



Spotify Analysis

Presented By: Group 5

Lecturer : Mr. Chan Sophal
Subject : Power BI



Our Team's



Eang Kimchhiv



Neang Rothmny



Gnep Vandara



Thear Sithysak

Table Of Content



Introductions

01

Objectives

02

Implementation

03

Data Cleaning

04

Data Modeling

05

Graph Briefing

06

Recommendations

07

Conclusion

08

Introduction



SpotGenTrack is a music dataset that was created for music popularity prediction, genre classification, automatic music tagging, music similarity, and other interdisciplinary applications within the music industry. The dataset is multimodal, including audio features, lyrics, and metadata from Spotify and Genius.

Objective

We will dig down to the most popular artists, tracks, genres, and albums..... to satisfy the listener and we do some insights into those data sets to make sure everyone is satisfied.



Implementation

1. Get dataset from open sources: kaggle.
2. Access to Wamp Server and open port for MySQL.
3. Import Dataset to MySQL server.
4. Connect Power BI to MySQL and import data.
5. Total Album, Artist, and Track
6. Follower of Artist
7. Popularity of Artist
8. Detect N-Word
9. Total time(ms) in all songs
10. Available Market



kaggle



Power BI



Data Cleaning



- Using MICE method to clean by Impute the wrong values of data by average value.
- But for column “sentence_similarity” we change from -1 to 0 because similarity is a percentage so it must between 0 and 1.

