# MeeT the TEAM

Nseke Ngilbus, SFDA 2023

Safiya Mirza, SFDA 2023

Michaela Bodie, SFDA 2023

Junwei Li, SFDA 2023


10
October | 2023
Spotify '23
Data
Visualization
Project

# AGENDA

## Business Challenge

What is the problem we are tackling?

## Analysis/Visualization

What did we find?

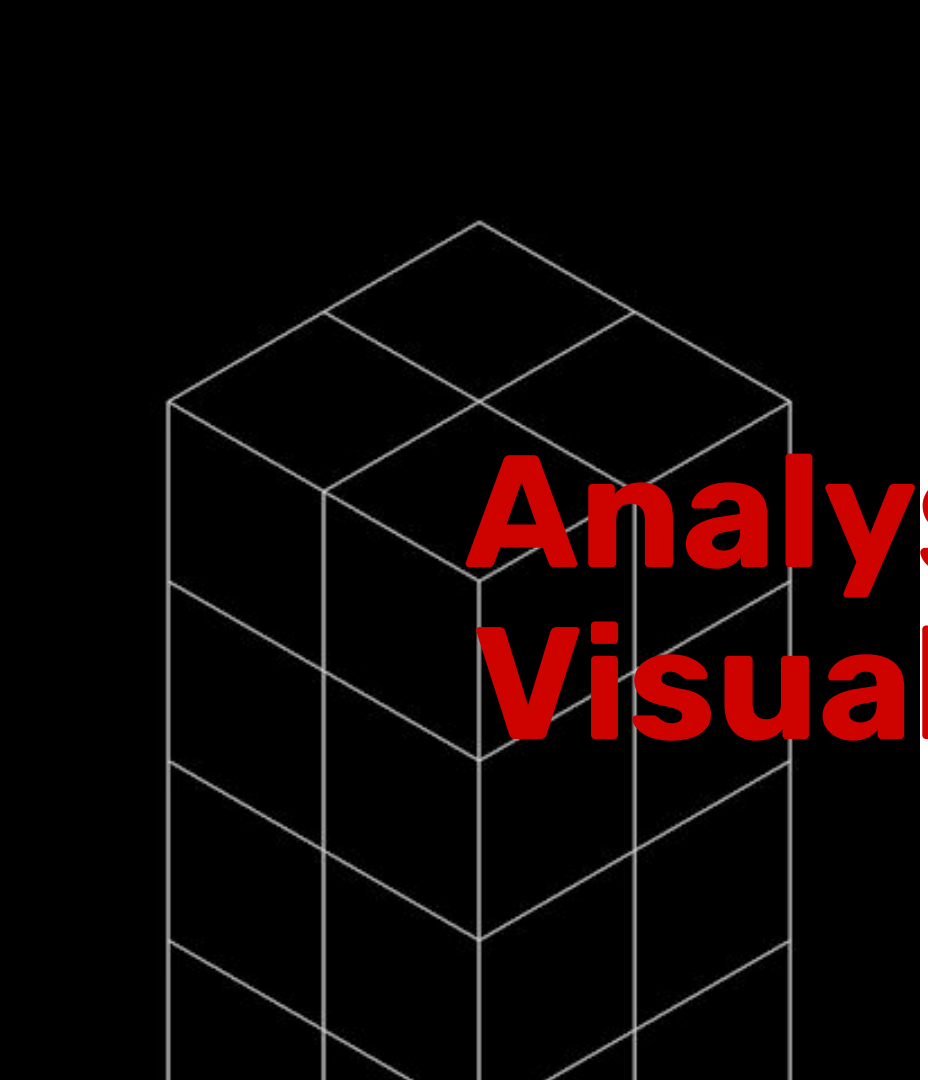
## Results

Summarizing our key findings

# The **Business Challenge**

# The Purpose

We want to perform a musical analysis on the top 50 most streamed songs of 2023 as listed on Spotify. We will examine certain attributes in audio features and analyze whether or not they account for trends in song popularity, as well as what was the most popular level of these attributes.



# Analysis & Visualization

# The Data

**Kaggle maintains records of the most streamed Spotify songs in 2023. The dataset offers 943 records of this years' most streamed songs as well as insights into their attributes, however for our analysis we focused on the top 50 most streamed songs.**

