

2023-06-09

GENDER DIVERSITY IN POP MUSIC CHARTS

Matthieu Baillard



OVERVIEW

- INTRODUCTION
- DATA COLLECTION
- DATA CLEANING
- EDA
- MYSQL / ERD
- VIZ & HIGHLIGHTS
- MACHINE LEARNING
- DEMO
- CHALLENGES
- LIMITATIONS
- NEXT STEPS



INTRODUCTION

- WOMEN LARGELY OVERLOOKED FROM MUSIC HISTORY (*Hildegard von Bingen, Fanny Mendelssohn...*)
- FEMALE POP STARS NOW AT THE FOREFRONT OF POP CULTURE
- DOES IT ACTUALLY REFLECT THE CURRENT STATE OF THE INDUSTRY?
LET'S EXAMINE THE CHARTS DATA AND FIND OUT!



▶ PLANNING

Trello Espaces de travail Récent Favoris Modèles Crée Parcourir Partager

Ironhack - Final Project Public Tableau

PROJECT PLANNING

- Brainstorm topics
- Trello Planning
- Business case
- Github
- + Ajouter une carte

DATA COLLECTION

- Billboard "Hot 100" Songs
- French charts 84-2020 (web scraping)
- Gender & Artist info: MusicBrainz API
- Bio : Last.fm API
- Optional to enrich missing bios: Wikipedia API
- Optional - Wikipedia lists
- Optional - Spotify charts
- Optional - Most popular Spotify playlists?
- Optional - More niche genres?
- Optional - Top 500 albums on Rate Your Music (NO: albums and not songs)
- + Ajouter une carte

DATA PRE-PROCESSING

- Data Cleaning - WIP
- + Ajouter une carte

SQL

- Database selection
- Database design / ERD
- Export Dataframes to MySQL
- SQL queries scripts
- + Ajouter une carte

EDA

- Python
- Vizualization Tableau
- + Ajouter une carte

PDF REPORT

- Report - WIP
- Markdown PDF
- Markdown to PDF?
- + Ajouter une carte

MAI

- Rec
- Test
- Feat
- Vec
- PCA
- Nor
- Nea
- Uf
- + Ajouter une carte

Power-ups Automatisation Filtre Partager



DATA COLLECTION & CLEANING



BILLBOARD HOT100 1958-2022

billboard

	chart_position	chart_date	song	performer	song_id	instance	time_on_chart	consecutive_weeks	previous_week	peak_position	worst_position	chart_debut	chart_url
0	84	1990-05-05	"B" Girls	Young And Restless	GirlsYoung And Restless	1.0	1	NaN	NaN	84	84	1990-05-05	https://www.billboard.com/charts/hot-100/1990...
1	78	1990-05-12	"B" Girls	Young And Restless	GirlsYoung And Restless	1.0	2	1.0	84.0	78	84	1990-05-05	https://www.billboard.com/charts/hot-100/1990...
2	68	1990-05-19	"B" Girls	Young And Restless	GirlsYoung And Restless	1.0	3	2.0	78.0	68			
3	60	1990-05-26	"B" Girls	Young And Restless	GirlsYoung And Restless	1.0	4	3.0	68.0	60			
4	58	1990-06-02	"B" Girls	Young And Restless	GirlsYoung And Restless	1.0	5	4.0	60.0	58			

RangeIndex: 336295 entries, 0 to 336294
 Data columns (total 15 columns):

#	Column	Non-Null Count	Dtype
0	chart_position	336295	non-null int64
1	chart_date	336295	non-null object
2	song	336295	non-null object
3	performer	336295	non-null object
4	instance	336295	non-null float64
5	time_on_chart	336295	non-null int64
6	consecutive_weeks	302987	non-null float64
7	previous_week	302987	non-null float64
8	peak_position	336295	non-null int64
9	worst_position	336295	non-null int64
10	chart_debut	336295	non-null object
11	chart_url	336295	non-null object
12	song_id	336295	non-null object
13	main_artist	336295	non-null object
14	artist_featured	336295	non-null object

 dtypes: float64(3), int64(4), object(8)
 memory usage: 38.5+ MB

Source : github.com/HipsterVizNinja/random-data/tree/main/Music/hot-100



TOP 50 / TOP 100 FRANCE 1984-2020 (SCRAPING)



Position	Position Evolution	Artist	Song Title	Year	Week	URL Week ID
1	1	Entrée	Peter et Sloane	Besoin de rien, envie de toi	1984	Semaine du 26 octobre 1984
2	2	Entrée	Cookie Dingler	Femme libérée	1984	Semaine du 26 octobre 1984
3	3	Entrée	Scorpions	Still Loving You	1984	Semaine du 26 octobre 1984
4	4	Entrée	Chris De Burgh	High On Emotion	1984	Semaine du 26 octobre 1984
5	5	Entrée	Stevie Wonder	I Just Called To Say I Love You	1984	Semaine du 26 octobre 1984
...
152547	96	Entrée	Bon Entendeur	Le temps est bon	2020	Int64Index: 152551 entries, 1 to 152551 Data columns (total 8 columns): # Column Non-Null Count Dtype --- ----- 0 Position 152551 non-null int64 1 Position Evolution 152551 non-null object 2 Artist 152548 non-null object 3 Song Title 152548 non-null object 4 Year 152551 non-null int64 5 Week 152551 non-null object 6 URL Week ID 152551 non-null int64 7 song_id 152548 non-null object dtypes: int64(3), object(5) memory usage: 10.5+ MB
152548	97	Entrée	Patrick Fiori	Les gens qu'on aime	2020	
152549	98	-28	Jean-Louis Murat	Baby Love D.c.	2020	
152550	99	Entrée	Barbara Pravi	Voilà	2020	
152551	100	Entrée	Laurent Garnier	Coloured City	2020	
152551 rows × 7 columns						



ARTISTS METADATA & BIO

(APLS)



