Predicting Music Emotion: Insights from Canadian Weather and Spotify Audio Features*

Danceability Dominates, while Temperature Plays a Minor Role in Shaping Positivity in the Top 50 Charts

Cristina Su Lam

December 3, 2024

This study explores how musical features and temperature shape emotional perceptions of music, measured by valence—a scale capturing a track's positivity. Using Spotify's top 50 Canadian charts and population-weighted mean temperatures from six major cities, we applied a multiple linear regression model. Danceability emerged as the strongest predictor of valence, with an effect size of 1.260, while scaled temperature had a smaller but significant effect of 0.007. These findings highlight the interplay between intrinsic audio features and environmental contexts in influencing listener sentiment, offering insights into music's emotional dynamics and practical applications for personalized playlists, mood-based marketing, and further research on cultural and environmental influences.

Table of contents

*Code and data are available at: https://github.com/cristinaasu/Spotify_Analysis

1 Introduction

Music is a deeply embedded and transformative part of people's lives, offering experiences that range from solitary headphone listening on a chilly day to attending global live events. It fosters empathy, social bonding, and cultural understanding, connecting individuals across cultural and linguistic differences (Clarke, DeNora, and Vuoskoski 2015). A UK study by Anglada-Tort et al. (2023) revealed that musical preferences are influenced by weather; warmer temperatures encourage songs with high intensity and positive emotions, while increased rainfall sees a decline in their popularity. This underscores how weather significantly shapes the musical landscape.

The digital age has provided unprecedented access to music data, with platforms like Spotify offering rich datasets on song characteristics (Thompson et al. 2024). Inspired by the previous research, this study aims to bridge the gap by examining the interplay of musical and environmental factors on listener emotions within the Canadian context. Using valence—a metric measuring emotions from 0 (negative or sad) to 1 (positive or happy)—the study evaluates how these elements collectively shape listener experiences.

Valence, the estimand of interest in this study, represents the true effect this research seeks to understand. While it itself is unobservable, this study estimates it through a structured analytical approach, ensuring that the methodology remains sharply focused on the central research question. As an abstract quantity, it captures the theoretical relationship independent of sample-specific variations, allowing the study to delineate true effects from noise and systematic biases inherent in observational data.

Using a multiple linear regression framework, this study analyzed data from Spotify's Top 50 Canadian charts, combined with temperature data from six major cities, spanning June to November 2024. Danceability emerged as the strongest predictor of valence, underscoring its dominant influence on listener emotions. Artist-specific effects, such as the positive contributions of Brenda Lee's tracks and the negative impact of Future's music, were also notable. Interestingly, temperature—despite being the primary variable of interest—had a relatively minor influence, suggesting that weather exerts a limited direct effect on valence. These findings highlight the need to integrate contextual and musical factors to design personalized recommendations and mood-based marketing strategies that resonate with diverse listener preferences.

The remainder of this paper is structured as follows. Section 2 describes the dataset and measurement process, while Section 3 outlines the methodology and model validation. Section 4 presents the results, highlighting the key predictors of valence and their implications. Finally, Section 5 discusses the findings, limitations, and potential avenues for future research.

2 Data

2.1 Overview

To analyze the relationship between song characteristics, temperature, and valence–data from Spotify and weather records from Canada's most populous cities were integrated. To be more specific, the dataset included daily Top 50 songs in Canada from Spotify Charts (Spotify 2024), covering the period from June 1, 2024, to November 16, 2024. Using the Spotify API

(Thompson et al. 2024), features such as Valence, Danceability, Acousticness, and Tempo were retrieved. Concurrently, weather data was obtained from Environment and Climate Change Canada (2024) for six major cities: Toronto, Montreal, Edmonton, Calgary, Ottawa, and Vancouver.

Data cleaning was conducted in R (R Core Team 2023) using the tidyverse (Wickham et al. 2019) and arrow (Richardson et al. 2024) packages. The cities mentioned were selected due to their large populations, and a new numerical variable, Mean Temp, was created by weighting each city's mean daily temperature by its population density and dividing by the total population density. This weighting was applied to account for the varying influence of each city based on its population size, ensuring that the derived temperature better reflects the overall climate experienced by the majority of the population. The resulting dataset consists of 8,450 observations and includes key variables such as Valence, Danceability, Acousticness, Tempo, Temperature, and Artist. The dataset was thoroughly checked for missing or incomplete values, confirming that all entries were complete and suitable for analysis. A sample of the cleaned dataset is displayed in Table 1.

For visualization and analysis, the knitr (Xie 2014), ggplot2 (Wickham 2016), and kableExtra (Zhu 2024) packages were utilized to generate professional tables and visualizations, effectively enhancing the presentation of results.

Table 1: Sample of Song's Characteristics

Artist	Song	Valence	Mean Temp	Dance	Acoustic	Tempo	Date
Post Malone, Morgan Wallen	I Had Some Help (Feat. Morgan Wallen)	0.731	17.1	0.638	0.008	127.986	Jun 01
Shaboozey	A Bar Song (Tipsy)	0.599	17.1	0.722	0.072	80.969	Jun 01
Eminem	Houdini	0.889	17.1	0.936	0.029	127.003	Jun 01
Sabrina Carpenter	Espresso	0.690	17.1	0.701	0.107	103.969	Jun 01
Kendrick Lamar	Not Like Us	0.214	17.1	0.898	0.011	101.061	Jun 01

Table 2 presents the summary statistics of the numerical variables, revealing key insights into the dataset's characteristics and patterns. Valence, ranging from 0 to 1, spans 0.036 to 0.981 with a mean of 0.532, reflecting an overall moderate mood among tracks. Mean Temperature, calculated as a population-weighted average across major Canadian cities, has a mean of 16.9°C, with values spanning from 3.5°C to 24.8°C, showcasing the climatic variation during the study period.

Danceability and Acousticness, both measured on a 0 to 1 scale, highlight contrasting patterns. Danceability values range from 0.234 to 0.943, with a mean of 0.632, indicating that most songs are moderately to highly danceable. Acousticness, ranging from 0.000 to 0.968, has a lower mean of 0.228, suggesting that non-acoustic tracks—those with more electronic or synthetic instrumentation—dominate the dataset.

Finally, Tempo, measured in beats per minute (BPM), has a mean of 126 BPM, with a wide range from 48.7 BPM for slower tracks to 203.8 BPM for faster ones. This diversity in tempo underscores the variety of musical styles, forming a robust basis for exploring their relationship with valence.

Table 2: 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
Acousticness	0.228	0.224	0.000	0.968
Tempo	126.157	29.921	48.718	203.812

2.2 Measurement

As noted in Section 2.1, the weather data originates from Environment and Climate Change Canada (2024), which collects meteorological observations from a comprehensive network of weather stations across the country. These stations measure key variables such as temperature, precipitation, and wind speed using standardized instruments and methodologies. Data is captured hourly or daily, depending on the station, and is validated through automated quality control processes and manual verification. This rigorous approach ensures reliable and precise measurements that meet international meteorological standards. In the context of this analysis, one weather station per city was selected based on relevance and proximity; additional details on this selection process are provided in Section C.1.

Similarly, the music data originates from Spotify Charts (Spotify 2024), which rank the most-streamed songs in a given region daily. These rankings are compiled by aggregating anonymized user streaming data across its platform. Streams are verified to ensure authenticity, eliminating artificial plays generated by bots or fraudulent activity. Additionally, Spotify employs machine learning models to analyze the audio features of each song, such as Valence, Danceability, Acousticness, and Tempo. These features are derived from the acoustic and digital signal processing of tracks, providing a quantitative representation of each song's characteristics.

It is important to note that in this dataset, daily temperature values are repeated across all fifty songs for a given day, as they represent an average measurement for the day rather than song-specific conditions. Similarly, artist entries recur within the dataset, reflecting the daily rankings of the Top 50 streamed songs, which often include repeated tracks for artists with sustained popularity or multiple charting songs.

The combination of these trusted data sources ensures that the dataset used in this study is grounded in accurate and validated measurements. Weather data reflects real-world environmental conditions, while Spotify's data captures both user engagement and precise audio analysis, enabling a comprehensive analysis of the relationship between temperature, song features, and valence.

