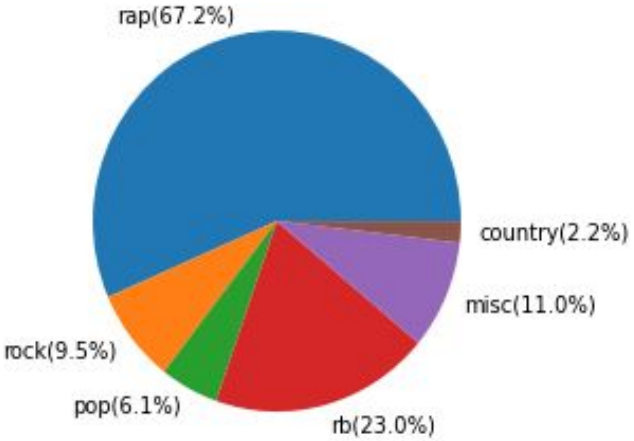
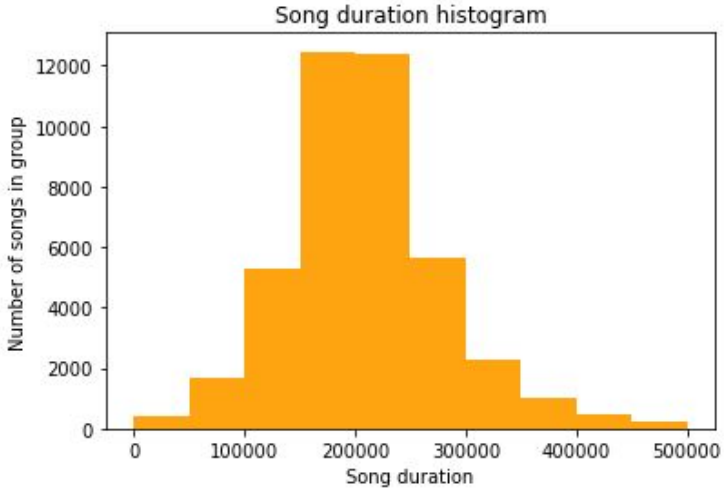
Spotify Track Statistics and Lyrics Analysis