# What even are the Attributes?

## List of Attributes

**BPM:** Beats per minute, a measure of song tempo

**Valence %:** Positivity of the song's musical content

**Danceability %:** How suitable the song is for dancing

**Acousticness %:** Amount of acoustic sound in the song

**Instrumentalness %:** Amount of instrumental content in the song

**Liveness %:** Presence of live performance elements

**Speechiness %:** Amount of spoken words in the song

**Energy %:** Perceived energy level of the song

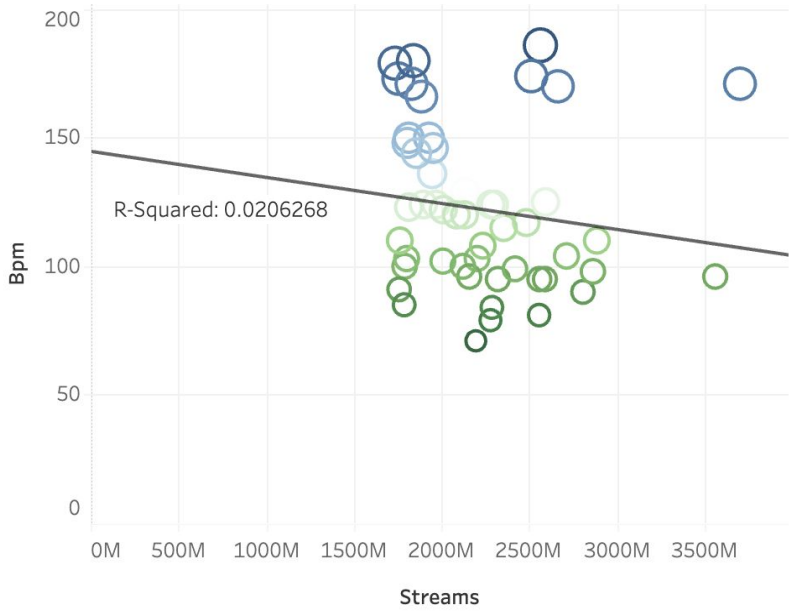
# What is a R-Squared value?

Definition

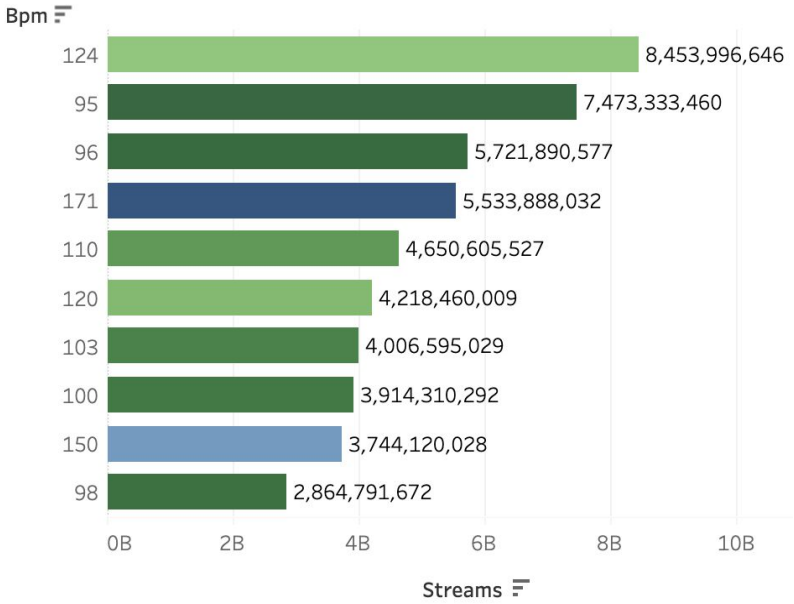
- ★ R-squared is the proportion of the variation in the dependent variable that is predictable from the independent variable
- ★ In terms of our analysis, R-squared tells us how much the musical attribute accounts for determining number of streams
- ★ The closer the R-squared value is to 1, the stronger the correlation is between the two variables

