

# My title\*

My subtitle if needed

Cristina Su Lam

November 26, 2024

This analysis explores the relationship between various musical and contextual features with the emotional tone (Valence) of songs on Spotify. Using a dataset of songs with features such as tempo, danceability, and acousticness, alongside contextual data like mean temperature, we aim to predict Valence—a measure of positivity in the music. The study applies multiple linear regression models, evaluates performance metrics, and identifies influential predictors. Results show that predictors like danceability and energy strongly relate to Valence, while contextual variables like temperature also provide moderate explanatory power. These findings highlight the potential for combining musical and environmental data to analyze listener sentiment and mood-related trends in music[need to be more detail about findings].

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data</b>	<b>2</b>
2.1	Overview . . . . .	2
2.2	Measurement . . . . .	3
2.3	Outcome variable . . . . .	4
2.4	Predictor variables . . . . .	4
2.4.1	Mean Temperature . . . . .	4
2.4.2	Tempo . . . . .	4
2.4.3	Danceability, Energy, and Acousticness . . . . .	5

---

\*Code and data are available at: [https://github.com/cristinaasu/Spotify\\_Analysis](https://github.com/cristinaasu/Spotify_Analysis)

<b>3</b>	<b>Model</b>	<b>7</b>
3.1	Model Structure . . . . .	7
3.1.1	Model justification . . . . .	7
<b>4</b>	<b>Results</b>	<b>7</b>
<b>5</b>	<b>Discussion</b>	<b>7</b>
5.1	First discussion point . . . . .	7
5.2	Second discussion point . . . . .	8
5.3	Third discussion point . . . . .	8
5.4	Weaknesses and next steps . . . . .	8
	<b>Appendix</b>	<b>9</b>
<b>A</b>	<b>Additional data details</b>	<b>9</b>
<b>B</b>	<b>Model details</b>	<b>9</b>
B.1	Posterior predictive check . . . . .	9
B.2	Diagnostics . . . . .	9
	<b>References</b>	<b>10</b>

```
# Load dataset
analysis_data <- read_parquet("../data/02-analysis_data/analysis_data.parquet")
mlm_model <- readRDS("../models/mlm_model.rds")
```

# 1 Introduction

Valence, a measure of emotional tone in music, is pivotal for understanding how listeners perceive and engage with songs. Ranging from 0 (negative or sad) to 1 (positive or happy), Valence provides insights into the emotive qualities of a track. This study aims to model and predict Valence using a blend of musical features such as tempo, danceability, and acousticness, alongside contextual variables like mean temperature.

Music serves as a universal language, evoking emotions and creating connections. By analyzing how intrinsic musical properties interact with external factors, such as environmental conditions, we can uncover patterns that shape listener sentiment. This approach not only deepens our understanding of the factors influencing emotional tone but also offers practical applications for playlist curation and music recommendation systems.

The estimand of interest is Valence, as it encapsulates the core emotional response elicited by music. Through multiple linear regression (MLR) models, we identify the most significant

predictors of Valence and assess their explanatory power. This approach highlights the interplay between quantitative attributes of music and listener experiences, aiming to improve the accuracy of mood prediction models.

In this analysis, we seek to bridge the gap between musical analytics and contextual data, contributing to the broader field of musicology and data-driven sentiment analysis. By refining predictive methods, we aim to better understand the emotional landscape of modern music and its relationship with the listener’s environment.

[Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section [2](#)....]

## 2 Data

### 2.1 Overview

We use the statistical programming language R (R Core Team 2023).... Our data (Toronto Shelter & Support Services 2024).... Following Alexander (2023), we consider...

The dataset contains key variables spanning musical features, environmental data, and outcomes. Below is a summary of numerical variables:

Table 1: Summary Statistics

Variable	Mean	SD	Min	Max
Valence	0.532	0.243	0.036	0.981
Mean Temp	16.929	5.298	3.497	24.822
Danceability	0.632	0.134	0.234	0.943
Energy	0.632	0.166	0.192	0.968
Acousticness	0.228	0.224	0.000	0.968
Tempo	126.157	29.921	48.718	203.812

These metrics reveal diverse song features and moderate variability in environmental temperature, offering rich explanatory potential for Valence.