MusicBrainz

	Artist	MusicBrainz ID	Type	Gender	Area	Begin Date	End Date	Tags	Other Tags
0	Boris	57652bf8-cfe8-42e7-b9a7-5572a7080d8d	Group	NaN	Japan	1992	NaN	['ambient', 'dissonant', 'doom metal', 'drone'...]	□
1	Richard "Groove" Holmes	726cf69-c905-4161-a10c-accb13d9ec26	Person	Male	United States	1931-05-02	1991-06-29	['hard bop', 'jazz and blues', 'soul jazz']	□
2	'Little' Jimmy Dickens	bd7589a3-f82f-4c3d-b7e2-e57e89552da1	Person	Male	United States	1920-12-19	2015-01-02		□ □
3	Pookie Hudson	29dc009-015f-47c4-bd17-ed2af6d2ae0c	Person	NaN	NaN	1934-06-11	2006-01-16		□ □
4	"Weird Al" Yankovic	7746d775-9550-4360-b8d5-c37bd448ce01	Person	Male	United States	1959-10-23	NaN	['accordion', 'american', 'comedy', 'comedy ro...']	□
...
655	À la recherche de la Nouvelle Star	46acab74-83fe-4d7a-90cd-fca626f229fb	NaN	NaN	France	NaN	NaN		□ □
656	Ärsenik	201ac6bb-c268-4e9d-9390-5305512d5ee2	Group	NaN	France	1992	NaN		['hip hop']
657	É o Tchan	04611209-2171-4652-98cd-e2c5e536ac04	Group	NaN	Brazil	1992	NaN		□ □
660	Émilie Simon	01252145-c9e8-4de5-a480-9b2bed05450a	Person	Female	France	1978-07-17	NaN	['arranger', 'arrangeur', 'composer', 'composi...']	□
662	Ólafur Arnalds	6655955b-1c1e-4bcb-84e4-81bcd9efab30	Person	Male	Iceland	1986-11-03	NaN	['ambient', 'composer', 'contemporary classica...']	□

10754 rows × 9 columns



MBID	Bio
57652bf8-cfe8-42e7-b9a7-5572a7080d8d	There are at least two different artists using...
726cf69-c905-4161-a10c-accb13d9ec26	Richard Arnold "Groove" Holmes, Born Richard A...
bd7589a3-f82f-4c3d-b7e2-e57e89552da1	James Cecil Dickens (born December 19, 1920 - ...)
29dc009-015f-47c4-bd17-ed2af6d2ae0c	Thornton James ("Pookie") Hudson (born Des Mo...
7746d775-9550-4360-b8d5-c37bd448ce01	Alfred Matthew "Weird Al" Yankovic (born Octob...
...	...
8cda32b5-1447-4ded-8a0e-f1cf8e6b3480	Élodie Frégé won France's Star Academy in 200...
01252145-c9e8-4de5-a480-9b2bed05450a	Émilie Simon is a French singer and composer b...
1e0de31c-4957-4649-9aa3-7b0f1d9d2c84	Étienne Daho (born January 14, 1956 in Oran, A...
6655955b-1c1e-4bcb-84e4-81bcd9efab30	Born 1986 in the suburban Icelandic town of Mo...
6efbf07-3346-4d30-af85-0abe646b97ba	Odyssey is a name of at least nineteen acts:\n...

last.fm

```
artist_bios_df = artist_bios_df.reset_index().drop_duplicates(subset=['MBID']).set_index('MBID')
```

10754 rows × 1 columns



DATA CLEANING



✓ M+ Hot 100 Billboard (US 1958-2022)

M+ Import existing Dataset

✓ M+ Pre-processing

M+ Clean 'song_id'

M+ From 'performer', create 'main_artist' and 'artist_featured'

M+ TODO Finetune Artist cleaning

M+ Pre-processing

M+ Create song_id

M+ From 'Week' to 'chart_date'

M+ Handle numerical columns if needed

M+ Handle Missing values in Artist / Song Title / song_id

M+ From 'Artist' to 'main_artist' and 'artist_featured'

M+ Make the dataset consistent with the Billboard Hot100

M+ Remove duplicates (from 13495 to 10754 artists)

✓ M+ Import Genre list and sort Tags (Genre_tags / Other_tags)

M+ Convert Date columns to dates (currently year or date)

M+ Clean Tags

M+ Fetch Artist Bio from Last.FM API

M+ Import previously fetched bios

M+ Remove duplicates (from 13495 artists to 10754)