2.3 Analysis of Variables

2.3.1 Outcome variable

The Valence variable exhibits a fairly uniform distribution, as shown in Figure 1, with values ranging from 0.036 to 0.981, as detailed in Table 2. This wide range captures a broad spectrum of emotional tones in the dataset.

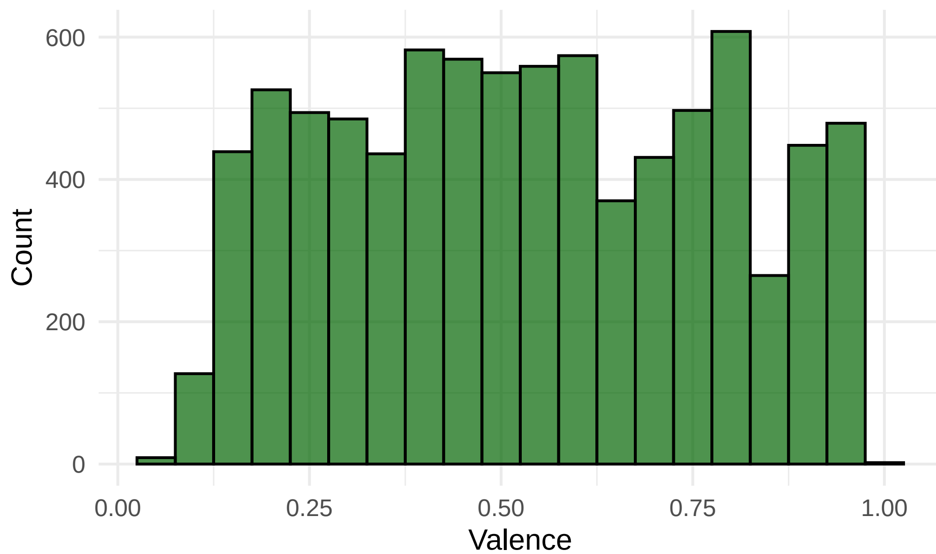

Figure 1: Uniform Distribution of Valence

Table 3 highlights the Top 5 songs with the highest Valence, led by "September" by Earth, Wind & Fire (0.981), alongside other highly positive tracks like "Apple" by Charli XCX (0.962) and "Super Freak" by Rick James (0.962). These songs represent the most uplifting and positive moods in the dataset.

Table 3: Top 5 songs with the Highest valence

Song	Artist	Valence
September	Earth, Wind & Fire	0.981

Table 3: Top 5 songs with the Highest valence

Song	Artist	Valence
Apple	Charli xcx	0.962
Super Freak	Rick James	0.962
Too Sweet	Hozier	0.960
HOT TO GO!	Chappell Roan	0.957

In contrast, Table 4 features the Top 5 songs with the lowest Valence, including "BLUE" by Billie Eilish (0.0365) and "Meteor Man" by Lil Uzi Vert (0.0384), which reflect darker or more somber tones. Together, these extremes illustrate the dataset's diversity, capturing a broad emotional spectrum across tracks.

Table 4: Top 5 Songs with the Lowest Valence

Song	Artist	Valence
BLUE	Billie Eilish	0.0365
Meteor Man	Lil Uzi Vert	0.0384
Drugs You Should Try It	Travis Scott	0.0729
Harry Potter - Trap Remix	Trap Remix Guys, Trap Remix Guy	0.0812
READY TO COOK UP	Future	0.0877

2.3.2 Predictor variables

- **Mean Temperature**

Figure 2 shows the average temperature over time, demonstrating a seasonal decline from approximately 25°C in the summer to below 5°C in late autumn. This pattern aligns with expected seasonal transitions in Canada and is further supported by Table 2, where the Mean Temperature averages 16.9°C with a standard deviation of 5.298°C. This notable variability provides an opportunity to investigate potential correlations between environmental factors, such as weather, and shifts in musical preferences reflected in Valence.

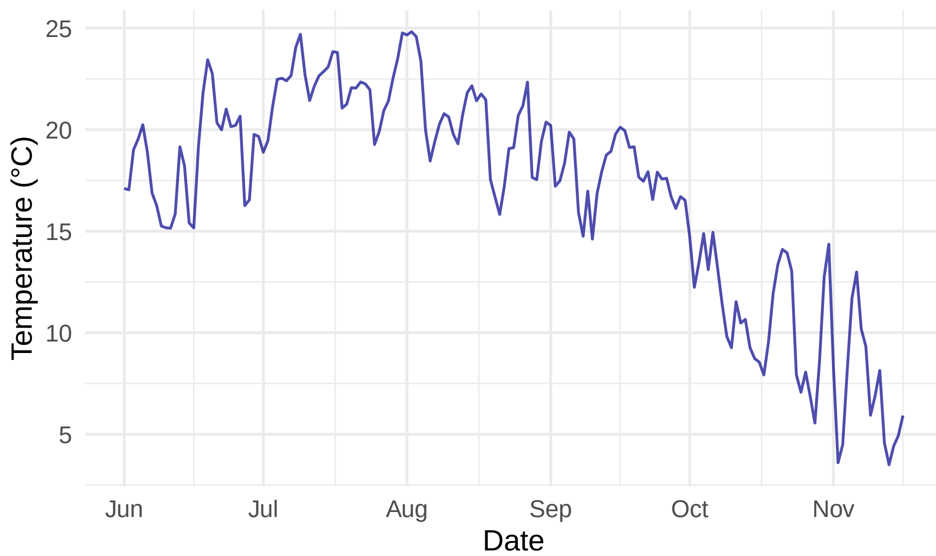

Figure 2: Average Temperature from June to November 2024

- **Artist**

The variable Artist reveals intriguing patterns in the dataset. Displayed in Figure 3, the top 10 artists, led by Sabrina Carpenter and Zach Bryan, contributed the highest number of entries to the dataset, with their songs appearing repeatedly in the daily Top 50 rankings, each exceeding 600 occurrences. This underscores their strong presence on the charts during the study period. However, despite their prominence, their songs do not feature in Table 3 or Table 4. In contrast, artists like Chappell Roan and Billie Eilish reflect a broader emotional range, appearing in the highest and lowest Valence categories, respectively. Hozier, contributing the fewest entries among the Top 10, stands out for appearing in Table 3, showcasing the emotional positivity of some of his music despite a smaller overall presence.

- **Tempo**

The distribution of tempo values, as depicted in Figure 4, shows a roughly normal shape, with most tracks falling between 100 and 130 beats per minute (BPM) and peaking around 120 BPM. This aligns with findings that today's popular music often falls within the 100–140 BPM range (MasterClass Staff 2024), highlighting rhythmic patterns that resonate with mainstream listener preferences.

In addition, the distribution reveals outliers on both ends of the spectrum, with slower tempos below 60 BPM and faster tempos above 180 BPM representing niche genres or unique musical styles.