## 2.2 Measurement

The data used in this analysis were collected from Spotify’s public API, which provides detailed metadata for songs. Each track is associated with multiple audio features and contextual information. The data was preprocessed to ensure consistency, with missing or erroneous values removed prior to analysis.

**Valence (Outcome):** Extracted directly from Spotify’s API, Valence is measured on a scale from 0 to 1, indicating the positivity or happiness conveyed by a song. This variable serves as the primary outcome of interest.

**Musical Features:**

**Danceability:** Quantifies how suitable a track is for dancing, based on a combination of tempo, rhythm stability, and beat strength. Values range from 0 to 1.

**Energy:** Measures the intensity and activity level of a track, combining attributes like loudness and dynamic range. Higher values indicate more energetic tracks.

**Acousticness:** Represents the likelihood of a track being acoustic, with values closer to 1 reflecting higher acoustic qualities.

**Tempo:** Captured in beats per minute (BPM), tempo describes the overall speed of a track. This feature was extracted directly from Spotify’s analysis tools.

**Contextual Feature:**

**Mean Temp:** Weather data was integrated using external sources and mapped to listening periods. This feature reflects the average daily temperature (°C) corresponding to the time a track was popular or listened to.

The integration of audio and contextual data allowed for a comprehensive analysis of factors influencing Valence, bridging the gap between intrinsic musical properties and external environmental conditions.

[Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.]

## 2.3 Outcome variable

The distribution of Valence shows a relatively uniform spread, with some concentration near the mid-range. This reflects the diversity of emotional tones in the dataset. The histogram of Valence illustrates this distribution, where most tracks lie between moderate to high levels of positivity.

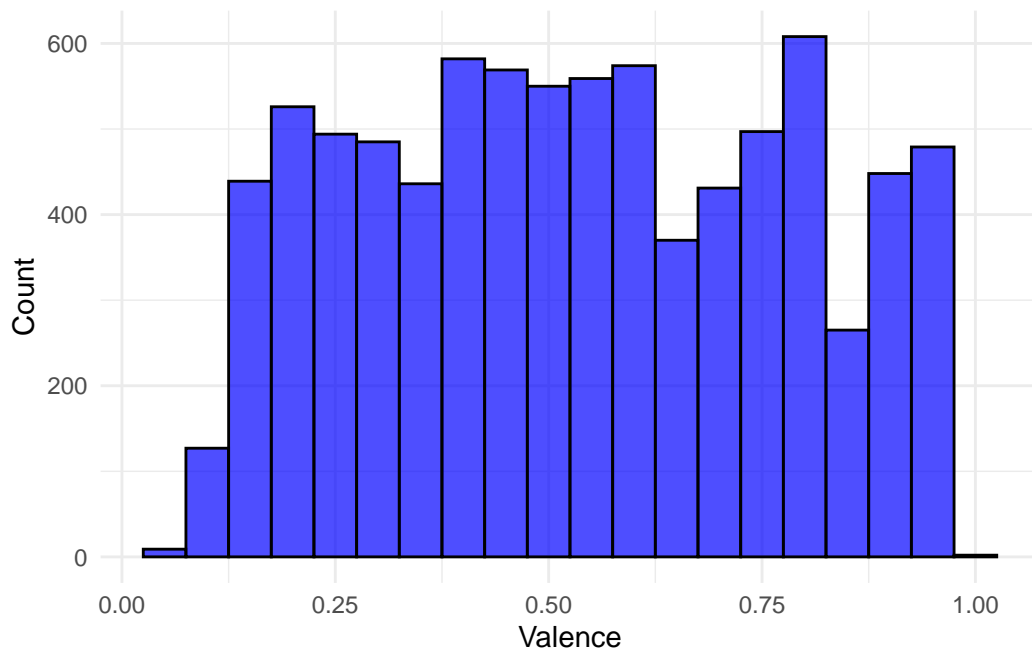


Figure 1: Distribution of Valence

## 2.4 Predictor variables

### 2.4.1 Mean Temperature

The average temperature over time demonstrates a seasonal decline, ranging from approximately 25°C in the summer to below 10°C in late autumn. This variable adds a contextual layer, capturing external influences that may correlate with song Valence and listening behavior.

### 2.4.2 Tempo

The distribution of Tempo reveals a central peak around 120 BPM, with a wider spread for slower and faster songs. This suggests a balanced mix of tempos, with a slight preference for moderate-speed tracks. Tempo's role in influencing Valence is analyzed alongside other musical features.