Data Cleaning

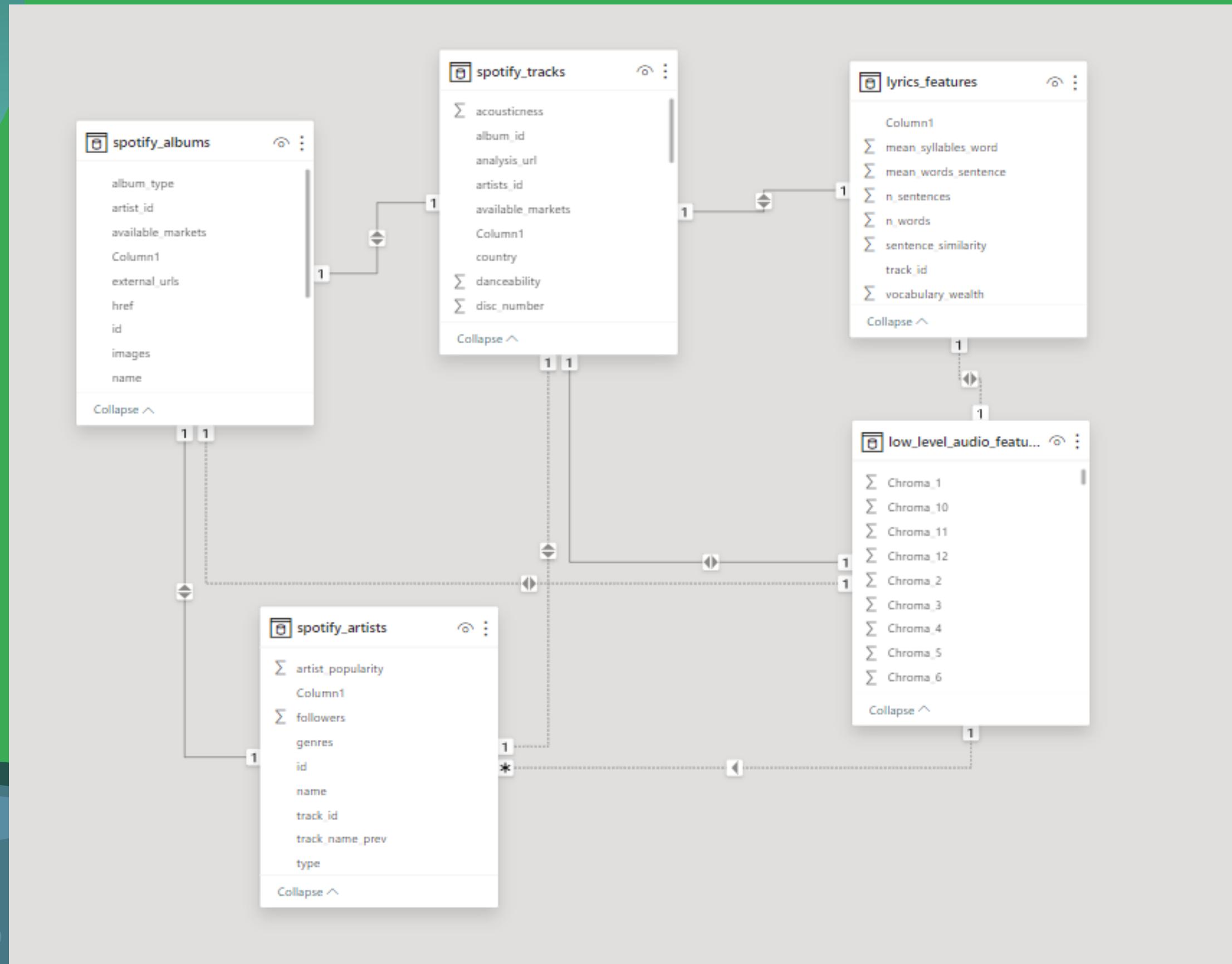


Column1	mean_syllables_word	mean_words_sentence	n_sentences	n_words	sentence_similarity
0	-1	-1	-1	-1	-1
36	-1	-1	-1	-1	-1
38	-1	-1	-1	-1	-1
41	-1	-1	-1	-1	-1
43	-1	-1	-1	-1	-1
45	-1	-1	-1	-1	-1
46	-1	-1	-1	-1	-1
48	-1	-1	-1	-1	-1
50	-1	-1	-1	-1	-1

To

Column1	mean_syllables_word	mean_words_sentence	n_sentences	n_words	sentence_similarity
0	1.05	3.6	42.3	317	0
36	1.05	3.6	42.3	317	0
38	1.05	3.6	42.3	317	0
41	1.05	3.6	42.3	317	0
43	1.05	3.6	42.3	317	0
45	1.05	3.6	42.3	317	0
46	1.05	3.6	42.3	317	0
48	1.05	3.6	42.3	317	0
50	1.05	3.6	42.3	317	0