M+ Merging all Artist info

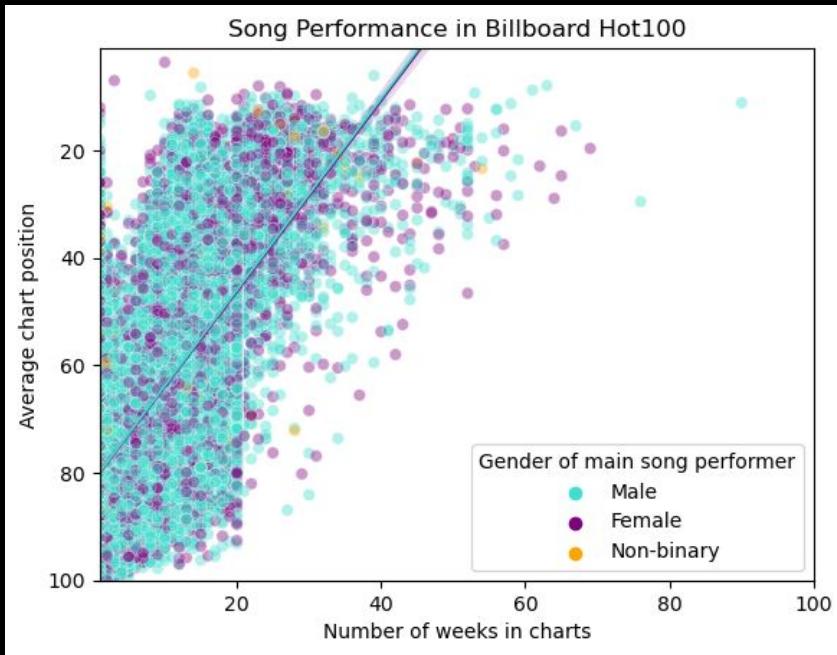
✓ M+ Data Cleaning

M+ Drop anonymous and homonymous artists

EXPLORATORY DATA ANALYSIS

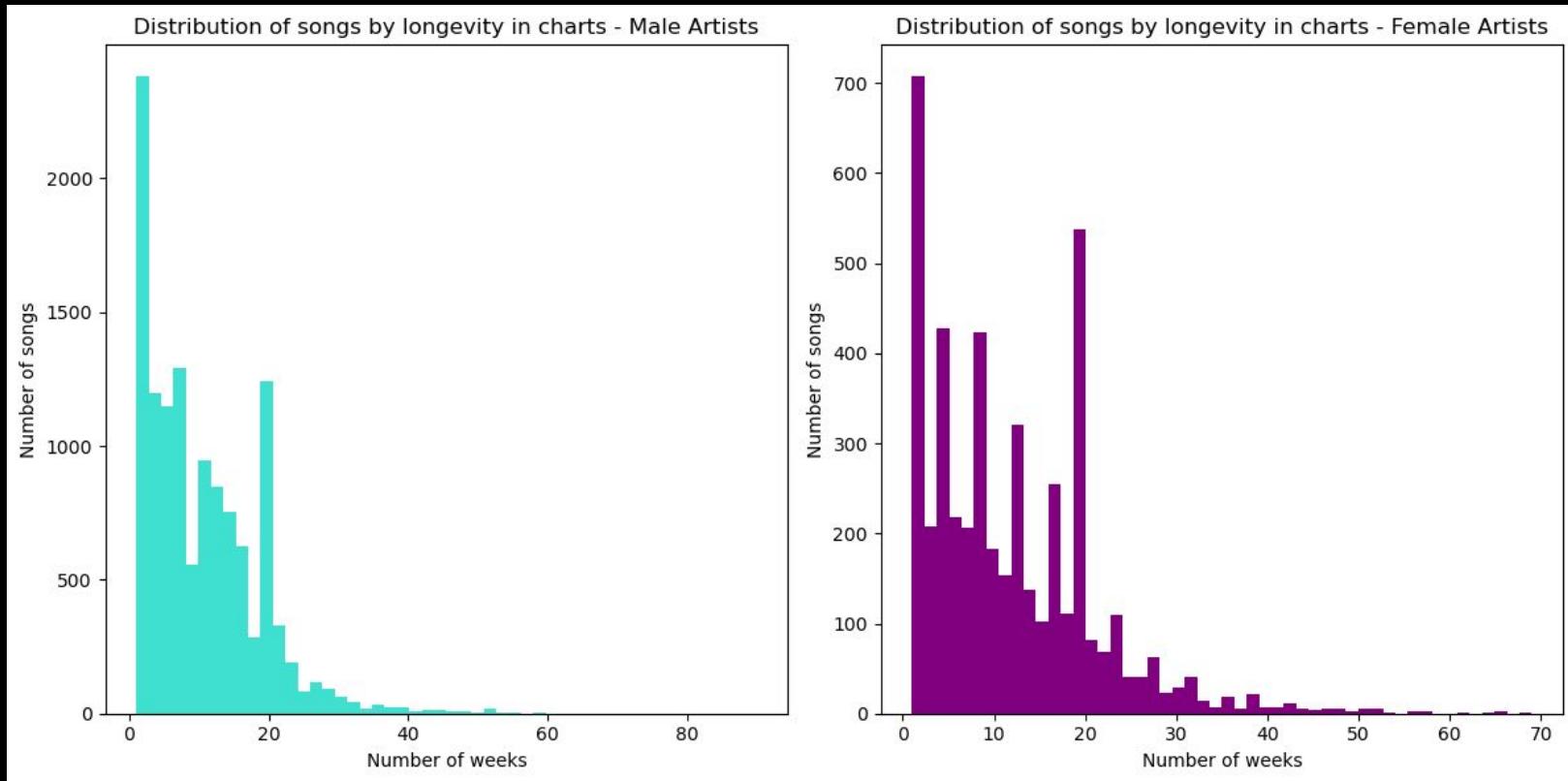


SONG PERFORMANCES BY GENDER



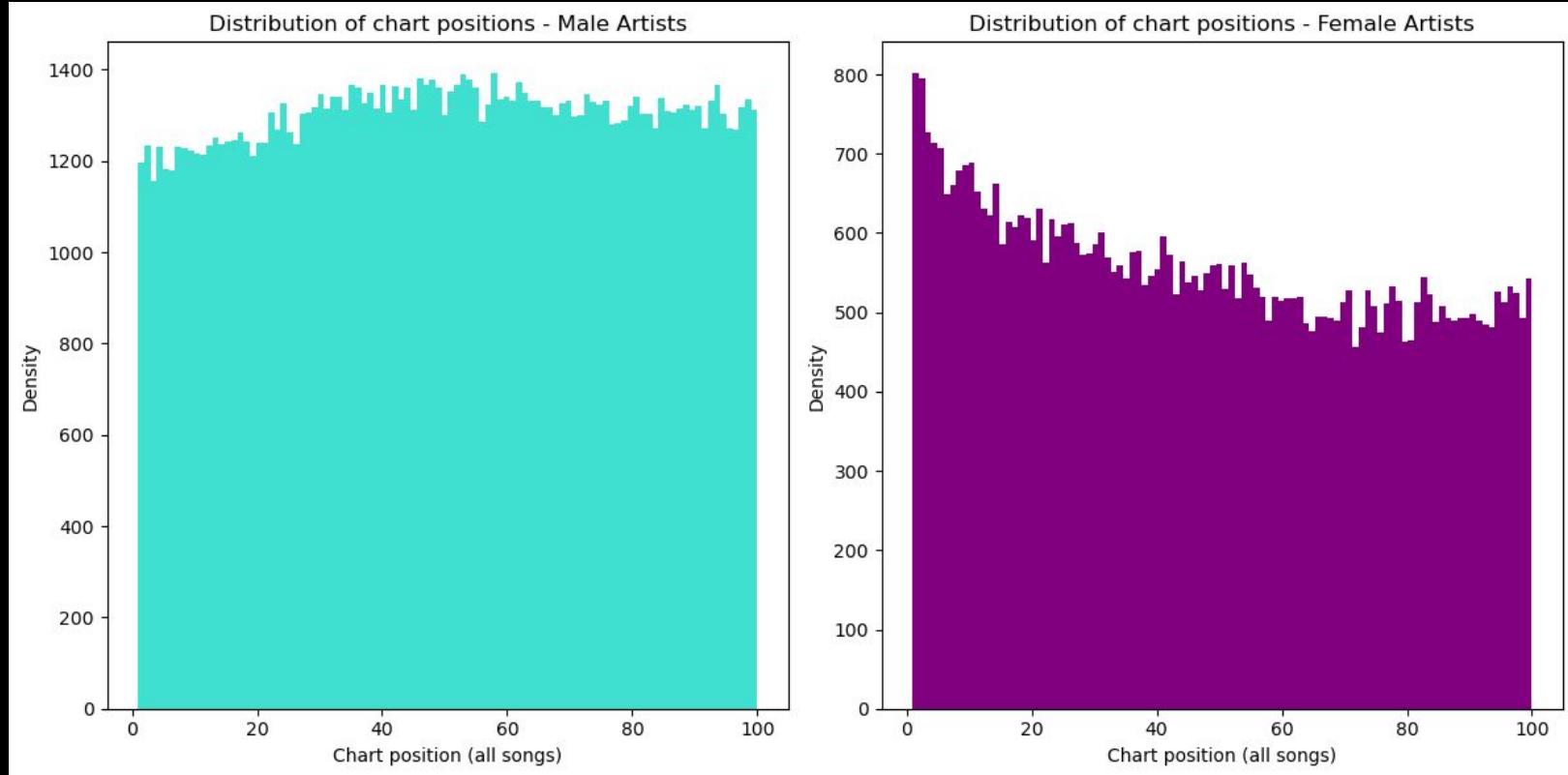


HOW LONG DOES A SONG STAY IN THE CHARTS?





