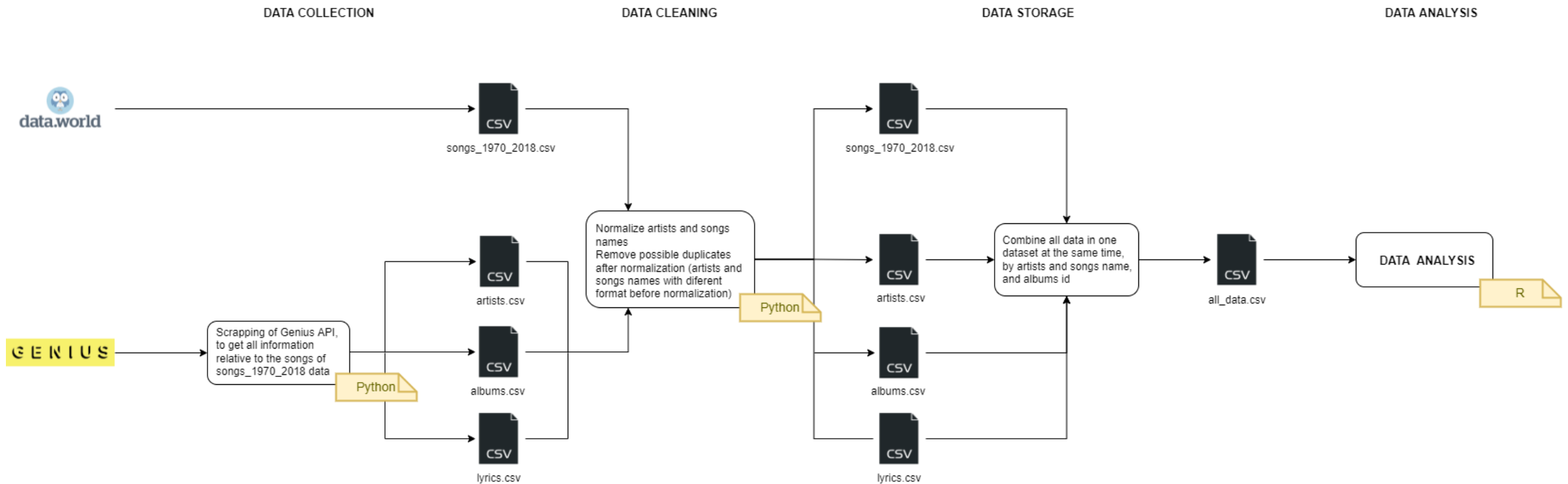
lyrics.csv:

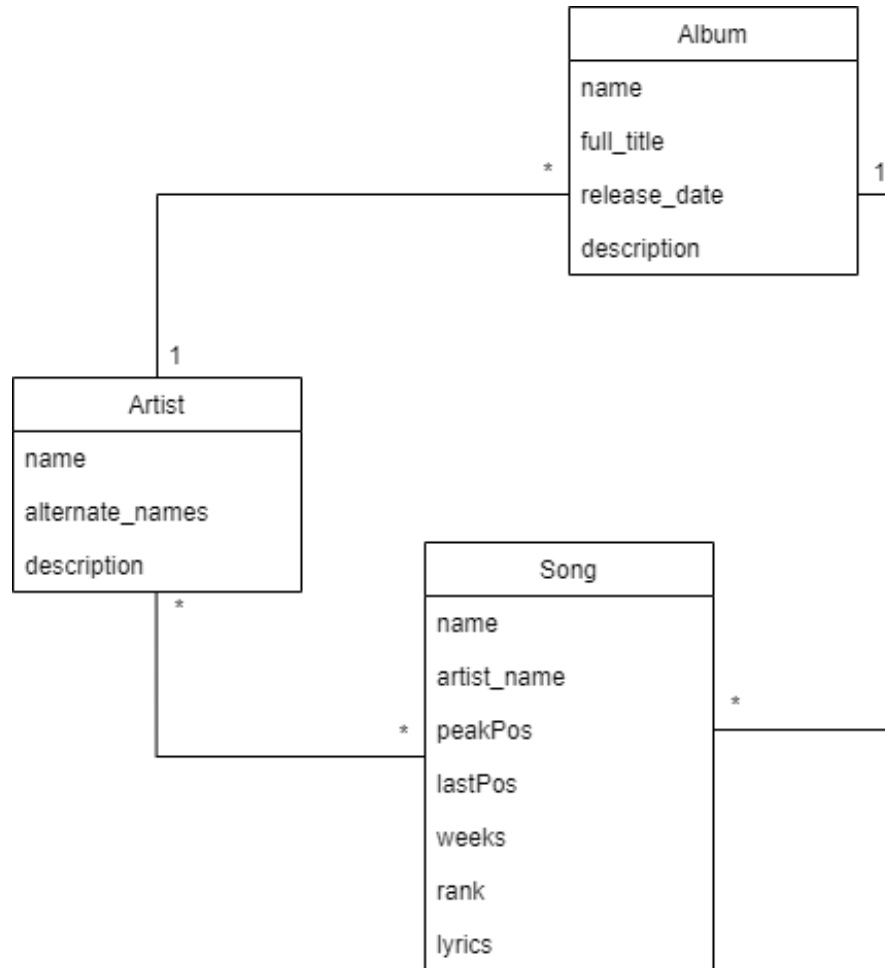
- Has the lyrics to the songs
- Contains Song Name and Lyrics

albums.csv:

- Has all the albums of the songs from Billboard
- Has around 3 500 albums
- Contains Album Name, Title, Release Date and Description

# Data processing pipeline





# Conceptual Model of the Domain

# Search Topics

The data search engine will consult a set of songs, originated between the years 1970 and 2018, their lyrics, albums they belong to, and the artists who produced them. For a more detailed search, you'll be able to filter the results based on year, artist, among other filter options.