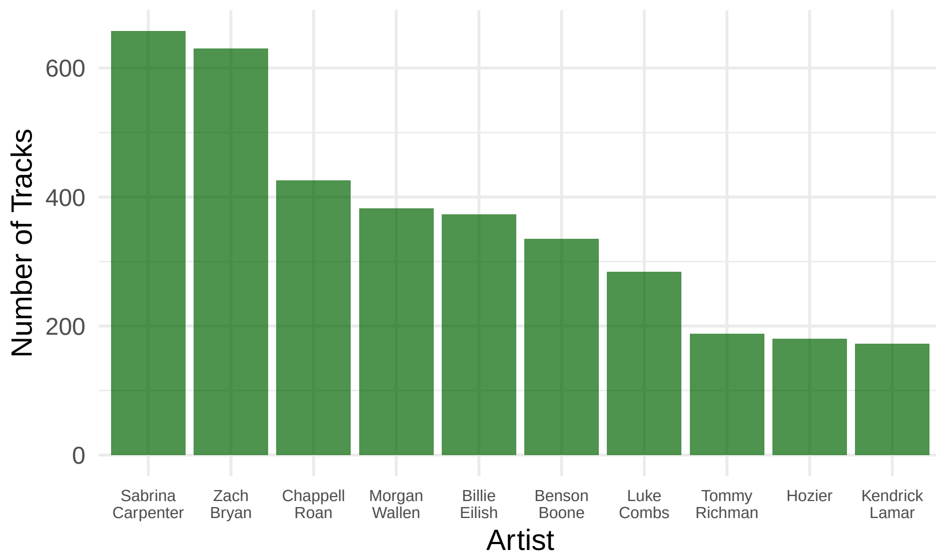

Figure 3: Sabrina Carpenter and Zach Bryan leads the Top 10 artists

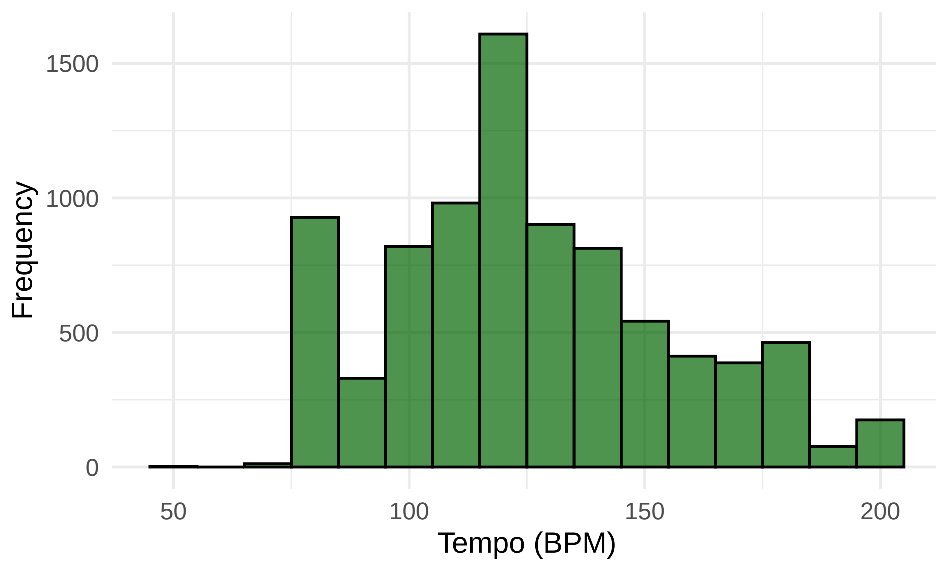

Figure 4: Normal Distribution of Tempo

- **Danceability and Acousticness**

The Figure 5 below highlights distinct patterns in the distributions of Danceability and Acousticness.

Acousticness displays a strongly right-skewed distribution, with most values concentrated near 0, reaffirming the prevalence of non-acoustic tracks in the dataset as noted in Section 2.1. This pattern aligns with its low mean of 0.228 and relatively high variability, as shown in Table 2, capturing a mix of purely electronic tracks and a smaller presence of acoustic elements.

In contrast, Danceability demonstrates a more uniform distribution, peaking between 0.6 and 0.8. With a higher mean of 0.632 and a standard deviation of 0.134, this indicates that the majority of tracks are moderately to highly danceable. This distribution underscores the dataset's emphasis on popular music, where rhythmic appeal and danceability are often prioritized.

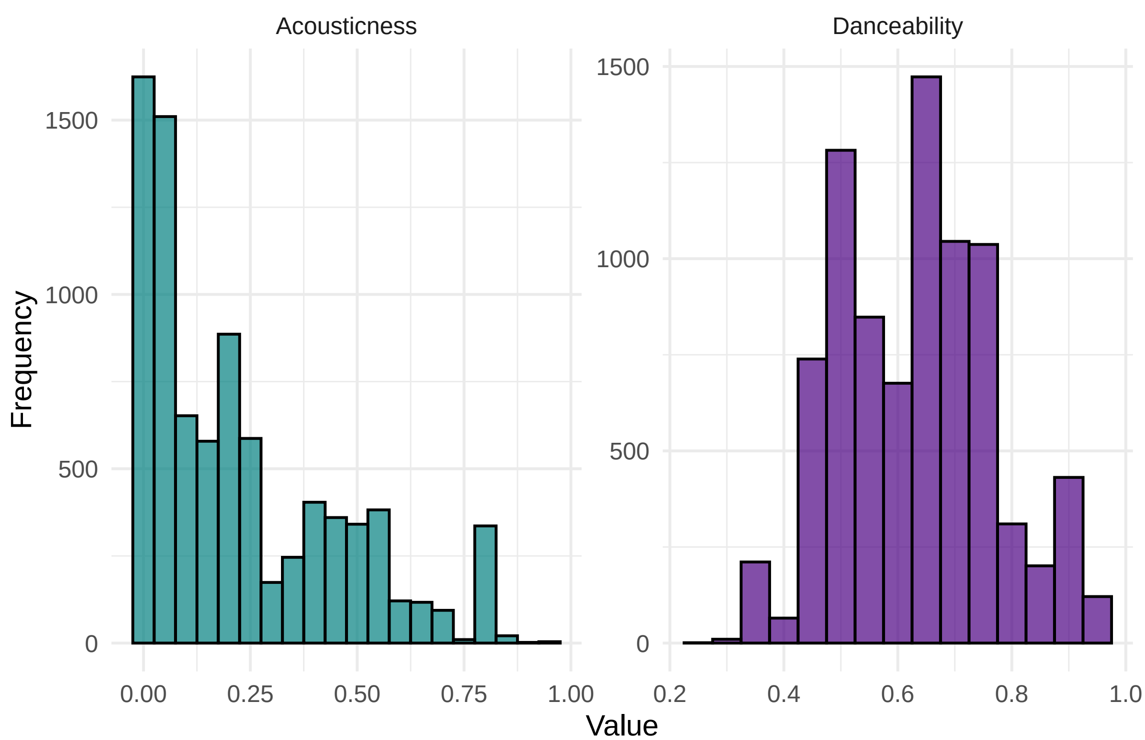

Figure 5: Acousticness highlights tracks with lower acoustic qualities, while Danceability shows a balanced distribution across rhythmic and energetic levels.

2.3.3 Exploring Relationships

- **Valence vs. Temperature**

The heatmap in Figure 6 illustrates that higher temperatures (20-25°C) are associated with a greater frequency of tracks exhibiting moderately positive valence values, clustering around 0.5-0.75. In contrast, at lower temperatures (5-15°C), songs are distributed more evenly between positive and negative valence values. Notably, the highest frequency is observed at a valence of approximately 0.25 when temperatures range between 15-20°C. The absence of data in the lowest temperature range (0-5°C) reflects a limitation of the dataset rather than an actual trend.

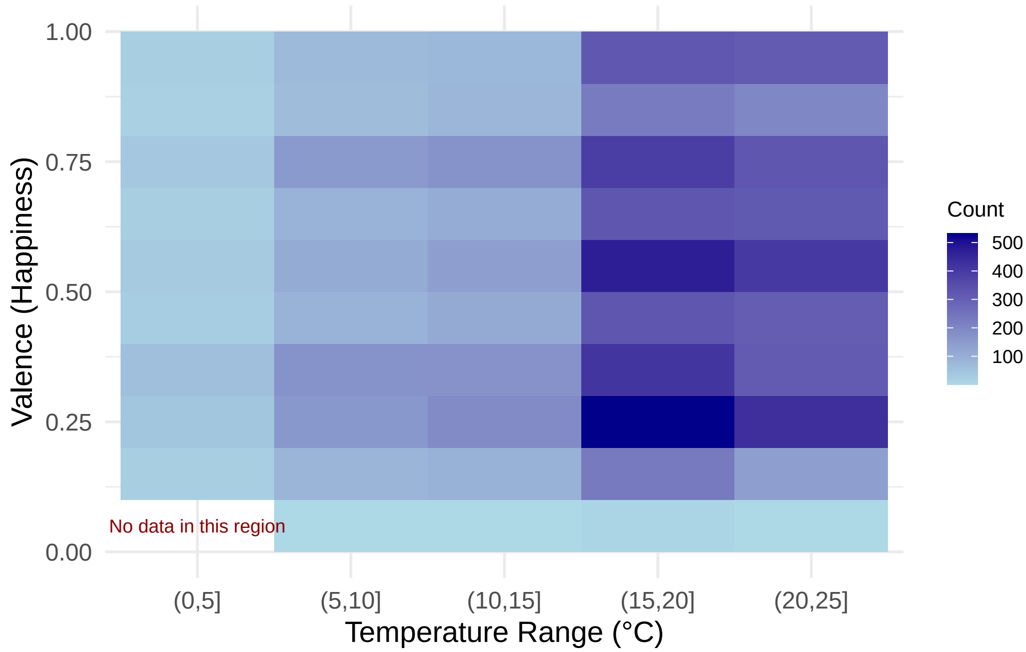

Figure 6: Higher temperatures align with positive valence values, while lower temperatures show a wider spread in emotional tones.