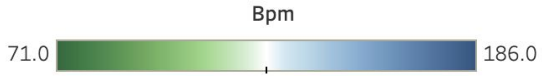
# BPM Statistics

## # of Streams vs. BPM 2023



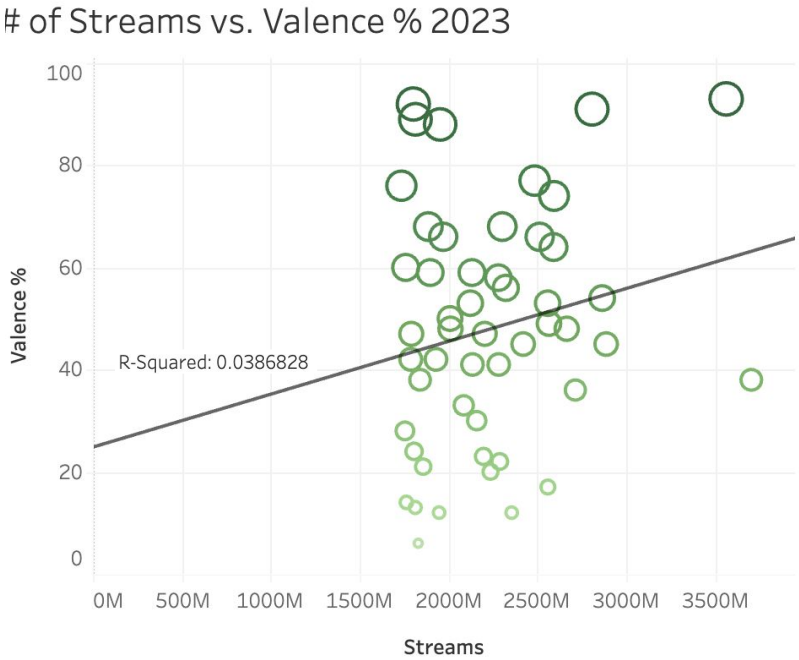

## Most Popular BPM 2023

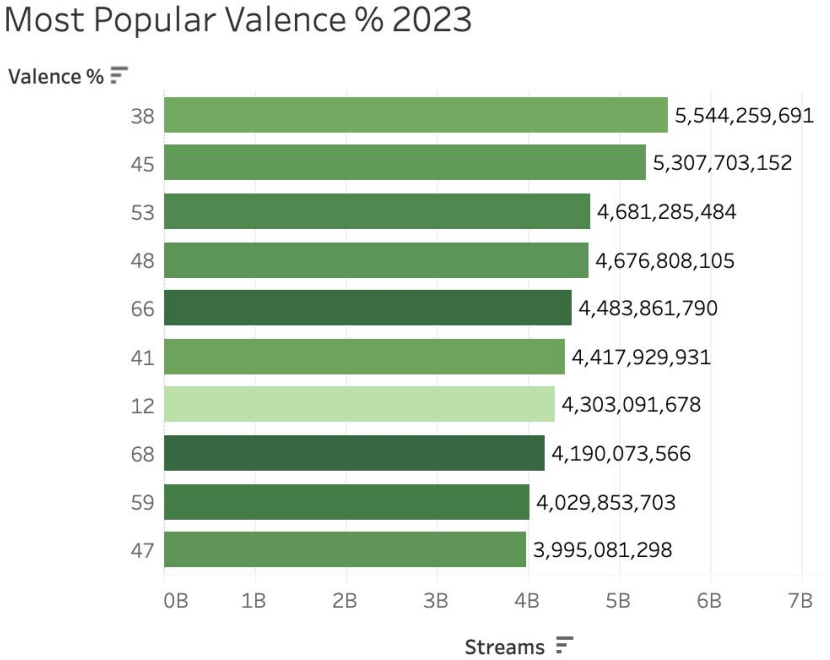



- **BPM and # of Streams has a very weak negative linear relationship**
- **The BPM level with the most amount of streams is a BPM of 124**