The engine will have the capacity to:

- Search for songs
- Search for artists
- Search for albums
- Search for quotes of letters
- ...

Table 1. Top 10 Billboard musics

Song Name	Artist Name	N° of weeks
Radioactive	Imagine Dragons	87
Sail	AWOLNATION	79
I'm Yours	Jason Mraz	76
How Do I Live	LeAnn Rimes	69
Counting Stars	OneRepublic	68
Party Rock Anthem & LMFAO Featuring Lauren Bennett	GoonRock	68
Rolling In The Deep	Adele	65
Foolish Games/You Were Meant For Me	Jewel	65
Before He Cheats	Carrie Underwood	64
Ho Hey	The Lumineers	62

Table 2. Top 10 artists on the Billboard

Artist Name	N° of songs
Glee Cast	183
Taylor Swift	67
Drake	63
Elton John	59
Madonna	57
Tim McGraw	50
Rod Stewart	48
Kenny Chesney	46
Chicago	44
Billy Joel	43

Table 3. Top 10 albums on the Billboard

Album Name	N° of songs
Ulysses	46
Finnegans Wake	28
Glee: The Music, The Complete Season Two	28
Glee: The Music - The Complete Season Three	23
More Life	21
beerbongs & bentleys	17
The Big Book of Song Lists	17
Starboy	16
Views	16
Purpose (Deluxe)	15

# Data Analysis

# Data Analysis

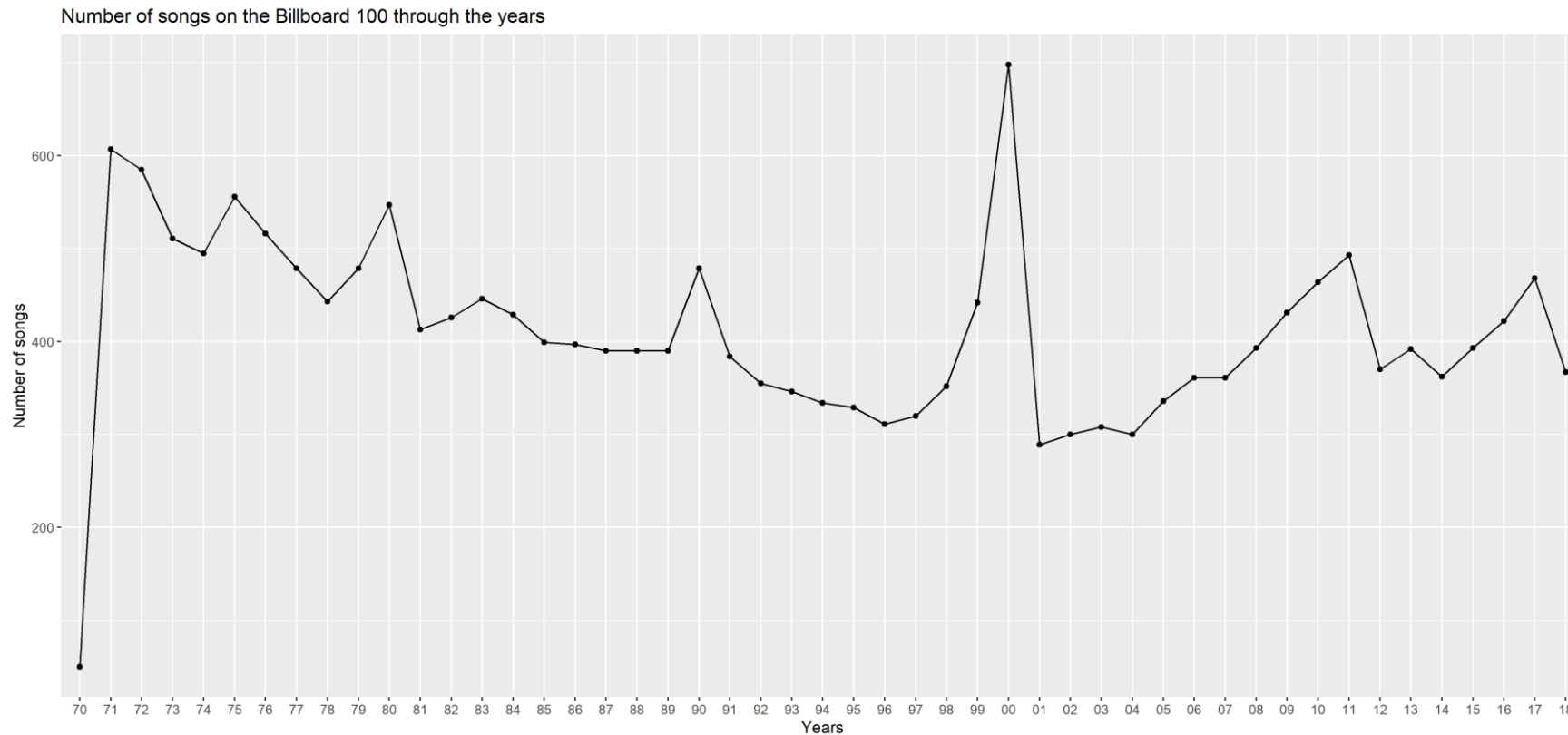


Figure 1: Evolution of the number of musics passing through Billboard over the years