- **Valence vs. Artist**

The chart in Figure 7 highlights notable contrasts in the emotional tone of music produced by the Top 10 artists. Hozier, Tommy Richman, and Chappell Roan stand out with the highest average Valence scores, suggesting their tracks lean toward more positive and uplifting moods. This trend aligns with their reputations for creating emotionally resonant music. Conversely, Kendrick Lamar and Zach Bryan, who rank lowest in average Valence, demonstrate a preference for deeper or more introspective themes, reflecting their storytelling-driven styles.

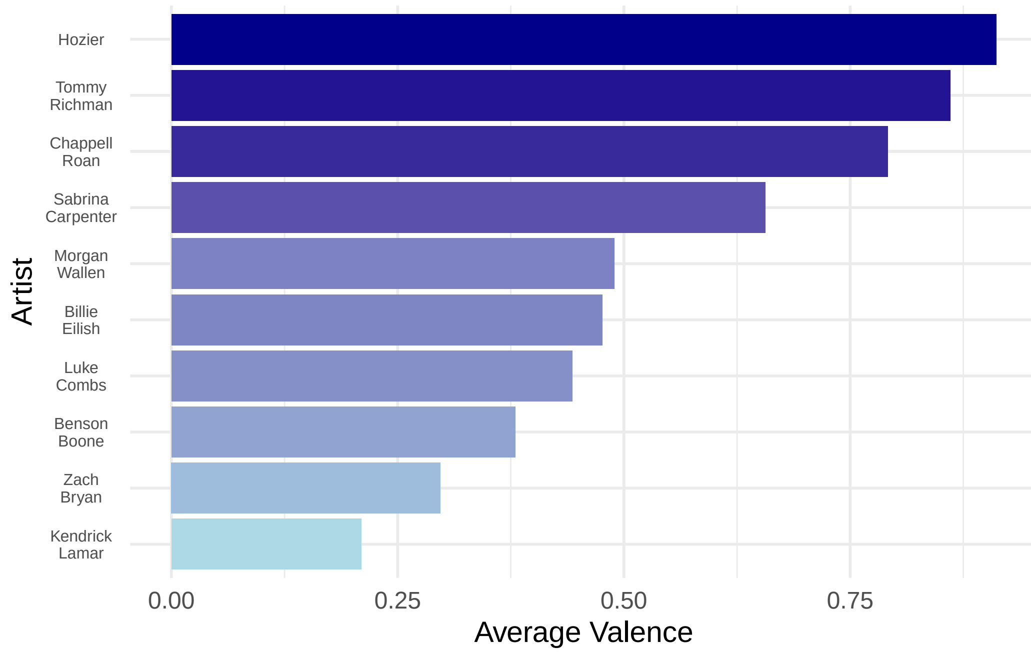

Figure 7: Hozier leads with the most positive emotional tone, while Kendrick Lamar features the lowest average valence among the top contributors.

3 Model

3.1 Overview

This section presents a multiple linear regression (MLR) model to predict song valence based on audio features and contextual factors. It evaluates how danceability, acousticness, scaled tempo, scaled mean temperature, and artist-specific effects influence valence, offering a clear framework for analyzing these relationships.

The analysis utilized the R packages stats (R Core Team 2023), MLmetrics (Yan 2024), and car (Fox and Weisberg 2019). The stats package enabled MLR modeling with the `lm` function, while MLmetrics calculated RMSE and R-squared for model evaluation. The car package assessed multicollinearity using Variance Inflation Factors (VIF).

3.2 Model Structure

Mathematical Representation:

$$
\begin{aligned}
y_i = \beta_0 + \beta_1(\text{Scaled Mean Temp})_i + \beta_2(\text{Artist})_i + \beta_3(\text{Danceability})_i \quad (1)\\
+ \beta_4(\text{Acousticness})_i + \beta_5(\text{Scaled Tempo})_i + \epsilon_i \quad (2)
\end{aligned}
$$

- y_i: Valence of the song, measuring perceived positivity.
- β_0: Baseline valence when predictors are at their reference levels.
- β_1: Effect of scaled mean temperature on valence.
- β_2: Artist-specific effects on valence.
- β_3: Effect of danceability on valence.
- β_4: Effect of acousticness on valence.
- β_5: Effect of scaled tempo on valence.
- ϵ_i: Unexplained variability, assumed to follow a normal distribution with mean 0 and variance.

3.3 Variable Selection

In constructing the model, careful consideration was given to selecting predictors that accurately capture the relationships influencing song valence:

- *Scaled Mean Temperature*: Represents the influence of environmental factors on emotional responses to music. Scaling normalizes its range for better interpretability and comparability across predictors.
- *Artist*: Captures the unique stylistic and emotional contributions of individual artists, as certain artists consistently produce music associated with higher or lower valence.
- *Danceability*: Captures a song's rhythm and energy, with higher values often evoking happiness and positivity, thereby increasing valence.
- *Acousticness*: Higher acousticness often linked to calmer, introspective tones that may lower valence, while lower acousticness corresponds to more electronic or dynamic tracks with higher positivity.
- *Scaled Tempo*: Reflects a song's speed, influencing its energy and emotional intensity, with faster tempos often enhancing valence. Scaling ensures consistency with other predictors, like Scaled Mean Temperature.

3.4 Model Validation

The model validation process involved out-of-sample testing, where the dataset was split into training and testing sets in a 70-30 ratio. Three variations of the multiple linear regression model were tested, differing in the number of predictors included. Model performance was evaluated using Root Mean Square Error (RMSE) and R-squared values. As shown in Table 5, the models demonstrate varying degrees of predictive accuracy and explanatory power.

Table 5: Comparison of Model Performance

Model	Variables	R-Squared	RSME
MLR 1	Scaled Mean Temp, Artist, Danceability, Acousticness, Scaled Tempo, Date	0.0868	0.8597
MLR 2	Scaled Mean Temp, Artist, Danceability, Acousticness, Scaled Tempo	0.0866	0.8601
MLR 3	Scaled Mean Temp, Danceability, Acousticness, Scaled Tempo	0.2031	-1.4267

After not taking into consideration Artist, MLR 3 includes fewer predictors, it performs significantly worse, as evidenced by its negative R-squared and much higher RMSE, indicating poor fit and predictive ability. On the other hand, MLR 2 demonstrates a slight performance advantage over MLR 1 in both RMSE and R-squared metrics.

3.5 Model Justification

The choice of MLR 2 is justified by its balance of predictive performance and model simplicity. The `Date` variable was excluded because it is indirectly related to `Scaled Mean Temp`, which already captures the temporal context more meaningfully, making its inclusion redundant and unnecessarily complex. Furthermore, this model includes the key variables of interest outlined in Section 3.3, ensuring it addresses the research objectives effectively. Lastly, MLR 2 slightly outperforms MLR 1 in terms of RMSE and R-squared, further supporting its selection as the final model for predicting song valence.

3.6 Model Diagnostics

To demonstrate why this model is optimal, the following tests and evaluations were conducted to assess its performance and reliability. A plot comparing the predicted and true valence values was generated to visually assess the model's performance, Figure 10. The dashed line represents the 1:1 relationship, indicating perfect predictions. Most points cluster around the line, showing that the model effectively captures the relationship between predictors and valence. However, some deviations are observed, which may indicate areas where the model could be improved.