CA FEWD WOMEN ON TOP OF THE CHARTS





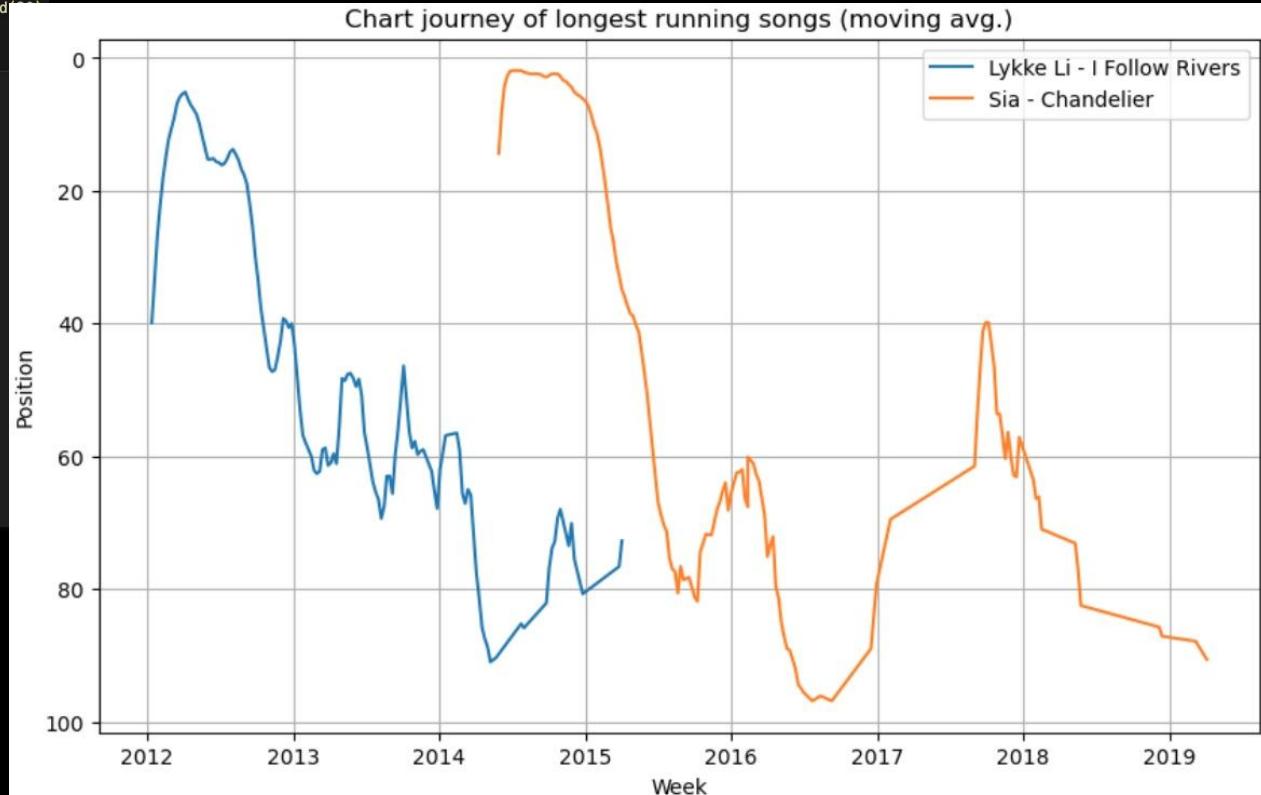
AN APPARENT BALANCE ON TOP OF FR CHARTS

```
Top20_FR_artist = FRcharts_df['performer'].value_counts().head(20)
Top20_FR_artist
```

Johnny Hallyday	1088
Mylène Farmer	1009
Madonna	977
Rihanna	807
Maitre Gims	796
David Guetta	711
Soprano	674
Céline Dion	612
Ed Sheeran	528
Lady Gaga	513
Britney Spears	505
Calvin Harris	472
Coldplay	467
Adele	466
The Black Eyed Peas	466
Indochine	459
Katy Perry	452
Queen	446
Sia	445
Kylie Minogue	444

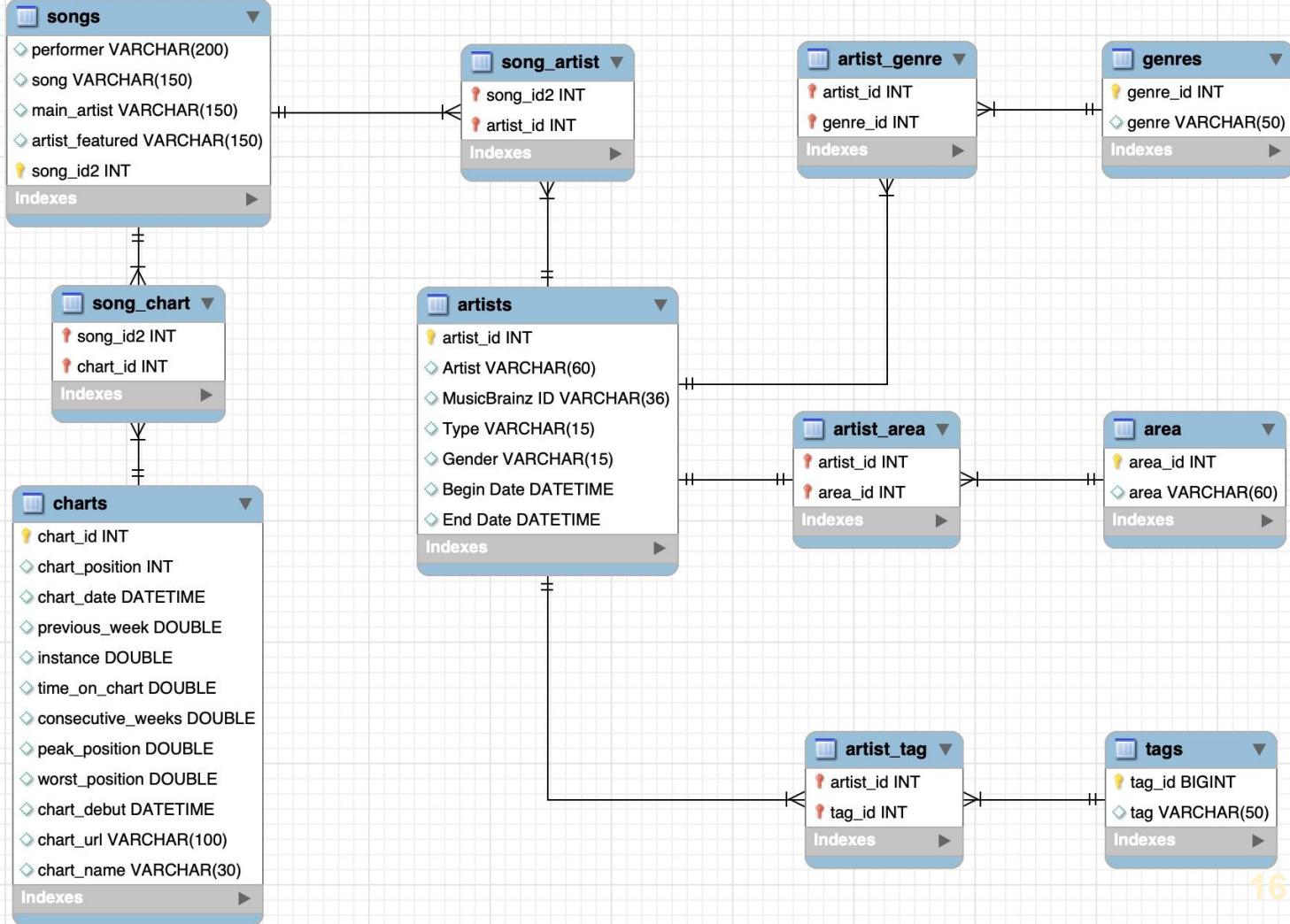
```
top_20_FR_songs = FR_song_counts.head(20)
top_20_FR_songs
```