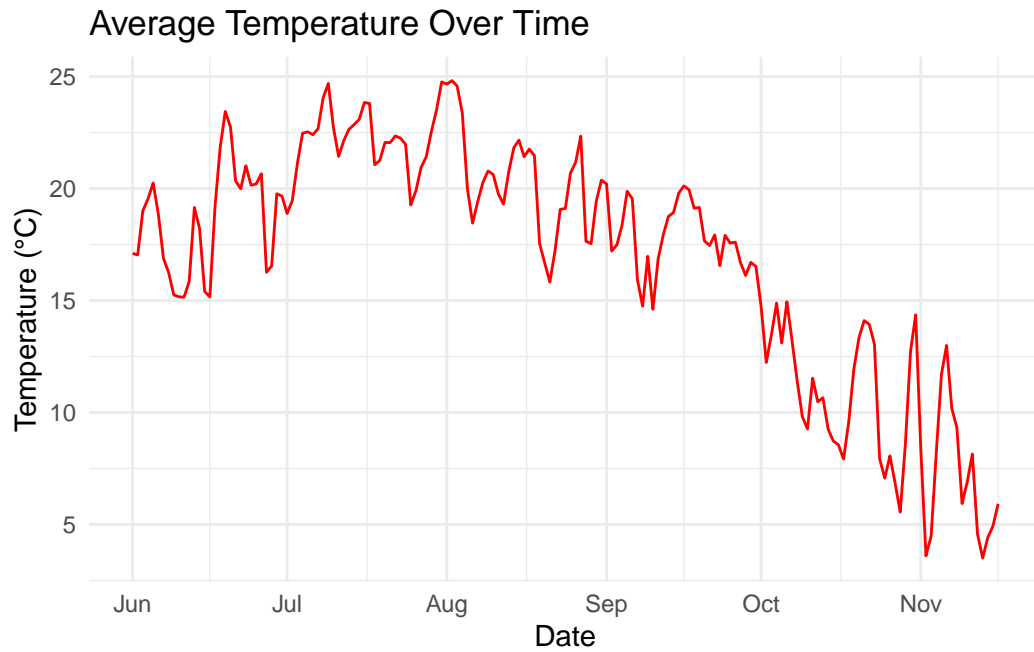


Figure 2: Average temperature Over Time

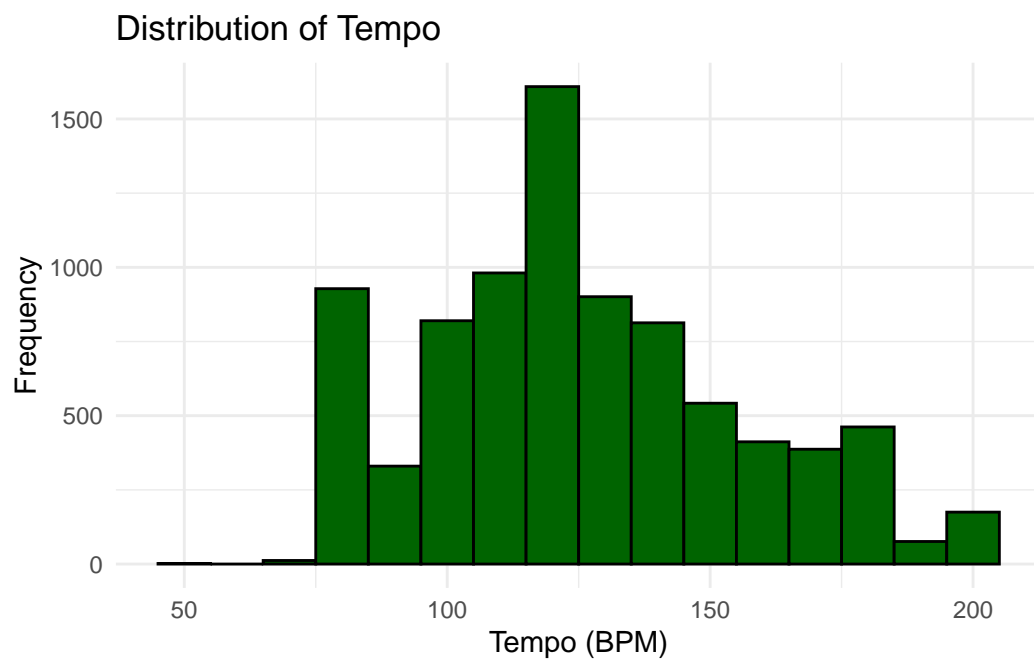


Figure 3: Distribution of Tempo

### 2.4.3 Danceability, Energy, and Acousticness

The comparative histograms of Danceability, Energy, and Acousticness highlight their distinct distributions:

Danceability: Peaks near 0.6 to 0.8, suggesting that most songs are moderately to highly danceable.

Energy: Similarly peaks in the mid-range, reflecting a tendency towards active, engaging tracks.

Acousticness: Displays a skewed distribution, with most tracks having low acoustic properties, indicating a focus on electronic or produced music styles.

Each of these variables provides unique explanatory power, contributing to the overall understanding of what drives the emotional tone in music.

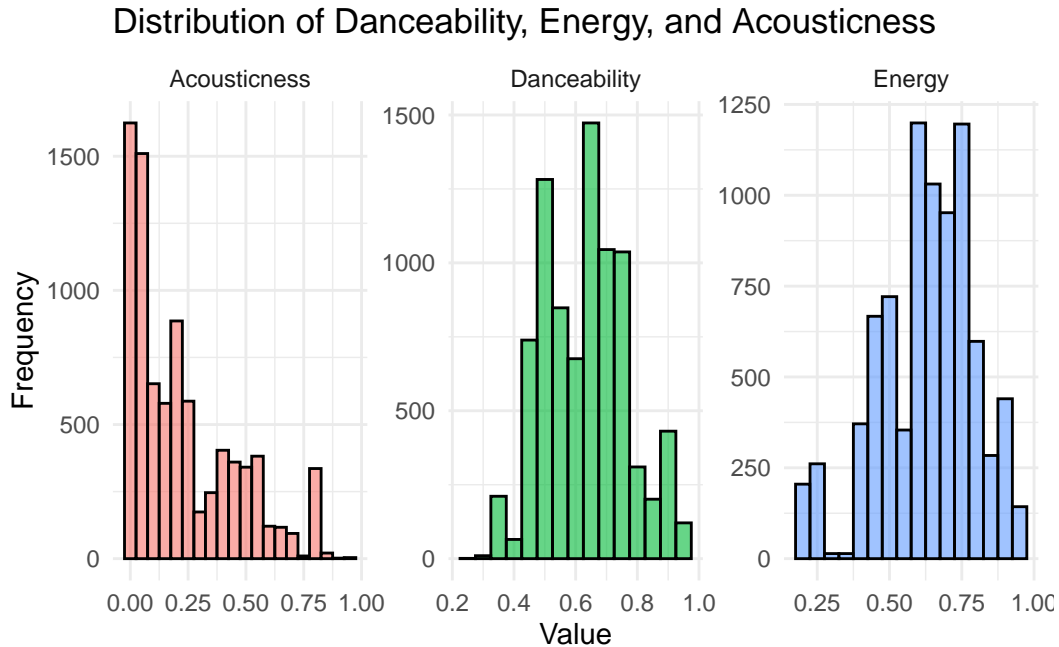


Figure 4: Distribution of Tempo

## 3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix [B](#).

### 3.1 Model Structure

- $y_i$ :
- $\beta_0$ :
- $\beta_1$ :
- ... [add beta's for the other predictors]
- ...
- $u_{\text{pollster}}$ :
- $u_{\text{state}}$ :

#### 3.1.1 Model justification

Model	# Variables	RMSE	R-Squared
MLR 1	6	0.0846	0.8676
MLR 2	5	0.0866	0.8600
MLR 3	4	0.2031	-1.4266

Comparison of Model Performance

## 4 Results

Our results are summarized in Table ??.

## 5 Discussion

### 5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

### 5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.



### **5.3 Third discussion point**

### **5.4 Weaknesses and next steps**

Weaknesses and next steps should also be included.

## Appendix

### A Additional data details

### B Model details

#### B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected  
by, the data

#### B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC algo-  
rithm

## References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. <https://open.toronto.ca/dataset/deaths-of-shelter-residents/>.