The Variance Inflation Factor values, shown in Table 6, evaluate potential multicollinearity among predictors. Although `Artist` has a high raw VIF of 61.91 due to its many levels, the adjusted value of 1.01 in the last column accounts for its degrees of freedom, confirming that multicollinearity is not a concern for this categorical variable. Other predictors—`Scaled Mean`

`Temp`, `Danceability`, `Acousticness`, and `Scaled Tempo`—have VIF values below 6, further indicating no significant multicollinearity in the model.

Additionally, residual plots and Q-Q plots were generated as part of the diagnostics to evaluate model assumptions. A more detailed analysis of these plots and their implications is provided in Section 5.

Table 6: Variance Inflation Factor for Model Predictors

Predictor	GVIF	Df	GVIF^(1/(2*Df))
`Scaled Mean Temp`	1.33	1	1.15
Artist	61.91	155	1.01
Danceability	5.43	1	2.33
Acousticness	4.31	1	2.08
`Scaled Tempo`	2.83	1	1.68

4 Results

The results from the multiple linear regression (MLR) analysis reveal significant insights into the predictors of Valence, as shown in the coefficient summary in Table 7 and the corresponding coefficient plot Figure 8. These findings build upon the descriptive patterns of the outcome and predictor variables, detailed in Section 2.3.

Danceability is identified as the strongest predictor, with a substantial positive coefficient 1.188 and a t-statistic of 75.469, demonstrating its strong and statistically significant association with higher Valence. Its adjusted R^2 value of 0.081 indicates that this variable alone explains 8.1% of the variability in Valence, highlighting the central role of rhythmic and energetic qualities in shaping positive emotional responses to music. In Figure 5, the distribution of Danceability values, peaking at moderate to high levels, further underscores its prominence in popular tracks.

Scaled Tempo demonstrates a statistically significant positive effect on Valence, with a coefficient of 0.054 and a precise estimate supported by a small standard error of 0.002. As illustrated in Figure 4, the original distribution of Tempo, approximately normal with a peak around 120 BPM, provides important context for interpreting this relationship. Scaling centers and standardizes the variable without altering its underlying patterns or associations. According to Table 8, Scaled Tempo explains 1.8% of the variability in Valence, highlighting its subtle yet meaningful role in influencing musical positivity.

Table 7: The modeling originally includes all artists, but for readability, only the top 5 artists with the highest positive and negative impact on valence are presented here, alongside all other variables of interest.

Term	Estimate	Std. Error	Statistic
(Intercept)	0.059	0.017	3.513
Scaled Mean Temp	0.003	0.001	2.473
Andy Williams	0.407	0.085	4.793
Bobby “Boris” Pickett, The Crypt-Kickers	0.210	0.044	4.731
Brenda Lee	0.381	0.033	11.418
Dean Martin	0.331	0.085	3.892
Eminem, JID	-0.986	0.033	-30.309
Future	-0.776	0.037	-21.177
Future, Metro Boomin, Kendrick Lamar	-0.778	0.018	-43.222
Gunna	-0.820	0.019	-42.503
Kendrick Lamar	-0.868	0.015	-56.843
Van Morrison	0.268	0.019	14.023
Danceability	1.188	0.016	75.469
Acousticness	-0.259	0.008	-30.727
Scaled Tempo	0.054	0.002	35.226

Scaled Mean Temperature exhibits the smallest yet statistically significant positive coefficient (0.003, $t = 2.473$). Its seasonal variability, transitioning from 25°C to below 5°C as summarized in Table 2, reflects a subtle relationship with Valence. This aligns with the prior study referenced in Section 1 of Anglada-Tort et al. (2023), which found that higher temperatures tend to be associated with music featuring happier tones. However, the impact observed in this study is smaller than anticipated, consistent with its modest adjusted R^2 value of 0.0007 in Table 8. This finding suggests that while temperature does play a role in shaping musical perception, its effect is considerably less pronounced compared to key audio features like Danceability and Tempo.

In contrast, Acousticness exhibits a significant negative effect on the outcome variable, with a coefficient of -0.259, accounting for approximately 1.3% of its variability. This suggests that tracks with pronounced acoustic features, often associated with introspective or somber tones, are perceived as less positive. This finding aligns with the nature of acoustic music, which typically evokes calmer and more reflective emotions, contrasting with the dynamic and energetic qualities linked to higher levels of Valence.

Table 8: Predictors' Adjusted R^2 Contributions

Variable	Adjusted R^2
Artist	0.5884628
Acousticness	0.0134552
Scaled Mean Temp	0.0000730
Scaled Tempo	0.0176875
Danceability	0.0812390

The categorical variable Artist accounts for the largest portion of Valence variability, with an adjusted R^2 value of 0.589 in Table 8. This reflects the diverse and unique qualities that different artists bring to their music, including stylistic variations, genre-specific elements, emotional themes, and even their popularity. Unlike continuous predictors (e.g., Acousticness or Tempo), Artist captures a range of unmeasured attributes—such as vocal style, lyrical content, and production techniques—that significantly shape listener perception.

The variable's numerous levels, representing individual artists, further enhance its ability to explain variance in the emotional tone of songs. Each artist's unique influence contributes significantly to the model's complexity, resulting in substantially higher predictive accuracy compared to other predictors.

Figure 8 visually complements the statistical findings by summarizing the significance, direction, and magnitude of the predictors. Each point represents a predictor's estimated coefficient, with horizontal lines indicating the 95% confidence intervals. Predictors with larger t-statistics, reflected by narrower confidence intervals, demonstrate greater reliability in their estimated effects.

While both artists exhibit strong positive effects, Brenda Lee's narrower confidence interval indicates greater reliability compared to Andy Williams, despite his slightly higher coefficient. Conversely, Kendrick Lamar demonstrates a pronounced negative association, reflecting contrasting musical styles and emotional themes. The figure further emphasizes Danceability as the most impactful predictor while highlighting the nuanced contributions of Scaled Tempo, Scaled Mean Temperature, and artist-level effects, further validating the robustness of the MLR model.

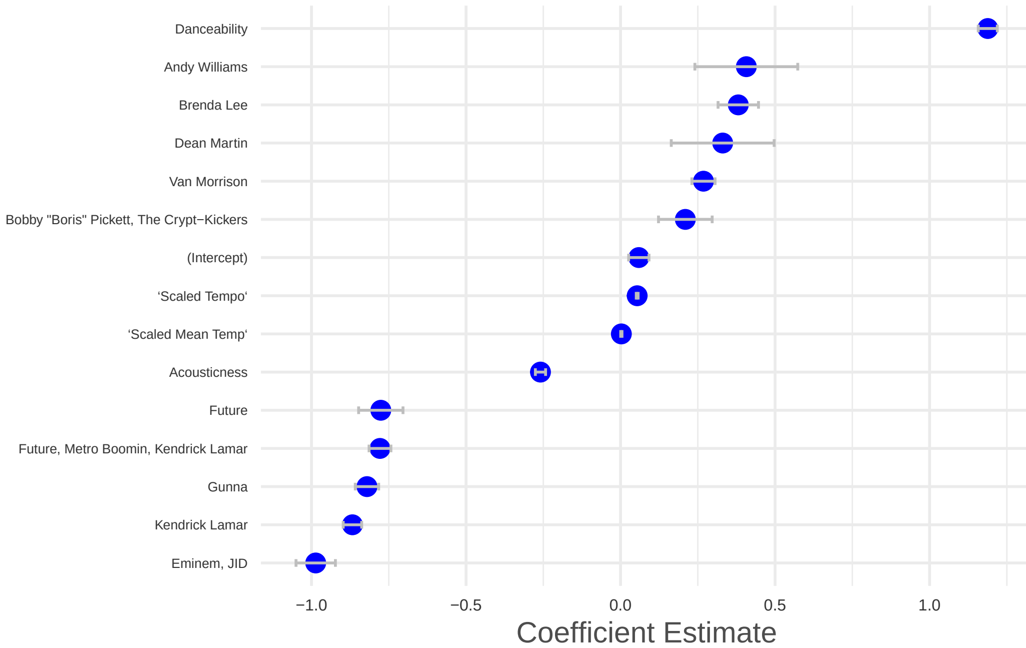

Figure 8: Impact of Key Features on Emotional Tone in Music

5 Discussion

This study applied a multiple linear regression (MLR) framework to explore the relationships between song characteristics, temperature, and valence, a metric representing emotional positivity in music. By integrating Spotify's audio feature data with population-weighted mean temperature, the analysis quantified the contributions of key predictors, including danceability, acousticness, tempo, and artist-specific effects, to valence. Statistically significant effects were identified for most predictors, while redundant variables like date were excluded to ensure model parsimony. The chosen model demonstrated a predictive power of 86% (R-squared), effectively capturing the observed relationships while leaving room for further refinement.