An Information Processing and Retrieval Study

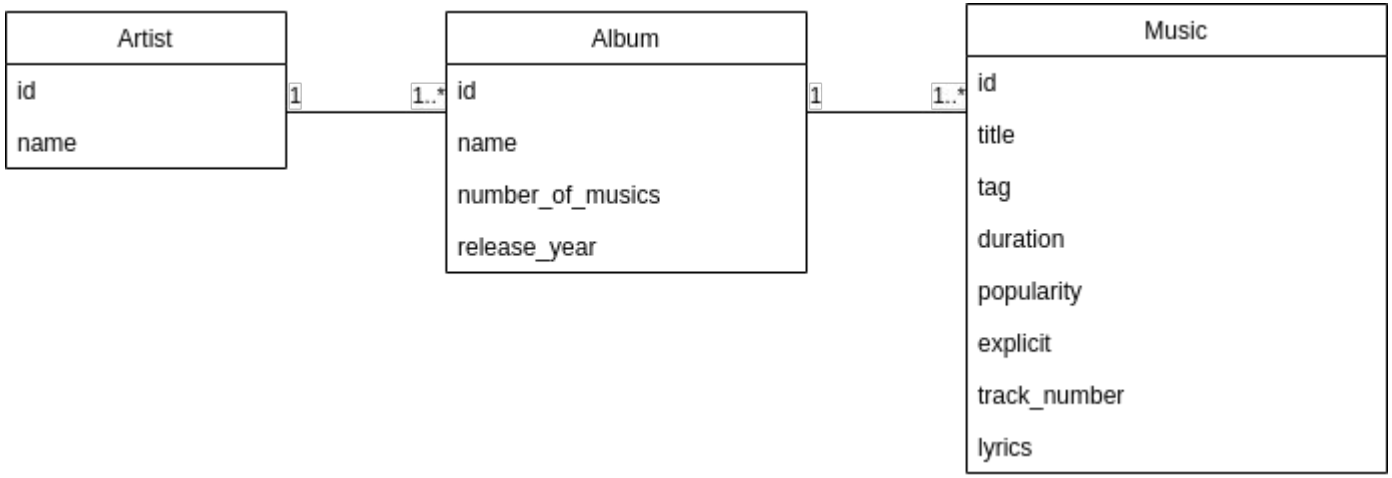
Introduction - Milestone 1

- Spotify Track Statistics and Lyrics Analysis
- Original dataset + Spotify API extra data
- Pipeline: duplicate removal, random trimming, column selection.





Conceptual Model



Documents

- File format: Json
- Tool used: Solr

All the text fields were indexed.

Some useful integer fields to search by range, or to sort and group the data, such as the popularity or album release year.

Format could be csv, but was already converted from previous delivery. This way it’s the same as the schema.

Indexing

Issue: The musics are global and have many different languages.

Most of the filters and tokenizers in Solr work differently according to the language passed on the arguments. Can’t have more than 1 language in a field.

Standard Tokenizer was used because there was no need for more complexity.

Filters used:

- ASCII Folding Filter
- Lower Case Filter
- Classic Filter
- Beider-Morse Filter

Retrieval

Different types of boosts applied:

- Field Boosts
- Term Boosts
- Phrase Match
- Wildcards
- Proximity search

Queries: a few examples

Scenario 1

- Searching for the remastered albums of The Beatles

Option	Values
q	artist: "The Beatles" album_name: Remastered
q.op	OR
fl	album_name, artist
qf	artist^2 album_name
mm	2
defType	dismax

Artist - Album	Relevance
The Beatles - Revolver (Remastered)	R
The Beatles - 1 (Remastered)	R
The Beatles - Live At The BBC (Remastered)	R
The Beatles - The Beatles - Inspirations	N
The Ballet - I Blame Society	N

Artist - Album	Relevance
The Beatles - Revolver (Remastered)	R
The Beatles - 1 (Remastered)	R
The Beatles - Live At The BBC (Remastered)	R

Scenario 1 - Evaluation

- Searching for the remastered albums of The Beatles

Metric	W/out boost	W/ boost
Average Precision	0.88	1
Precision at 5 (P@5)	0.60	–
Precision at 3 (P@3)	–	1

Scenario 2

- Searching for Brazilian Funk Songs

Option	Values
q	artist: Mc* lyrics: amor
q.op	OR
fl	lyrics, artist, title, album_name
qf	artist^3 lyrics
mm	2
ps	3
defType	dismax

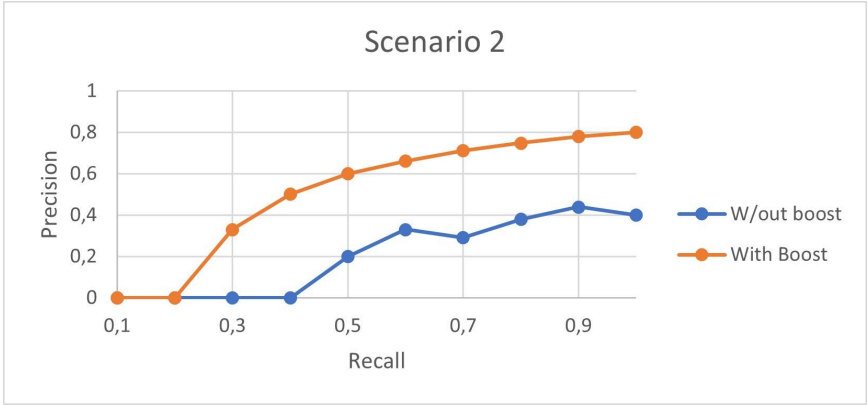
Artist - Title	Relevance
Mr Jukes & Barney Artist - Autumn Leaves	N
Lyrics Born - Before and After	N
Lyrics Born - ir Racha	N
Heartist OTOD - Tango In Paris	N
MC Lars - Download This Song	R
MC Cabelinho - O Preço	R
Ana Carolina - Dias Roubados	N
Chocolate MC - Aveces	R
MC Nathan ZK - Maloka Firmeza	R
Mario Biondi - Bom de Doer	N