# Valence % Statistics



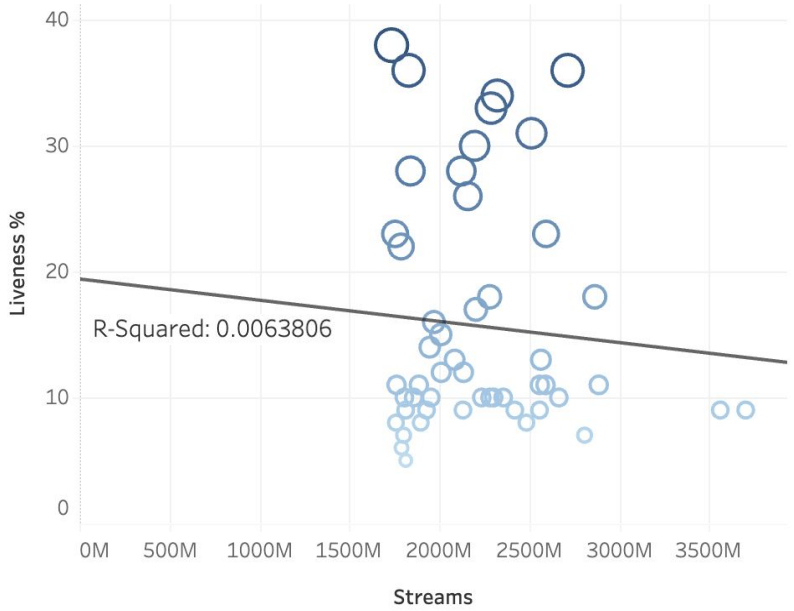





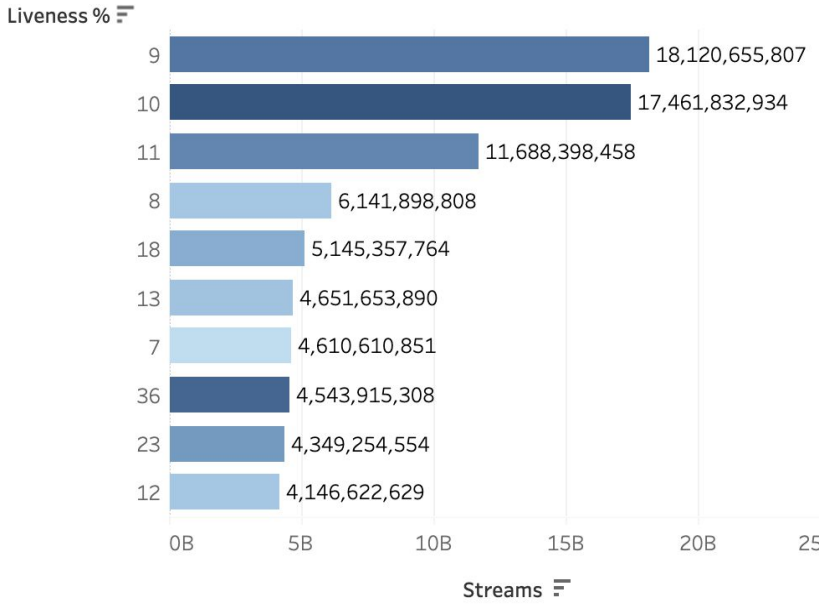
- **Valence % and # of Streams has a very weak positive linear relationship**
- **The Valence percentage with the most amount of streams is 38%**

# Liveness % Statistics

## # of Streams vs. Liveness % 2023




## Most Popular Liveness % 2023

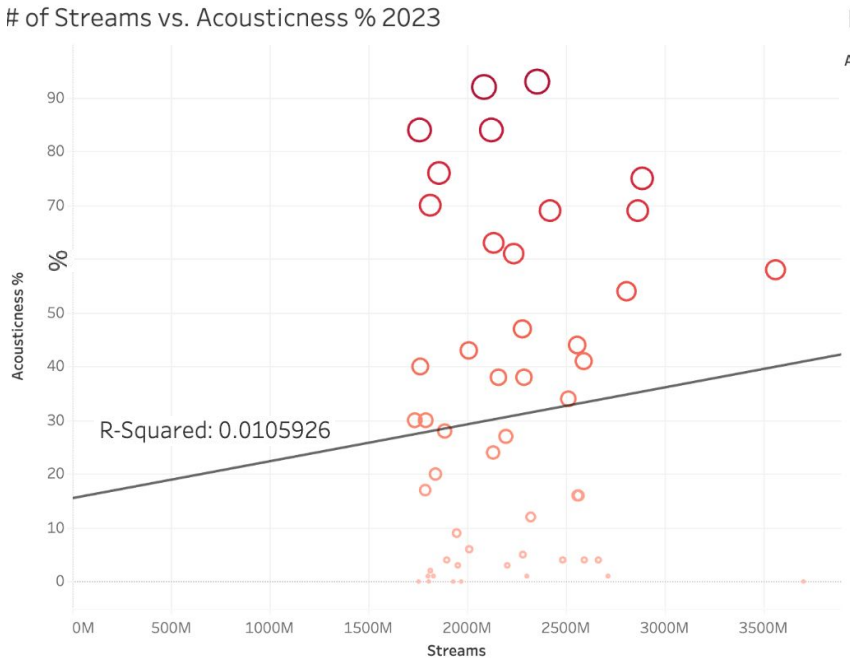


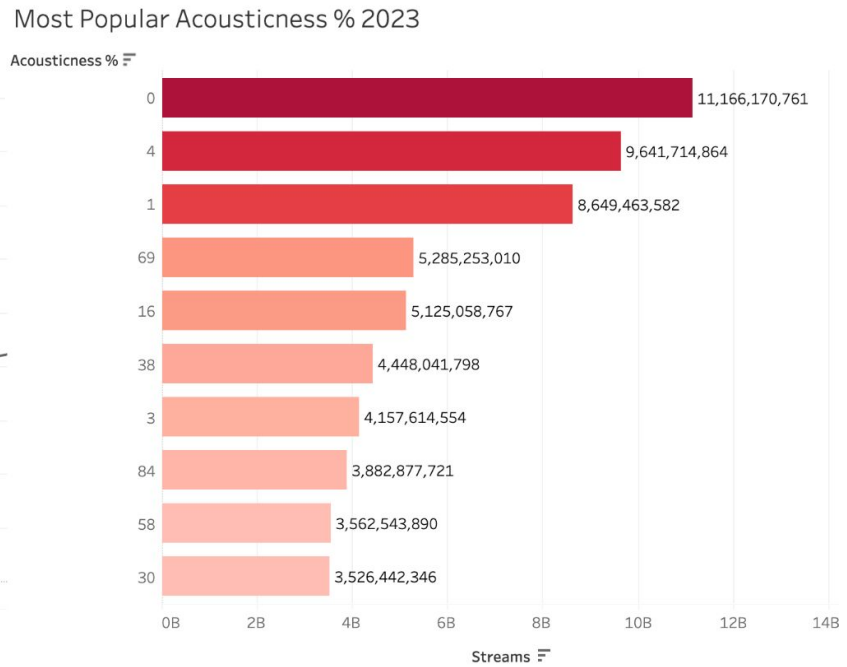

Liveness %

5 38

- **Liveness % and # of Streams has a very weak negative linear relationship**
- **The Liveness percentage with the most amount of streams is 9%**