The findings highlight how intrinsic audio features and contextual factors shape listeners' emotional perceptions of music. Danceability emerged as the most influential predictor, underscoring the strong resonance of rhythmic and energetic qualities with perceptions of positivity. Tempo and acousticness also exhibited statistically significant effects: faster tempos aligned with more positive emotional tones, while acoustic tracks correlated with calmer or introspective moods. These results deepen our understanding of how specific musical attributes elicit emotional responses.

Beyond intrinsic features, contextual factors such as mean temperature subtly influenced listeners' musical experiences. Although its contribution was modest compared to predictors like danceability, the significant relationship between temperature and valence suggests that environmental contexts shape musical perception. This opens pathways for further exploration of how external factors like weather and cultural context influence musical preferences and emotional engagement.

5.1 Weaknesses of the Approach

Despite the robustness of the MLR framework, several limitations remain. In Figure 9, Residual and Q-Q plots revealed deviations from normality and heteroscedasticity, indicating that the model may not fully capture nonlinearities or interactions among variables. While the results are interpretable within the current framework, adopting alternative modeling techniques, such as generalized additive models, could better account for these complexities and improve predictive accuracy.

The reliance on population-weighted mean temperature introduces potential biases by overlooking localized environmental variations. Although this metric accounts for major population centers, it does not fully represent the diverse conditions experienced by listeners across Canada. Incorporating city-level or real-time weather data could provide a more precise understanding of temperature's impact on valence.

The analysis of artist-specific effects is confined to the Top 50 songs in Canada, excluding contributions from less prominent artists. This limitation restricts the generalizability of the findings, as lesser-known artists may introduce stylistic and emotional themes not captured in this dataset. Expanding the scope to include a wider range of artists would provide a more comprehensive perspective.

Finally, the study's six-month timeframe limits the ability to explore seasonal or long-term trends. While this period offers a snapshot of recent patterns, extending the dataset to include multiple years would allow for a deeper understanding of how musical preferences evolve over time and provide insights into potential seasonal influences.

These constraints emphasize the importance of future research refining and expanding upon this approach by incorporating advanced models, broader datasets, and more granular contextual information to better capture the intricate interplay of music, context, and emotion.

5.2 Future Directions and Next Steps

To address the identified limitations, future research could incorporate more granular contextual data, such as city-level weather patterns or regional cultural influences, to better capture localized environmental factors shaping listeners' emotional responses to music. Expanding the dataset to examine variations across different religions, cultural backgrounds, or linguistic groups within Canada would further enhance the findings' generalizability, reflecting the country's rich stylistic and cultural diversity.

Incorporating temporal trends presents another promising direction. Extending the dataset to span multiple years would allow for the investigation of seasonal patterns and long-term shifts in musical preferences, similar to the approach by Anglada-Tort et al. (2023). Adding variables such as lyrical content and genre could further enrich the analysis by providing insights into how lyrical themes and musical styles interact with listener sentiment, fostering a deeper understanding of the interplay between song characteristics and emotional perception.

Finally, adopting advanced modeling techniques, such as machine learning or Bayesian frameworks, could significantly improve predictive accuracy and reveal complex, nonlinear relationships among variables. These approaches would build upon the current findings, enabling a more sophisticated exploration of the multifaceted influences of music, context, and emotion while addressing the limitations of traditional regression-based models.

Appendix

A Survey, Sampling & Observational Data

A.1 Spotify Data Collection and Handling Process

Spotify collects streaming data from its extensive user base to generate insights into music consumption trends, as detailed in ("Spotify Charts," n.d.). A stream is recorded when a listener plays a song or music video for at least 30 seconds, and streams from downloaded content are included only if the listener goes online at least once every 30 days. These streams contribute to metrics such as all-time streams, release stream counts, and individual track plays. However, not all streams are chart-eligible; Spotify applies a proprietary formula to ensure that the charts accurately reflect legitimate user-driven activity.

Data Validation and Fraud Prevention

Spotify employs advanced machine learning algorithms to identify and exclude streams suspected of being generated fraudulently, such as those created by bots or through artificial manipulation. This ensures that charts are not distorted and accurately represent genuine user preferences. Importantly, this filtering process does not affect the royalties paid to artists, maintaining financial transparency.

Types of Charts

Spotify offers several chart types, updated daily and weekly, to capture different aspects of music trends:

- Global and Regional Charts: Reflect overall listening patterns across regions using aggregated, eligible streams.
- City Charts: Focus on popular tracks within specific cities, highlighting localized listening behaviors.
- Viral Charts: Rank songs gaining social media traction, based on rising play counts, sharing activity, and new listener discoveries.

Daily charts are published by 6 PM EST the day after the charting period, while weekly charts are released globally after the week ends on Thursday. Chart eligibility includes any song or album live on Spotify during the respective time period.

Handling Missing or Anomalous Data

Although Spotify does not disclose specific methodologies for addressing missing or incomplete data, it ensures data integrity by:

- Filtering irregular streams using machine learning algorithms.
- Applying consistent validation thresholds, such as the 30-second playback rule.

- Regular updates that minimize potential reporting gaps.

Strengths

- Granular Insights: Spotify provides detailed data at global, regional, and city levels, enabling nuanced analysis of user behavior.
- Fraud Detection: Machine learning algorithms effectively exclude non-genuine streams, preserving the credibility of its charts.
- Comprehensive Metrics: Including all-time streams, release-specific counts, and daily/weekly charts ensures a thorough view of music trends.

Weaknesses

- Active User Bias: Charts primarily reflect preferences of highly active users, potentially underrepresenting less frequent listeners or those engaging with niche genres.
- Exclusion of Non-charting Songs: The analysis focuses on a limited selection of songs included in Spotify's charts, such as the Top 100 for daily charts or the Top 50 for other charts, which restricts insights into underrepresented musical tastes and niche genres that do not appear in these rankings.
- Opaque Methodologies: Spotify's proprietary formula and data-cleaning processes are not publicly disclosed, hindering external validation.

This structured approach allows Spotify to capture meaningful trends in music consumption while addressing challenges such as fraudulent activity and regional representation. However, improvements in methodological transparency and broader inclusion of non-charting songs could further enhance its analytical scope.

A.2 Weather Data Collection and Handling Process

Environment and Climate Change Canada (ECCC), as detailed in "Basics of Measuring Climate Change" (n.d.), collects comprehensive weather and climate data through an extensive network of observation stations, satellite systems, and historical archives. These datasets form the foundation for climate analysis, weather forecasting, and research on environmental impacts, ensuring a robust and reliable basis for understanding Canada's diverse climatic conditions.

Data Sources and Methodology

1. **Observation Stations**
    - Over 600 weather stations across Canada adhere to World Meteorological Organization (WMO) standards, ensuring global comparability.
    - These stations record key atmospheric parameters, such as temperature, precipitation, and wind speed, at hourly, daily, and monthly intervals.

   - Advanced automated systems, like the Lansdowne House station, feature specialized instruments, including tipping bucket gauges (rainfall), snow depth sensors, and Stevenson screens (temperature and humidity).

2. **Satellite and Remote Sensors**
   - Satellites and weather balloons provide supplementary data, particularly in remote regions, capturing atmospheric profiles and surface observations.

3. **Historical Records**
   - Canada's climatological record dates back to 1840, with daily climate data available since 1870 and hourly observations since 1953.