DATA MODELING



Graph Briefing

75.51K

Total Album

75.51K

Total Artist

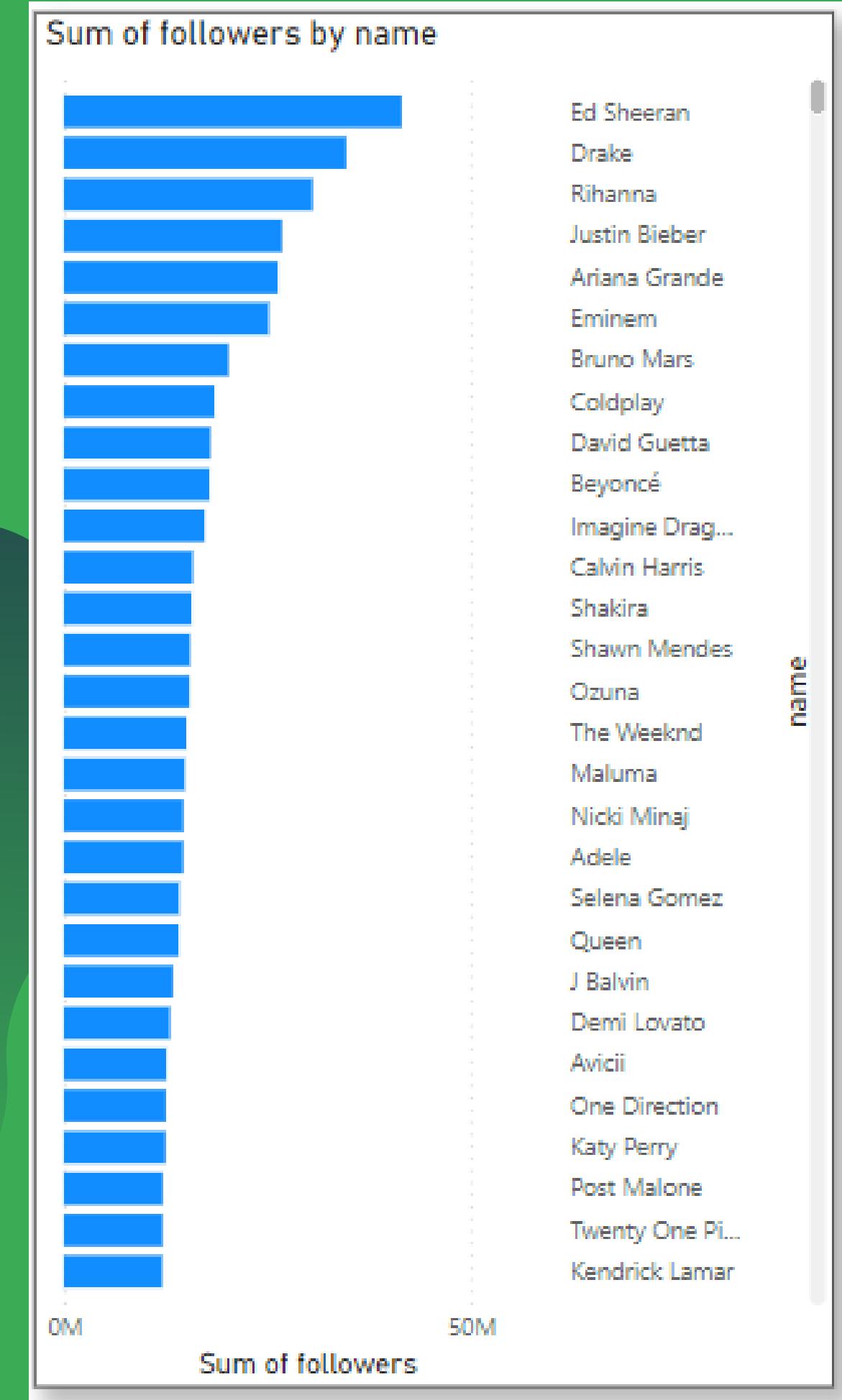
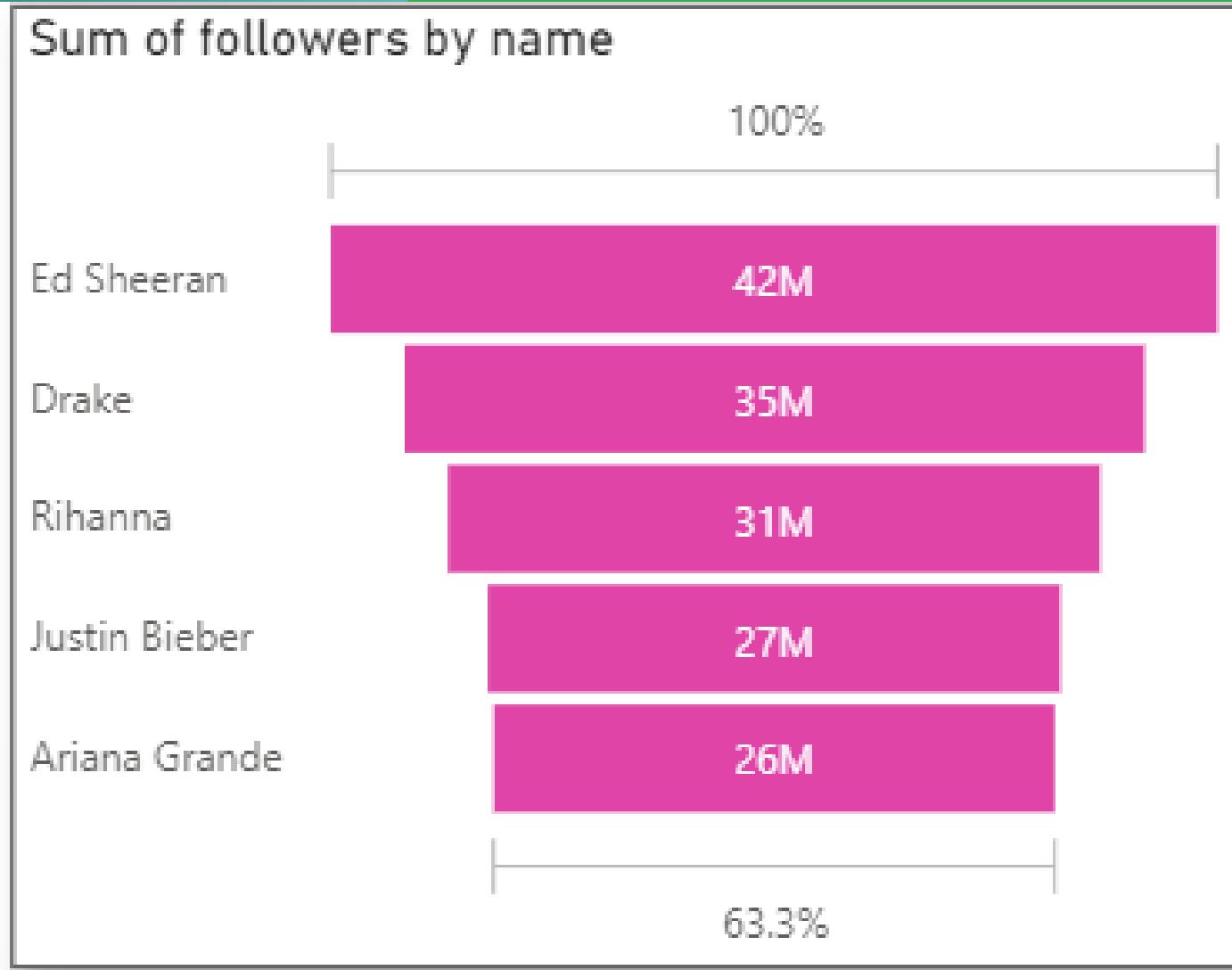
622K