# Acousticness % Statistics

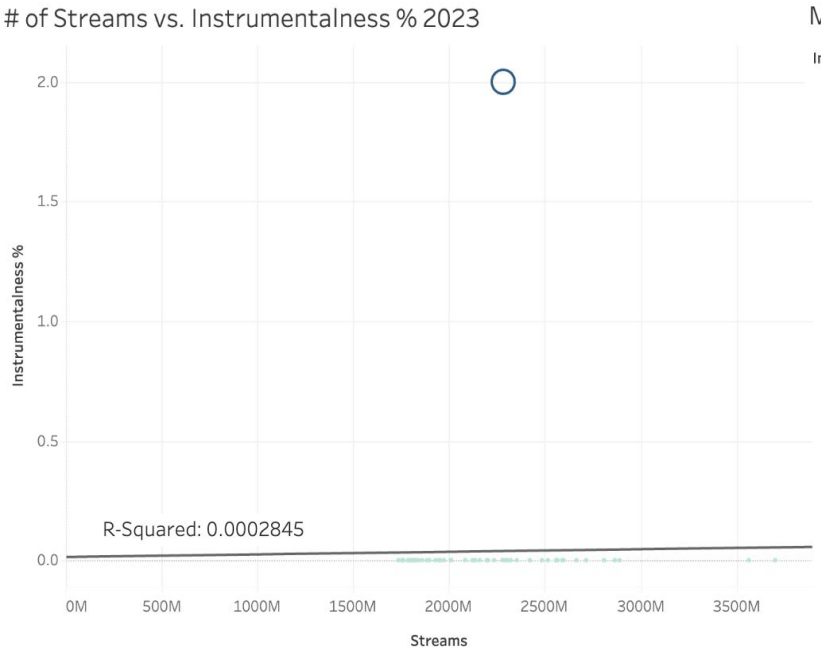


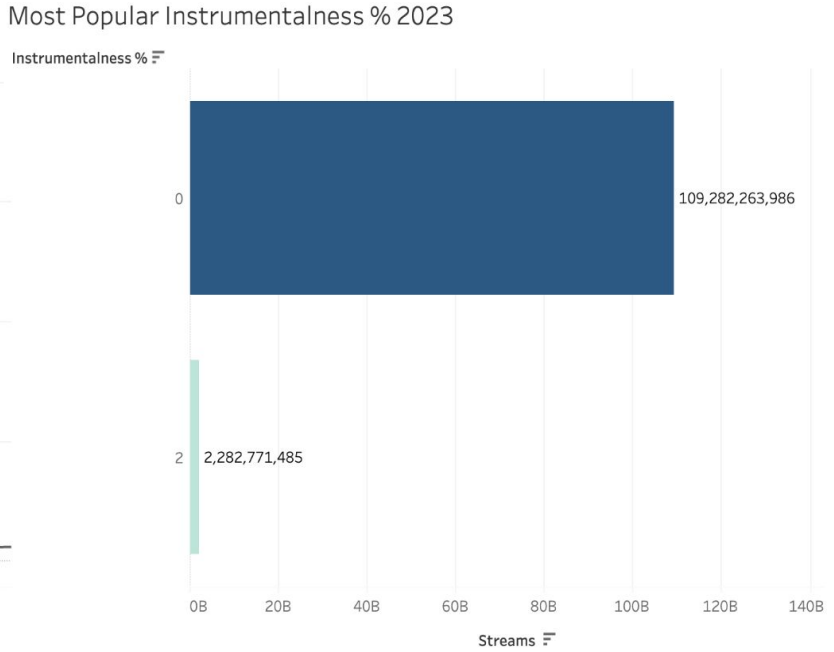





- **Acousticness % and # of Streams has a very weak positive linear relationship**
- **The Acousticness percentage with the most amount of streams is 0%**