4. **Gridded and Point Data**
   - Data is available as point data (individual stations) or gridded datasets, where spatial interpolation estimates conditions between stations.

Data Validation and Handling

1. **Quality Control**

- ECCC applies automated checks to identify anomalies or missing values, flagged with markers (e.g., "-99999" for missing data).
- Manual reviews correct biases introduced by factors such as station relocations, instrument changes, or urbanization effects.

2. **Missing and Anomalous Data**

- Missing values are flagged and interpolated using spatial techniques where appropriate.
- Metadata documents causes of data gaps, such as operational disruptions or equipment malfunctions.

3. **Spatial Interpolation**

- Non-station observations are estimated using gridded datasets, which incorporate data from nearby stations through spatial modeling techniques.

4. **Data Formats**

- Weather records are stored in various formats (e.g., hourly [HLY01], daily [DLY02, DLY04], monthly [MLY04]) with detailed metadata for traceability.

Strengths

- Comprehensive Coverage: Regularly updated data from a dense network of stations provides high temporal and spatial resolution.

- Rigorous Quality Control: Automated and manual quality control ensures reliability, even for long-term climate studies.
- Accessibility: Data availability in multiple formats (point and gridded) supports diverse research and policy applications.
- Transparency and Metadata: Extensive documentation enhances usability and fosters transparency.

Weaknesses

- Urbanization Effects: Changes in urban landscapes around stations can introduce biases in long-term temperature or precipitation records, requiring careful adjustments.
- Sparse Coverage in Remote Areas: Northern Canada's limited station density reduces granularity despite interpolation efforts.
- Complex Formatting: Intricate flagging systems may pose challenges for non-expert users.
- Reliance on Adjustments: Historical comparability relies on accurate corrections for station relocations, instrument upgrades, and methodological changes.

Linkages to Literature and Methodological Insights

- The use of 30-year reference periods aligns with WMO standards, ensuring international comparability and consistency in climate analyses.
- Spatial interpolation techniques, widely supported in climate research, enhance the usability of datasets for environmental modeling and forecasting.
- Peer-reviewed studies emphasize the importance of adjustments for urbanization and station relocations in preserving the integrity of long-term climate records.

This detailed framework underscores the robustness and challenges of managing extensive weather datasets. It highlights the importance of methodological rigor, continuous refinement, and technological advancements to ensure the accuracy and relevance of climate data.

A.3 Conclusion and Future Directions

This appendix highlights the methodological rigor and challenges associated with integrating observational data from diverse sources. By addressing these challenges and leveraging advanced data validation techniques, the study lays the groundwork for future research exploring the intersection of music, environment, and emotion. Future studies could enhance representativeness by incorporating additional data sources, such as user surveys or alternative streaming platforms, to validate and expand upon these findings.

B Idealized Survey & Methodology - $100K Budget

The aim of this research is to investigate the interplay between weather conditions and music consumption preferences across Canadian cities, leveraging Spotify's streaming data and Environment and Climate Change Canada's (ECCC) weather datasets. With a $100,000 budget, the hypothetical survey outlined below is designed to address some of the key limitations identified in the current research methodology, such as underrepresentation of smaller cities, lack of granular cultural insights, and temporal patterns in music preferences.

B.1 Sampling Approach

Target Population

- Primary: Active Spotify users in Canada, aged 18–65, residing in all Canadian cities with ECCC weather stations.
- Secondary: Broader music listeners engaging with other platforms, to ensure diversity and comparability.

Sampling Strategy

1. Stratified Random Sampling:

- City: All cities with weather stations to capture diverse climatic and cultural conditions.
- Age Groups: 18–24, 25–34, 35–44, 45–54, 55–65.
- Gender: Male, Female, Non-binary.
- Cultural and Linguistic Diversity: Representation of English, French, Indigenous, and multilingual speakers.

2. Sample Size: 3,000 participants, proportionally distributed across cities based on population size. This balance ensures representation from smaller cities, addressing their underrepresentation in previous datasets.

3. Data Collection Period: 12 months, capturing seasonal variations and long-term trends in weather and music preferences.

4. Temporal Trends and Contextual Variables: Incorporate temporal patterns by comparing seasonal shifts across cities and genres, enriching the dataset with cultural events and holidays.

B.2 Survey Implementation

Recruitment Strategy

1. Online Recruitment:

- Use Spotify ads, Google Ads, and social media platforms tailored to regional audiences.
- Partner with local organizations to boost participation in smaller cities.
- Budget: $30,000.

2. Incentives:

- Participants receive $10 gift cards, with targeted outreach to underrepresented demographics.
- Budget: $30,000.

3. Inclusive Recruitment:

- Outreach campaigns in Indigenous communities and multilingual regions to enhance representation.

Survey Design

Questions structured to elicit insights on music consumption, emotional responses, and cultural factors:

- How does weather affect your choice of music (e.g., calm, energetic, reflective)?
- Do you prefer certain genres or lyrics during specific seasons or weather conditions?
- How do cultural or linguistic backgrounds shape your music preferences?
- How does time of day (morning, afternoon, night) impact your music choices?
- What external factors (e.g., holidays, work environment, commuting) influence your listening habits?

B.3 Observational Data Collection

Integration with ECCC Weather Data

- Hourly Weather Metrics: Link survey responses to city-level hourly weather data, ie. temperature, precipitation, snow depth, humidity, and wind speed.
- Cultural Events & Holidays: Cross-reference with music trends during regional holidays and events.
- Air Quality Index (AQI): Analyze how air quality influences mood and listening preferences.
- City-Level Data: Match survey responses to Spotify's city-level charts, including valence, energy, danceability, and genre diversity.
- Temporal Trends: Examine changes in listening patterns by day, week, and season.

B.4 Addressing Limitations

This survey design tackles key limitations by broadening its geographic and demographic representation. By including all Canadian cities with weather stations, it ensures coverage beyond the usual focus on six major urban centers, offering a more comprehensive understanding of city-level variations in music preferences. Moreover, stratifying the sample by linguistic and cultural groups mitigates biases toward dominant cultural narratives, capturing a richer, more nuanced picture of regional and cultural influences on music consumption.

Additionally, the year-long data collection period enables the analysis of seasonal and temporal trends, uncovering how weather impacts music consumption across different times of the year. The inclusion of granular contextual variables, such as cultural events, air quality, and commuting behavior, further enhances the analysis by addressing non-climatic factors that shape music preferences. This approach provides a holistic view of the interplay between weather and music consumption patterns. Additional Plots

B.5 Data Validation and Modeling

To ensure data quality, the survey incorporates robust validation techniques. Attention checks, such as instructing respondents to select a specific answer, help identify inattentive responses. Logic checks are employed to catch inconsistencies, such as self-reported outdoor listening during severe weather conditions. Additionally, post-stratification weighting adjusts for demographic imbalances, ensuring the sample accurately reflects Canada's population distribution and enhances the reliability of the findings.

Advanced modeling approaches further refine the analysis by addressing complex relationships within the data. Bayesian Hierarchical Models capture city-level variability and nonlinear interactions, providing nuanced insights into regional differences. Machine learning models are also leveraged to predict how external factors like weather influence music preferences and emotional responses, enabling a deeper understanding of the interplay between climate and music consumption.

B.6 Budget Breakdown

Expense Category	Budget Allocation
Online Recruitment	30000
Incentives for Participants	30000
Data Processing & Validation	20000
Survey Software (e.g., Qualtrics)	10000
Miscellaneous Costs	10000

B.7 Conclusion

With a \$100,000 budget, this survey framework leverages stratified sampling, city-level weather data, and advanced modeling techniques to address current research limitations. By incorporating broader geographic representation, cultural diversity, and temporal trends, the proposed methodology enhances the understanding of how external factors shape music preferences, providing a richer and more credible foundation for future studies.

C Supporting Information

C.1 Weather Station Selection and Mean Temperature Calculation

To ensure the weather data accurately reflects conditions across Canada's major urban centers, a population-weighted sampling strategy was employed. This approach incorporates temperature data from six key cities: Toronto, Montreal, Vancouver, Calgary, Edmonton, and Ottawa. These cities were chosen based on their significant populations and their role as representative urban hubs in different regions of the country.