Artist - Title	Relevance
Mr Jukes & Barney Artist - Autumn Leaves	N
Heartist OTOD - Tango In Paris	N
MC Lars - Download This Song	R
MC Cabelinho - O Preço	R
Chocolate MC - Aveces	R
MC Nathan ZK - Maloka Firmeza	R
MC Lars - Hurricane Fresh	R
Under MC - ¿Por qué Será?	R
Lis MC - Bad Bitch	R
Kid MC - Oração	R

Scenario 2 - Evaluation

- Searching for Brazilian Funk Songs

Metric	W/out boost	W/ boost
Average Precision	0.20	0.51
Precision at 10 (P@5)	0.40	0.80



Scenario 3

- Searching for summer songs

Option	Values
q	lyrics: (summer beach hot sea) title: (summer beach hot sea)
q.op	OR
fl	lyrics, artist, title, album_name
qf	lyrics^3 title
mm	2
ps	3
defType	dismax

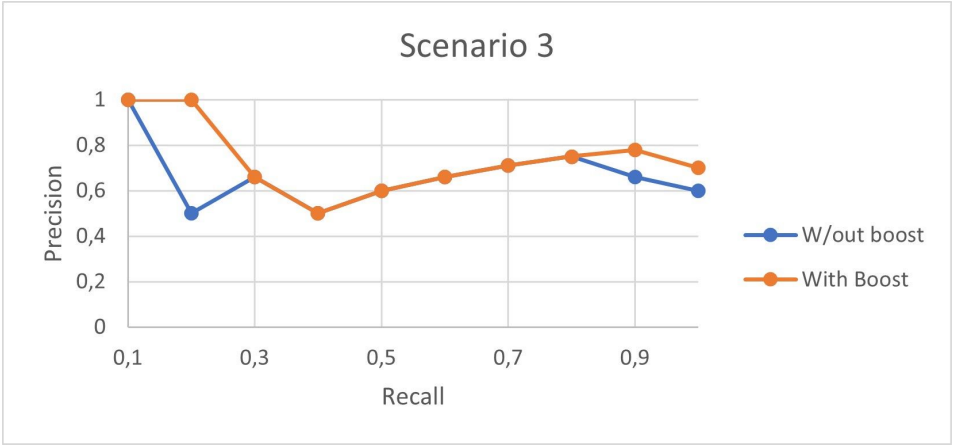
Artist - Title	Relevance
Flower Boy - Lemonade	R
Hewhocorrupts - Linguistic Violations	N
Louis Dunford - Summer in the Manor	R
James Bourne - Alone In Paradise	N
Frank Turner - Little Life	R
Diego Mar - Maloka Firmeza	R
burningforestboy - Days Before Summer	R
Spain - World Of Blue	R
Clutch - Red Alert Boss Metal Zone	N
punii+ - Battle	N

Artist - Title	Relevance
Louis Dunford - Summer in the Manor	R
Flower Boy - Lemonade	R
Hewhocorrupts - Linguistic Violations	N
James Bourne - Alone In Paradise	N
Frank Turner - Little Life	R
Diego Mar - Second Chance	R
SWENDAL - Sunrise In Miami	R
Diego Mar - Second Chance	R
Spain - World Of Blue	R
punii+ - Battle	N

Scenario 3 - Evaluation

- Searching for summer songs

Metric	W/out boost	W/ boost
Average Precision	0.67	0.74
Precision at 10 (P@5)	0.60	0.70



Conclusions and Future work