# Instrumentalness % Statistics

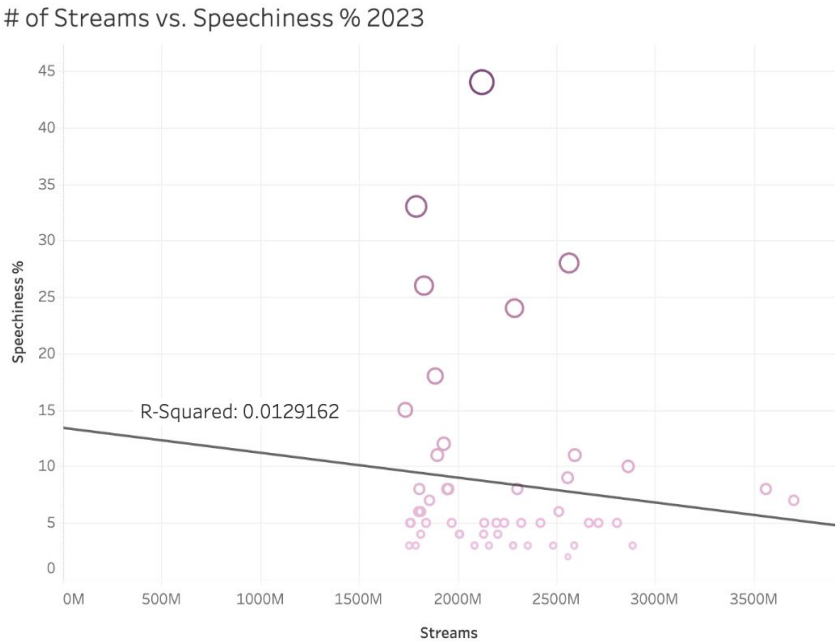


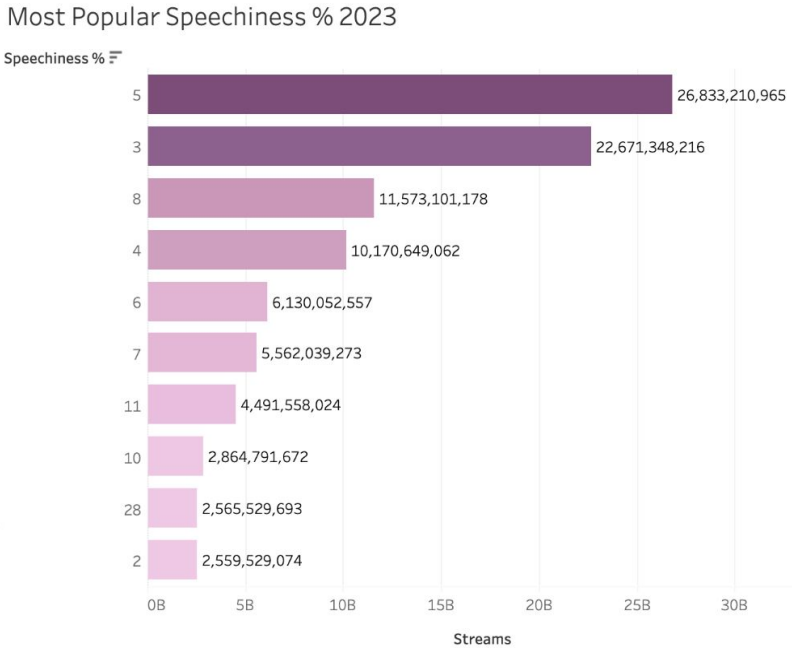




Instrumentalness %
0.000 2.000

- **Instrumentalness % and # of Streams has nearly no correlation at all**
- **The Instrumentalness percentage with the most amount of streams 0%**