Daily temperature readings were sourced from the nearest reliable weather stations to each city: Toronto City for Toronto, McTavish for Montreal, Vancouver Harbour CS for Vancouver, Calgary International CS for Calgary, Edmonton Blatchford for Edmonton, and Ottawa CDA RCS for Ottawa. These stations were selected for their proximity to the city centers and their consistent, high-quality temperature data.

The mean temperature was calculated using a population-weighted average, ensuring that each city's contribution to the overall calculation was proportional to its population. The formula used was:

$$\text{Mean Temperature} = \frac{\sum(\text{City Population} \times \text{Temperature})}{\sum \text{City Population}} \tag{3}$$

Table 10 provides details on the population and weather station used for each city in the calculation:

Table 10: Population and Weather Stations for Major Canadian Cities

City	Station	Population
Toronto	Toronto City	6200000
Montreal	McTavish	4200000
Vancouver	Vancouver Harbour CS	2700000
Calgary	Calgary Int'l CS	1600000

Edmonton	Edmonton Blatchford	1500000
Ottawa	Ottawa CDA RCS	1400000

This method emphasizes urban areas, capturing the majority of Canada's population while balancing representativeness with practical feasibility. However, it may not fully account for weather variations in smaller cities or rural regions.

By prioritizing proximity and reliable data collection, this strategy ensures an accurate and regionally representative calculation of average weather conditions.

C.2 Summary of Related Research

The study Here Comes the Sun: Music Features of Popular Songs Reflect Prevailing Weather Conditions by Anglada-Tort et al. (2023) analyzed the relationship between weather conditions and audio features of songs that appeared in the United Kingdom's weekly top charts from 1953 to 2019, comprising 23,859 unique entries. The researchers examined three key weather variables: daily maximum temperature, hours of sunshine, and days with rainfall, alongside audio features such as energy, danceability, and valence.

The study found that high-arousal, positive music (e.g., energetic, danceable, and emotionally uplifting songs) was positively associated with higher temperatures and sunshine hours but negatively correlated with rainfall. These relationships were strongest during colder months when changes in weather conditions were most noticeable. By contrast, low-arousal, negative music (e.g., calm or melancholic songs) showed no significant relationship with weather. This research highlights how weather conditions can influence population-level preferences for music, particularly for songs with features reflecting positivity and activity.

C.3 Model's Assumptions

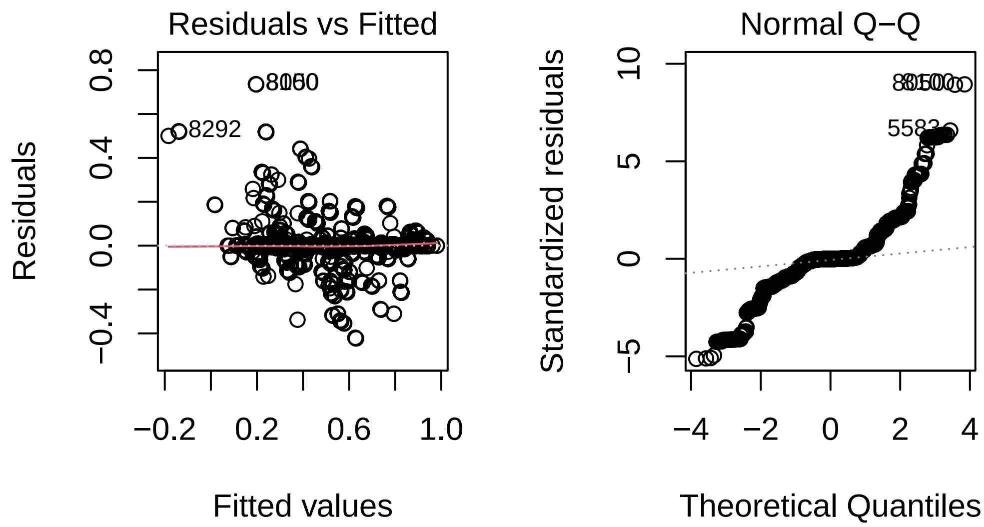

Figure 9: Residuals vs Fitted and Normal Q-Q Plot

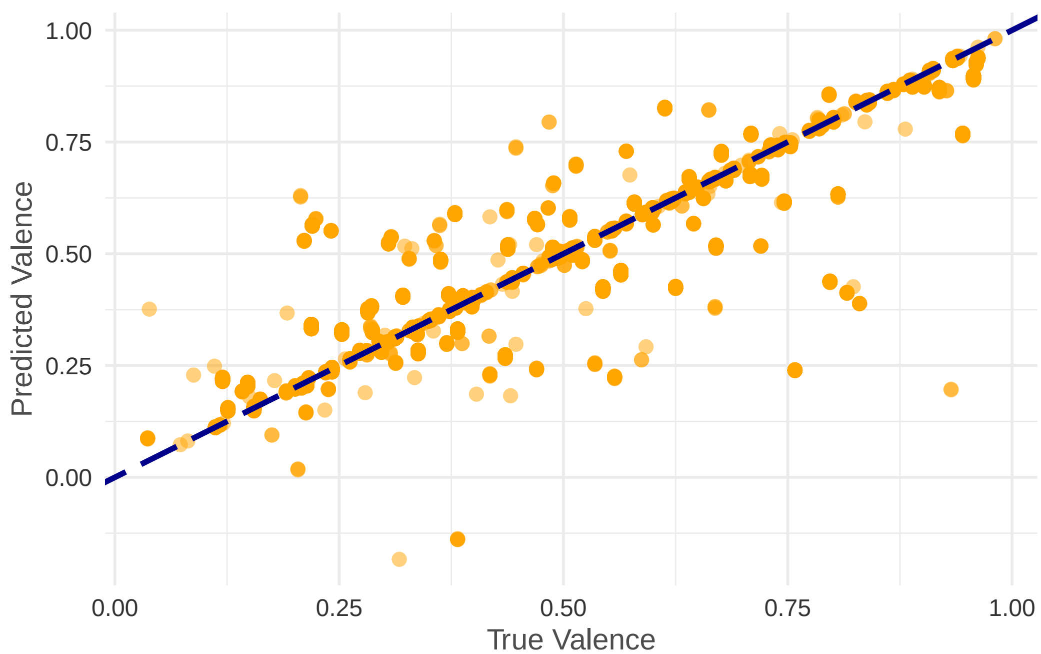

Figure 10: Comparison of Predicted and True Valence

References

Anglada-Tort, Manuel, Harin Lee, Amanda E. Krause, and Adrian C. North. 2023. "Here Comes the Sun: Music Features of Popular Songs Reflect Prevailing Weather Conditions." *Royal Society Open Science* 10 (221443). https://doi.org/10.1098/rsos.221443.

"Basics of Measuring Climate Change." n.d. Environment; Climate Change Canada. https://www.canada.ca/en/environment-climate-change/services/climate-change/canadian-centre-climate-services/basics/measuring.html.

Clarke, Eric F., Tia DeNora, and Jonna K. Vuoskoski. 2015. "Music, Empathy and Cultural Understanding." *Physics of Life Reviews* 15: 61–88. https://doi.org/https://doi.org/10.1016/j.plrev.2015.09.001.

Environment and Climate Change Canada. 2024. "Historical Climate Data." https://climate.weather.gc.ca/historical_data/search_historic_data_e.html.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

MasterClass Staff. 2024. "How to Find the BPM of a Song." https://www.masterclass.com/articles/how-to-find-the-bpm-of-a-song.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Spotify. 2024. "Spotify Charts: Global Overview." https://charts.spotify.com/charts/overview/global.

"Spotify Charts." n.d. Spotify. https://support.spotify.com/us/artists/article/charts/.

Thompson, Charlie, Daniel Antal, Josiah Parry, Donal Phipps, and Tom Wolff. 2024. *Spotifyr: R Wrapper for the 'Spotify' Web API*. https://CRAN.R-project.org/package=spotifyr.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in r." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

Yan, Yachen. 2024. *MLmetrics: Machine Learning Evaluation Metrics*. https://CRAN.R-project.org/package=MLmetrics.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.