chandelier-sia	143
i follow rivers-lykke li	138
who the fuck are arctic monkeys ?-arctic monkeys	123
alors on danse-stromae	116
petit papa noël-tino rossi	115
rolling in the deep-adèle	113
someone like you-adèle	110
shallow-lady gaga	109
shape of you-ed sheeran	106
nightcall-kavinsky	103



MYSQL DATABASE

ENTITY RELATIONSHIP DIAGRAM

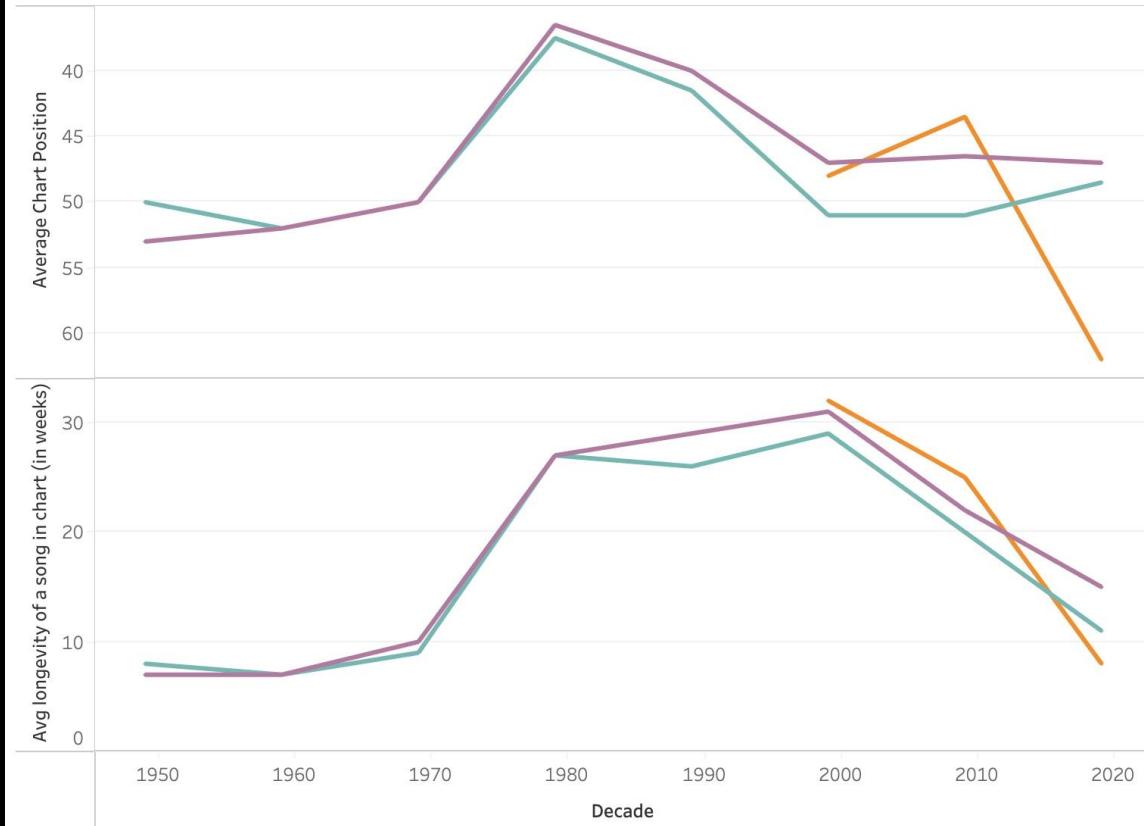