# Speechiness % Statistics

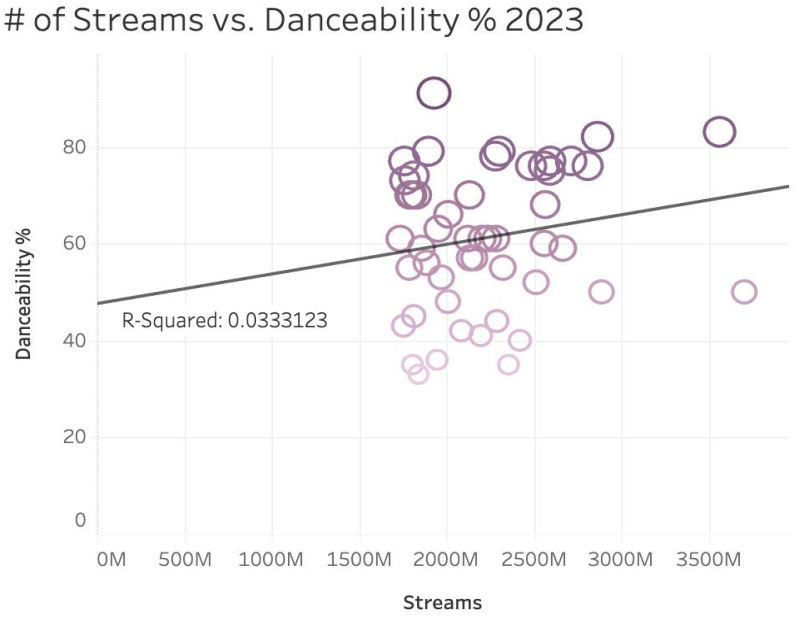


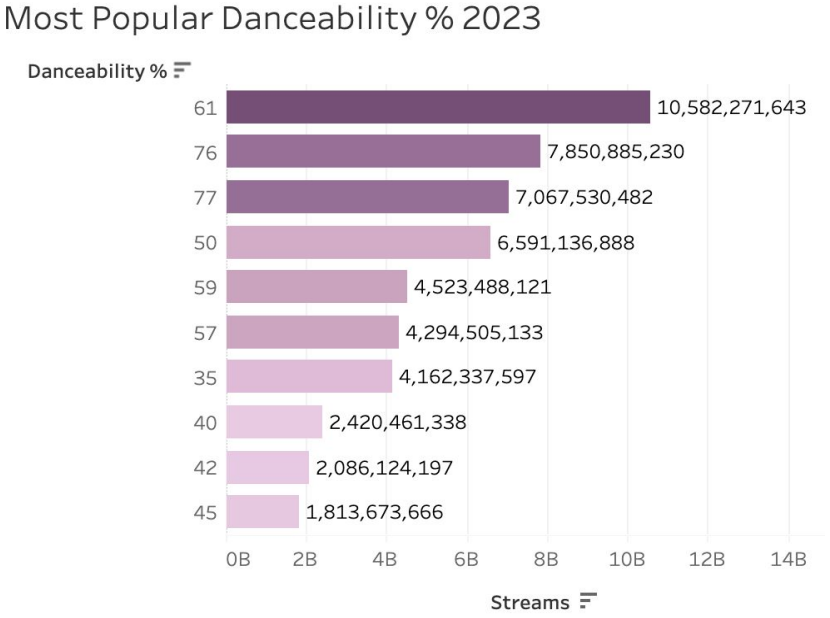





- **Speechiness % and # of Streams has a very weak negative linear relationship**
- **The Speechiness percentage with the most amount of streams is 5%**

# Danceability % Statistics



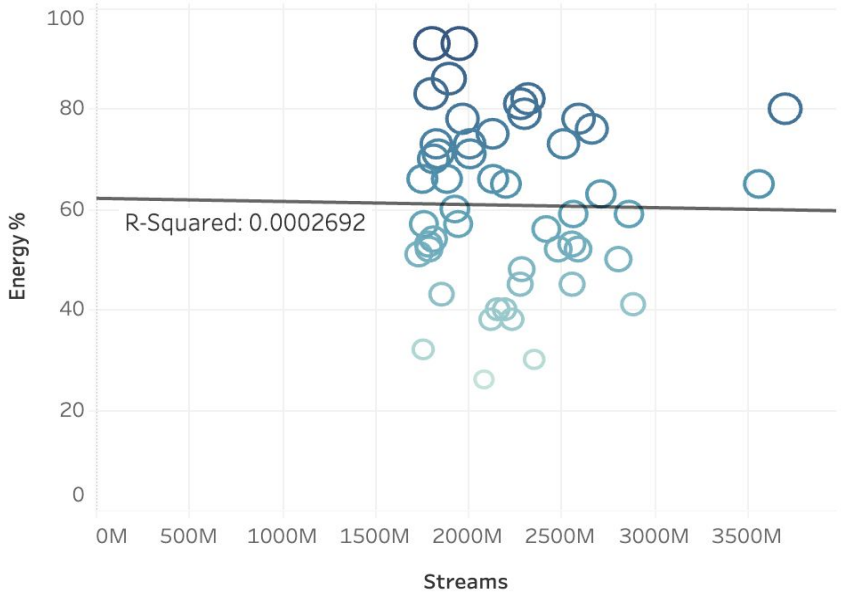





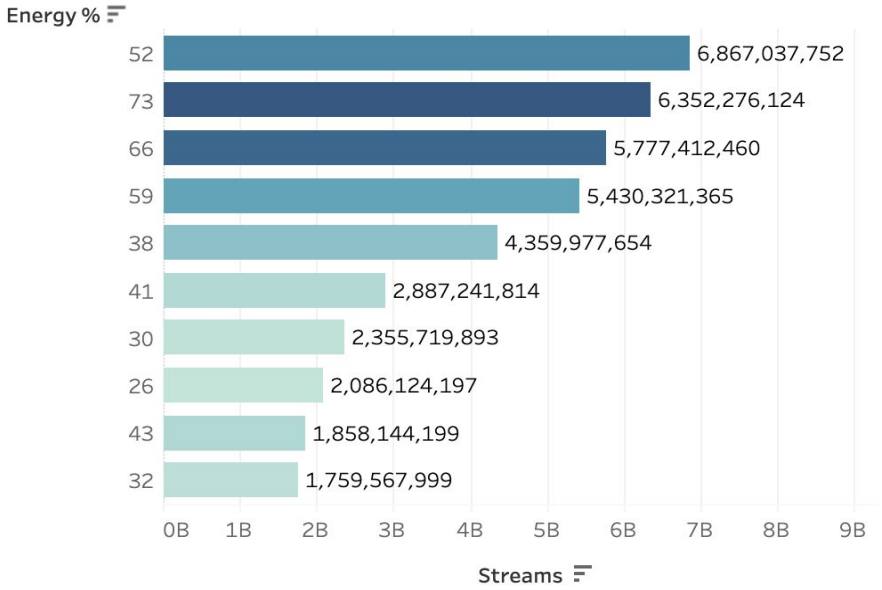
- **Danceability % and # of Streams has very weak positive linear relationship**
- **The Danceability percentage with the most amount of streams is 61%**