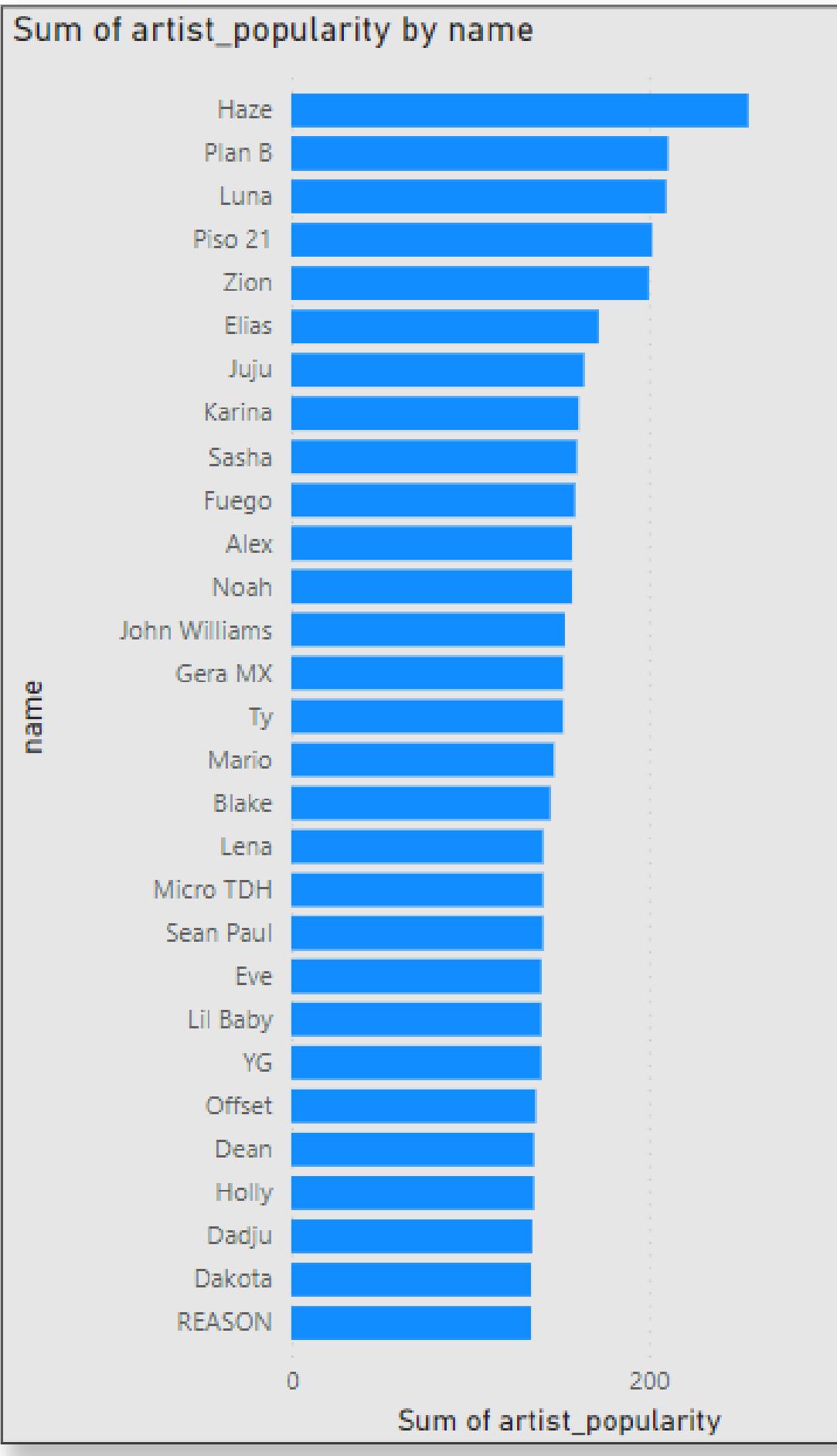
Total Tracks