VISUALIZATION & FINDINGS



EVOLUTION SONG PERFORMANCES BY GENDER

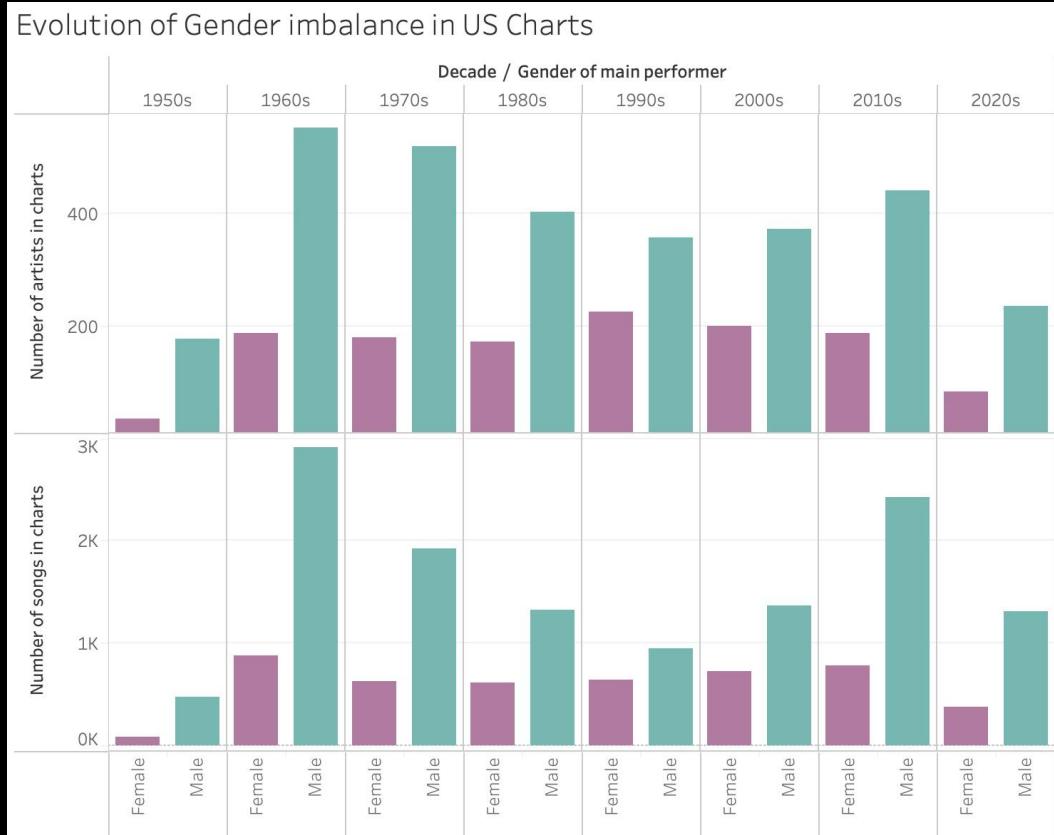
Average performances in US charts by gender of main performer artist



Gender of main perfor...
Female
Male
Non-binary

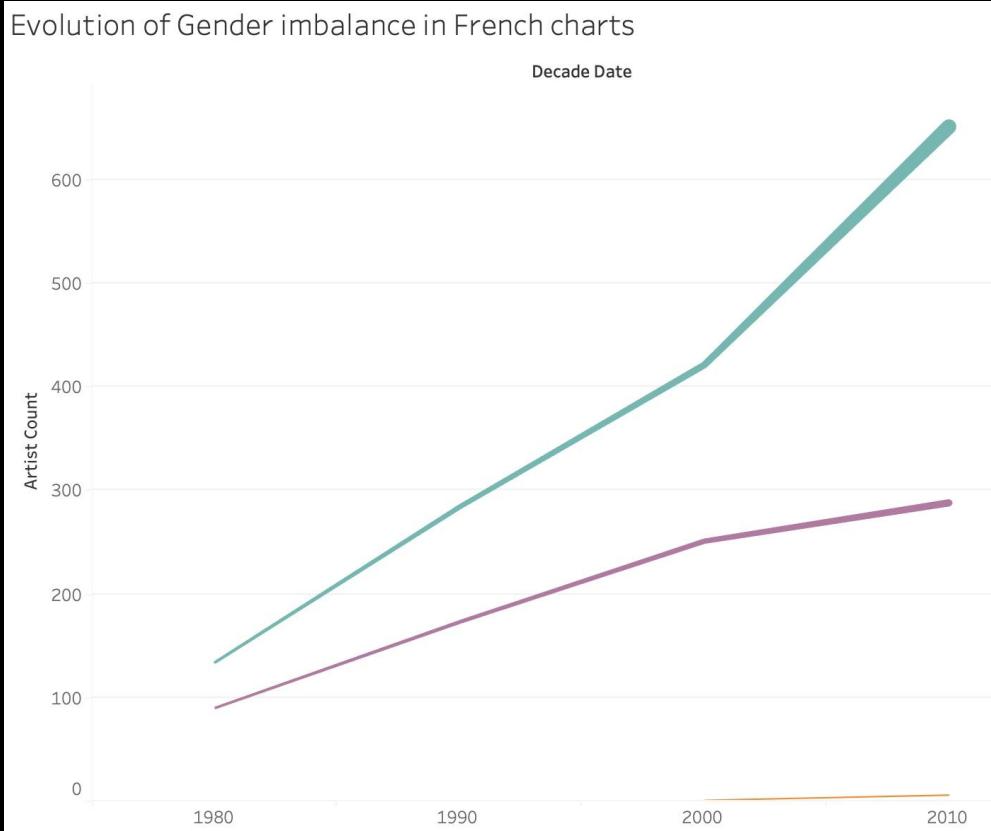


IT DOESN'T GET MUCH MORE BALANCED!!



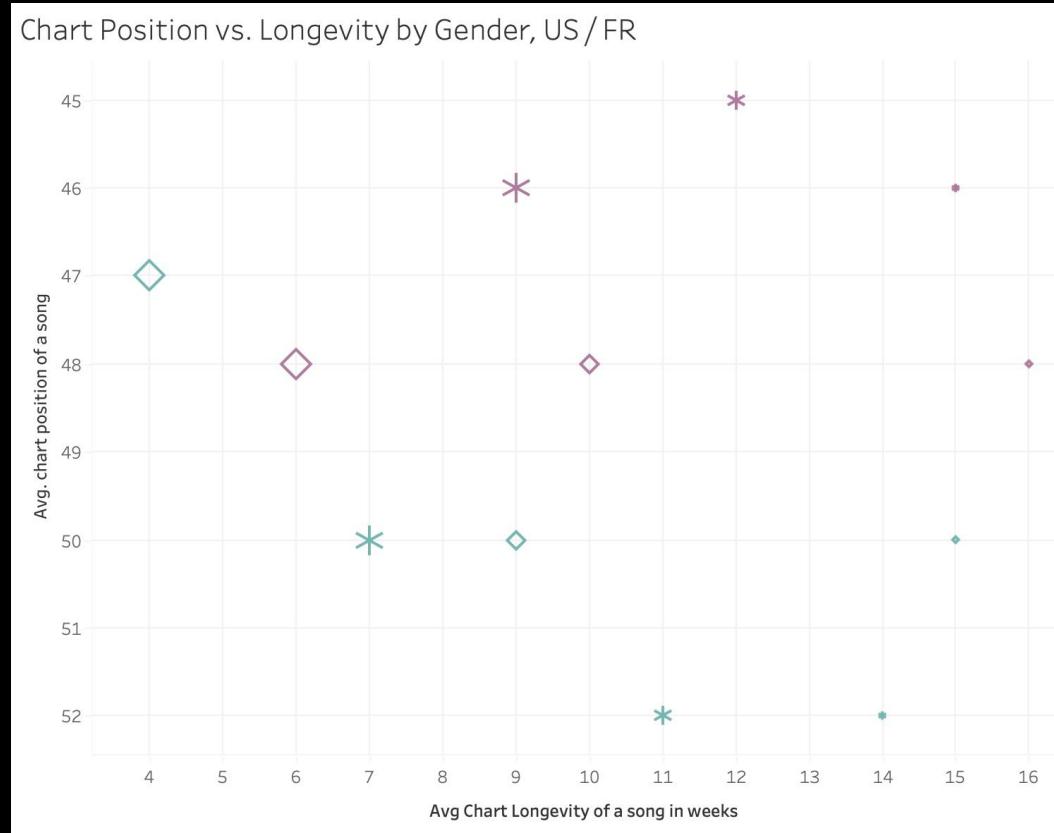


THE GAP EVEN WIDENS IN FRANCE!