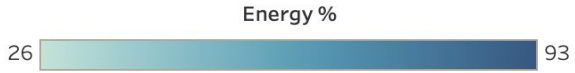
# Energy % Statistics

## # of Streams vs. Energy % 2023




## Most Popular Energy % 2023




- **Energy % and # of Streams has nearly no correlation at all**
- **The Energy percentage with the most amount of streams is 52%**

# RESULTS

# Our Findings

*The Data Story*

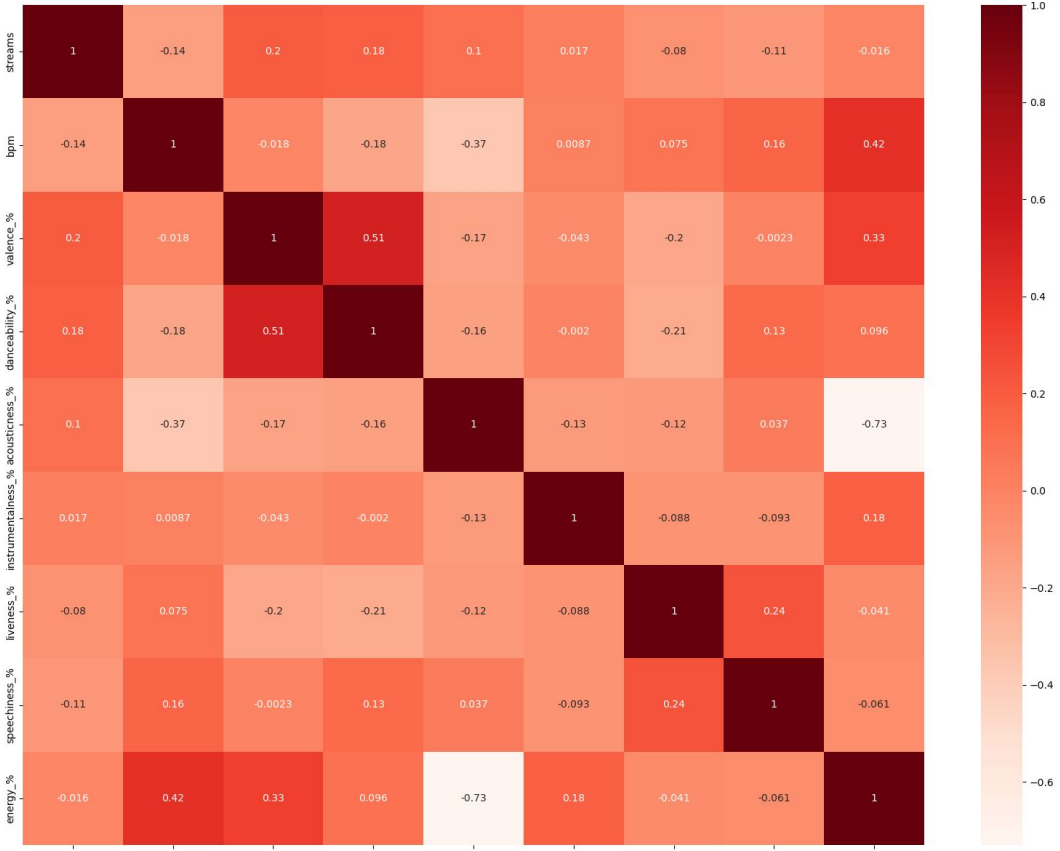
1. **All of the attributes selected have a very minimal correlation with number of streams within the Top 50 Tracks of 2023 (look to R-squared values)**

2. **Out of all the attributes analyzed, Valence Level and Danceability Level has the highest positive correlation with number of streams (look to heatmap, next page)**

3. **The Top 50 Tracks are characterized by quite high values of BPM, Valence, Energy, and Danceability**

4. **The Top 50 Tracks are characterized by quite low values of Acousticness, Speechiness, Instrumentalness, and Liveness**

# IN SUMMARY




- ★ **BPM, Danceability %, Valence %, Acousticness %, Instrumentalness %, Liveness %, Speechiness %, Energy % have little to no correlation with the number of streams a song receives**
- ★ **It can be hypothesized that top songs have higher BPM, Valence, Energy, and Danceability levels, and lower Acousticness, Speechiness, Instrumentalness, and Liveness levels**
- ★ **For further studies, a multi-stage regression model can be ran to analyze how the attributes work together to determine number of streams, as well as a logistic regression model to account for potential nonlinearity while determining which predictors are significant in influencing number of streams**

*Correlation Heatmap*

# THE END