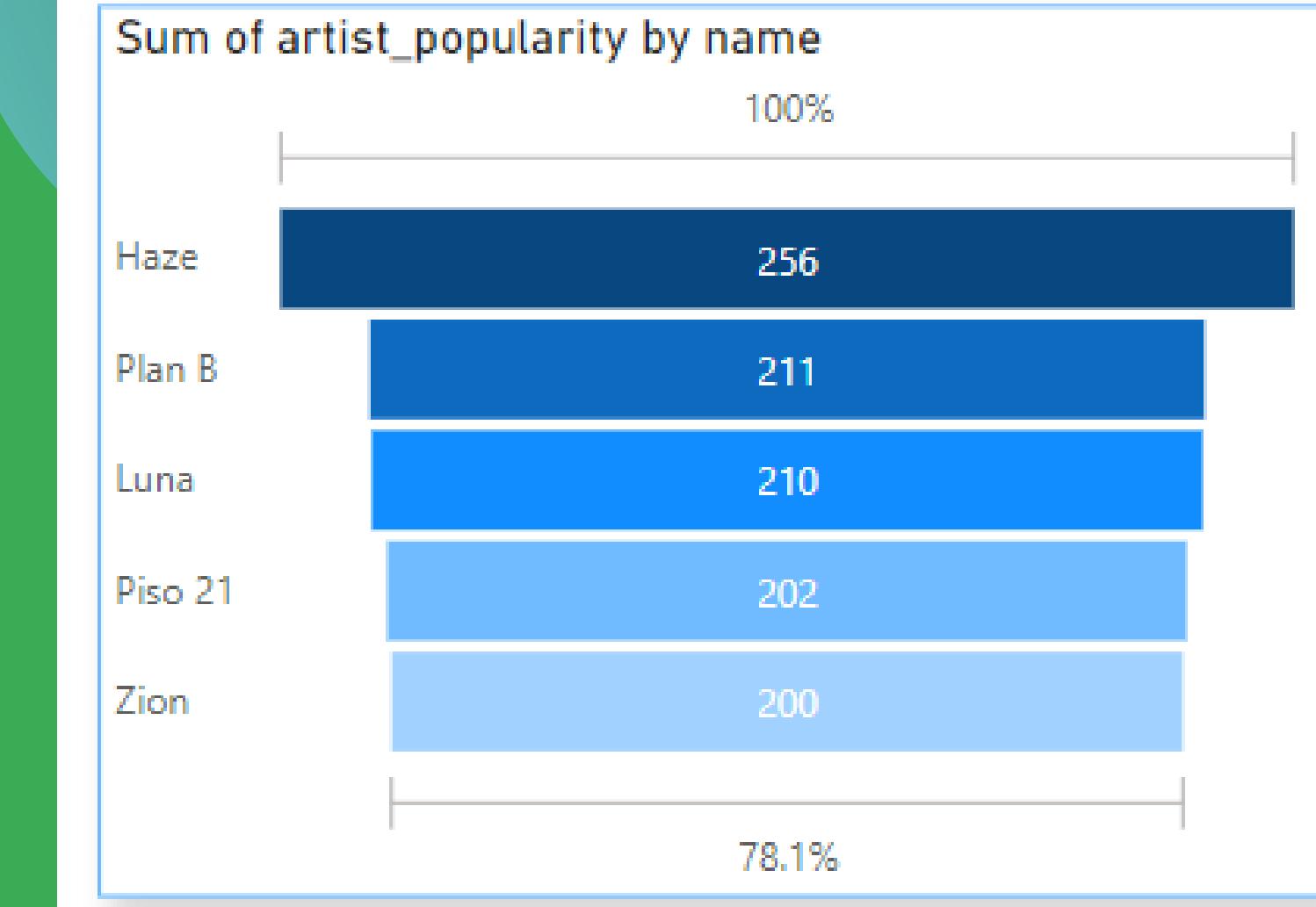
Describe about the **followers** of artist.