EVOLUTION SONG PERFORMANCES BY GENDER



- Gender of main performer
- Female
 - Male
- Decade
- 2000
 - 2010
 - 2020
- Country of chart
- * US
 - ◊ FR

RECOMMEND
WOMEN ARTISTS
THROUGH
MACHINE LEARNING



UNSUPERVISED MACHINE LEARNING

- Vectorize Bios (TfidfVectorizer)
- Vectorize Genres & Tags (CountVectorizer)
- Concat & weight genres & tags
- Normalize (Normalizer)
- NearestNeighbors



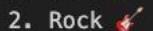
ARTIST RECOMMENDER



Select a genre:



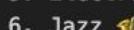
1. Pop 🎵



2. Rock 🎸



3. Hip hop 🎤



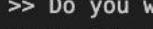
4. R&B 💍



5. Electronic 🎶



6. Jazz 🎵

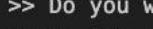


7. Folk 🪕



8. Soul 🎵

9. Alternative Rock 🙌



>> Do you want us to recommend an artist similar to >> Alanis Morissette

Okay, let our recommender find a similar artist to Alanis Morissette

gendered_artists.iloc[90] ⚡	
✓	0.0s
Artist	Alanis Morissette
MusicBrainz ID	4bdcee62-4902-4773-8cd1-e252e3e31225
Type	Person
Gender	Female
Area	Canada
Begin Date	1974-06-01 00:00:00
End Date	NaT
genre_tags	[alternative pop, alternative rock, folk rock, ...]
other_tags	[1990s, adult alternative, english, english, ...]
Bio	Alanis Nadine Morissette (born June 1, 1974) ...
Name:	90, dtype: object



```
#Iterate through the neighbors to return the Female artists
female_neighbors = []
for index in indices[0]:
    if gendered_artists.loc[index, 'Gender'] == 'Female':
        female_neighbors.append(index)
closest_female_neighbors = gendered_artists.loc[female_neighbors]

✓ 0.1s

female_neighbors
✓ 0.1s

[90, 3724, 3818]

gendered_artists.iloc[3724] ⚡
✓ 0.0s

Artist                               Sinéad O'Connor
MusicBrainz ID                      c78a77fa-507c-4c07-947a-035502945bd
Type                                Person
Gender                               Female
Area                                 Ireland
Begin Date                          1966-12-08 00:00:00
End Date                            NaT
genre_tags                           [alternative rock, celtic rock, folk, pop, pop...
other_tags                           [alternative, ireland, irish, irlandais, irlan...
Bio                                  Sinéad Marie Bernadette O'Connor (/ʃe'neid o'...
Name: 3724, dtype: object

gendered_artists.iloc[3818] ⚡
✓ 0.0s

Artist                               Stevie Nicks
MusicBrainz ID                      b7f2cca2-72c6-41fb-ae33-53370fc62fe7
Type                                Person
Gender                               Female
Area                                 United States
Begin Date                          1948-05-26 00:00:00
End Date                            NaT
genre_tags                           [country rock, pop, pop rock, rock, singer-song...
other_tags                           [country, rock, pop, rock, singer-songwriter, ...
Bio                                  Stephanie Lynn Nicks (born May 26, 1948)[1] is ...
Name: 3818, dtype: object
```

▶ DEMO

- 1) Let's get a random artist first from the US or French charts (U / F)
- 2) Select a musical genre
- 3) Confirm if you like the artist and would like us to recommend a similar one (Y) or if you would prefer another artist to start with.





CHALLENGES

The screenshot shows a SourceTree interface with a commit dialog open. The commit message is:

```
Committing
> Afficher la sortie complète
```

Annuler

Matthieu Baillard <matthieu.baillard@gmail.com>

concat vectors
update variable names for NB consistency
split NB in 2 parts with pickle exports
add NB project-part2 for ML
update UI & add Spotify player for recommender

✓ Pousser immédiatement les commits vers origin/main

gender-diversity-in-music-charts (Git)

.gitignore

Fragment 1: Lignes 1-12

```
1 + # CSV files too heavy to load to Github
2 + artist_bios.csv
3 + artist_bios.json
4 + scrobbles-goodmornincaptn_230528.csv
5 + Hot100_HipsterVizNinja.csv
6 +
7 + # TESTS
8 + projet_draft.ipynb
9 +
10 # Byte-compiled / optimized / DLL files
11 __Pycache__/
12 *.py[cod]
```

Pousser Récupérer Branche Fusionner Stash

View Remote Afficher dans le Finder Terminal Préférence Rechercher Indexer le fragment Rejeter le bloc

A large black box contains the text:

Nous n'avons pas pu exécuter des cellules, car le kernel 'base (Python 3.9.13)' est mort.

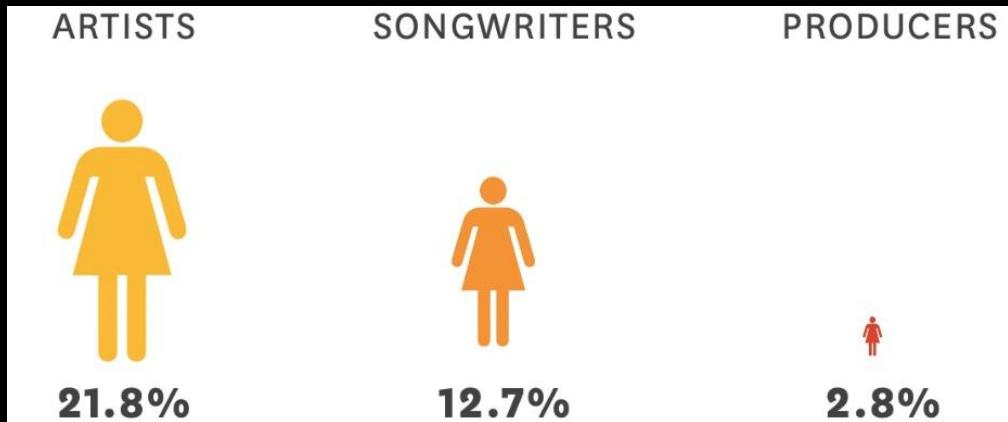
Annuler Archiver





LIMITATIONS / NEXT STEPS

- Improve cleaning functions, check in detail to plan exceptions
- Considering only performing artists doesn't show the whole issue in the industry





LIMITATIONS / NEXT STEPS

- Include bands
- More varied sources and genres for the recommender
(though my old laptop may not survive through
vectorizing long texts at larger scale)
- Improve recommender's UI





Ella Fitzgerald - Ella Fitzgerald Sings the Cole Porter Song Book
Etta James - Lady In Satin
Billie Holiday - Lady Day At Emerson's Bar & Grill
The Kinks - Lola vs. Powerman and the Moneygoround, Part One
Aretha Franklin - Aretha Now
Robbie Robertson - Robbie Robertson
Veronica Hart - Veronica Hart Sings
Nina Simone - Nina Simone
Astrud Gilberto - Astrud Gilberto Album
Aretha Franklin - Aretha Now
Robbie Robertson - Robbie Robertson
Nico - Chelsea Girl
Brigitte Fontaine - Brigitte Fontaine Est...
Buffy Sainte-Marie - Buffy Sainte-Marie - Illuminations

Dusty Springfield - Dusty In Memphis (Deluxe Edition)
Alice Coltrane - Journey in Satchidananda
Joni Mitchell - Blue
Lionel Richie - Hello (Legacy Edition)
Patti Smith - Horses
Jean Michel Jarre - Oxygene
Katy Perry - California Dreams
Marianne Faithfull - Broken English
Madonna - Material Girl
Donna Summer - Bad Girls
Dolly Parton - 9 To 5 And Odd Jobs
Diana Ross - Diana

Jean Jett & The Blackhearts - I Love Rock 'N' Roll (Expanded Edition)
Grace Jones - Nightclubbing
Go-Go's - Beauty and the Beat
Madonna - Like a Prayer
Stevie Nicks - Bella Donna (2016 Remastered)
Dusty Springfield - Dusty Springfield
Elton John - Don't Go Breaking My Heart
Hootie & the Blowfish - Maniac
Hurythmics - Touch
Cyndi Lauper - She's So Unusual
Madonna - Material Girl
Tina Turner - Private Dancer (30th Anniversary Issue)
Miharu Kohki - バラリズム

Whitney Houston - Whitney
Kate Bush - Hounds of Love (2018 Remaster)
Janet Jackson - Control
U2 - Rattle and Hum
Diana Ross - The Divine Punishment & Saint of the Fit
Suzanne Vega - Solitude Standing
Whitney Houston - All I Wanna Do
Tracy Chapman - Tracy Chapman ...
Tori Amos - Little Earthquakes
Madonna - Material Girl
Bikini Kill - the first two records
Yoko Ono - Yes, I'm A Witch

Liz Phair - Exile in Guyville
Robbie Williams - This
Portishead - Dummy
Adele -莫里索特- Jagged Little Pill
Paris Hilton - Why You Love My Love
Spice Girls - Spice
Missy Elliott - Supa Dupa Fly
Rita Ora - I Will Never Let You Go
Ms. Lauryn Hill - The Miseducation of Lauryn Hill
Fergie - Apologize - Felt Mountain...
Goldfrapp - Black Cherry
Kylie Minogue - Fever

Aaliyah - Aaliyah
Britney Spears - Britney (Digital Deluxe Version)
Norah Jones - Come Away With Me
Erykah Badu - Worldwide Underground
MC Hammer - U Can't Touch This
Joanna Newsom - Yes
Amy Winehouse - Back to Black
Cassadee Pope - Singing in the Way of Control
M.I.A. - Kala
Yeon Sun Nah - Nanyang Lane
David Bowie - Station To Station
Lady Gaga - The Fame Monster (Deluxe Edition)

Robyn - Body Talk
Jade Bird - The ArchAndroid
Adele - 21
Taylor Swift - Red
Lana Del Rey - Paradise
Lykke Li - I Never Learn
Peaches - R.U.B.
Sia - Chandelier
Fever Ray - Plunge
Charlotte Gainsbourg - Rest
Anne von der Lippe - Magic
Sophie - Oil of Every Pearl's Un-Insides

Lizzo - Cuz I Love You
LILONG - I GONE
Rihanna - HOMECOMING: THE LIVE ALBUM
Dua Lipa - Future Nostalgia
Kanye West - Ye
Drake - Scorpion
Wu-Tang Clan - Wu-Tang Clan Is Back!
H.E.R. - park bae jin
Little Mix - Sometimes I Might Be Introvert
Dua Lipa - New Rules
Charli XCX - CRASH
Rosalia - MOTOMAMI
Shyguy - Nymp
Hildur Guðnadóttir - Tár (Music from and Inspired by the Motion Picture)

REFERENCES

[**Inclusion in the Recording Studio? Gender and Race/Ethnicity of Artists, Songwriters & Producers across 1,000 Popular Songs from 2012-2021**](#), Karla Hernandez, Dr. Stacy L. Smith & Dr. Katherine Pieper, *USC Annenberg Inclusion Initiative*, 2022

[**État des lieux de la présence des femmes dans la filière musicale**](#), Centre national de la musique 2023

[**Gender In The Music Industry - Rock, Discourse and Girl Power**](#), Marion Leonard 2007

[**Gender, Branding, and the Modern Music Industry: The Social Construction of Female Popular Music Stars**](#), Kristin J. Lieb 2018

[**Facing the music: Stereotyping of and by women in US music journalism**](#), Kelsey Whipple and Renita Coleman, *Journalism*, 2021

[**Towards Gender Equality in the Music Industry: Education, Practice and Strategies for Change**](#), Catherine Strong · Sarah Raine 2019

[**A Feminist Analysis of Popular Music - Power Over, Objectification of, and Violence Against Women**](#), Brook Bretthauer MS,Toni Schindler Zimmerman & James H. Banning, *Journal of Feminist Family Therapy*, 2007

[**The influence of misogynous rap music on sexual aggression against women**](#). Barongan, C. and Hall, G. C., *Psychology of Women Quarterly*, 1995

[**Impact of music lyrics and rock music videos on children and youth**](#). Committee on Communications, American Academy of Pediatrics. 1996.

[**Constructions of Femininity in 1990s Girl-Group Music**](#), Nicola Dibben, *Feminism & Psychology*, 2002