4bn

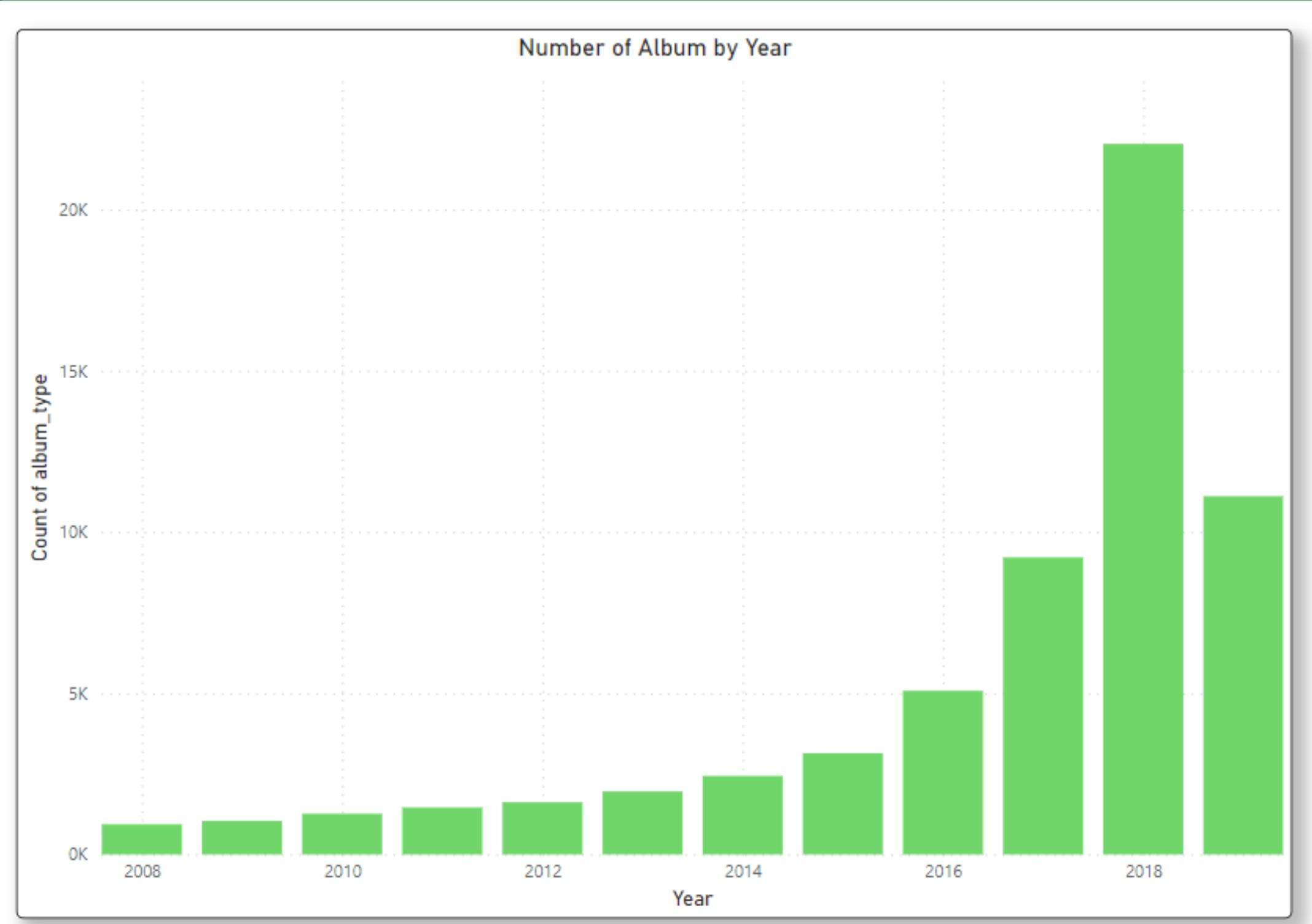




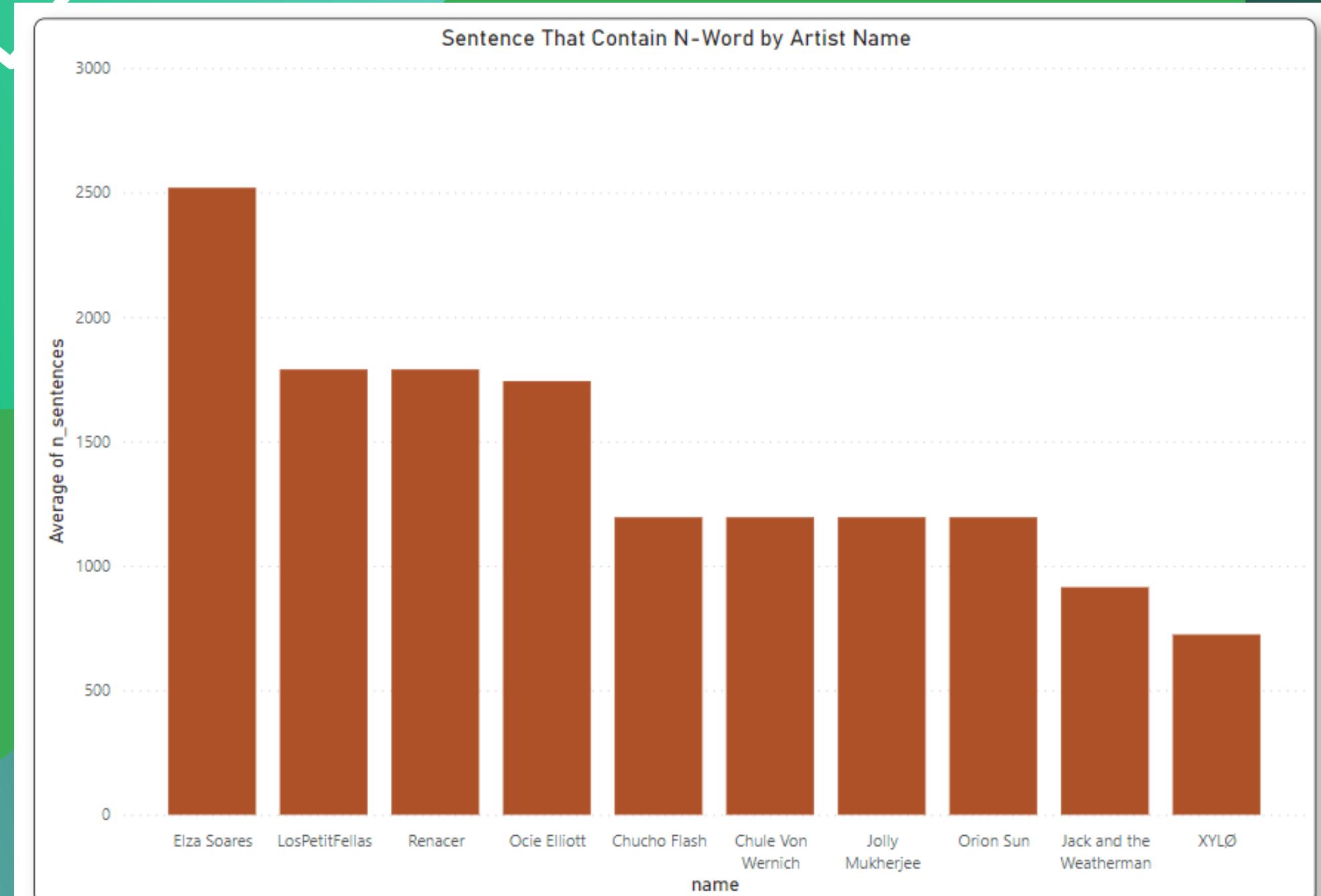
Describe about **popularity** of artist by name.



Describe about **number of albums** by year.



Describe about sentence that contain **N-word**
by artists.



Total time(ms) in all songs by Top 10 artists

Artist Name	Duration (ms)
Danny Avila	5505831
Ernesto Amstein	4694040
Cor Veleno	4667397
Santhanam	4508318
Yall	4500037
Familia Torres	4462995
Deeb	4437892
Malk De Kojn	4423000
Zuma Dionys	4373004
	216000
Total	216000



Regions that spotify spread



Recommendations

01

Expand the popularity of the artist based on their genre song.

02

Implementation ML to do the recommendation based on their mutual listening.

03

Have the mode for users to use Explicit free content so they don't get offended.

04

Allowing users to specify their preferences, such as their favorite genres and artists, and allowing them to filter out songs that they do not want to hear.



CONCLUSION



As technology evolves and user preferences shift, ongoing improvements in areas such as user experience, music discovery, and content management will likely contribute to an even more personalized and enjoyable music streaming experience.



