Music and World Happiness: Analyzing Cultural Preferences Through Spotify
Data

Emily Aiken

Springboard Capstone 2

January 2025

1 INTRODUCTION

The relationship between music and happiness has long been studied at an individual level, but less is known about how national musical preferences relate to societal well-being. Using data from Spotify and the 2024 World Happiness Report, this project explores whether musical characteristics and diversity can indicate different levels of national happiness.

2 DATA

This study integrates multiple datasets to analyze the relationship between Spotify musical features and measures of World Happiness. Our analysis draws from two primary data sources:

1. World Happiness Report 2024 (Data for Figure 2.1) – National happiness scores and related metrics.

2. Top Spotify Songs in 73 Countries – Musical metrics from 73 countries, including features like danceability, energy, mode, and tempo (October-December 2023).

3 METHOD

The analysis approached the question by classifying countries into two groups (binary classification) based on whether their happiness scores were above or below the median happiness score of 6.278:

1. **Feature Engineering**:

   - Base musical characteristics: Core Spotify metrics (danceability, energy, tempo, mode) weighted by song popularity to better represent what people actually listen to

   - Musical variability: Standard deviations of musical features, measuring how much variation exists in a country's musical preferences (e.g., do they listen to songs with similar energy levels, or a mix of high and low energy songs?)

   - Musical diversity measures: Number of distinct levels of each musical feature, indicating the breadth of musical choices (e.g., do they listen to songs with many different tempos, or stick to a narrow range?)

2. **Classification Models**:

   - Random Forest

   - Logistic Regression

4 DATA CLEANING

The data cleaning process focused on ensuring temporal alignment and proper feature engineering:

Problem 1: Potential data leakage with happiness scores from 2024 and music data extending into 2025.
Solution: Filter Spotify data to Oct-Dec 2023 to align with happiness data collection period.

Problem 2: Basic musical metrics didn't capture the full picture of musical preferences.
Solution: Engineered new features:
1. Weighted averages by song popularity
2. Standard deviation metrics to capture variability (musical variability)
3. Diversity measures for each musical characteristic (musical diversity)

Problem 3: Different scales and units across features.
Solution: Standardized all features using StandardScaler before modeling.

5 EXPLORATORY DATA ANALYSIS

Key findings from exploratory data analysis:

1. Base Musical Features:
   - Strong negative correlation between danceability and happiness (-0.368, $p < 0.01$) (See Figure 1.)
   - Positive correlation with songs in major keys (0.366, $p < 0.01$) (See Figure 2.)
   - Moderate positive correlation with tempo (0.266, $p < 0.05$) (See Figure 3.)
2. Musical Variability:
   - Strong positive correlations between happiness and variability in:
     - Energy (0.359, $p < 0.01$)
     - Acousticness (0.334, $p < 0.01$)
     - Liveness (0.334, $p < 0.01$) (See Figure 4.)
3. Musical Diversity:
   - Strongest correlations of all categories
   - Liveness diversity (0.436, $p < 0.001$)
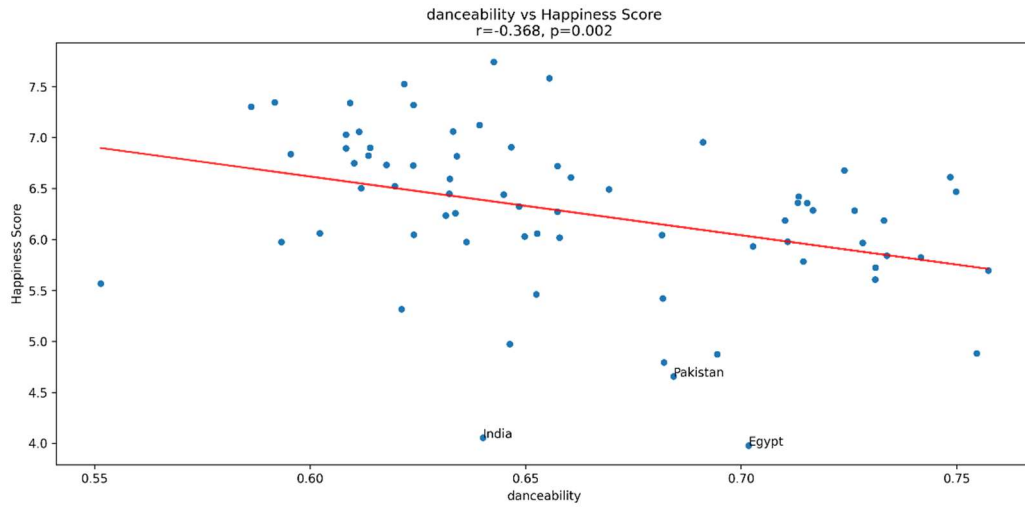   - Speechiness diversity (0.426, $p < 0.001$) (See Figure 5.)

Figure 1. Our analysis revealed a negative correlation between danceability and happiness.
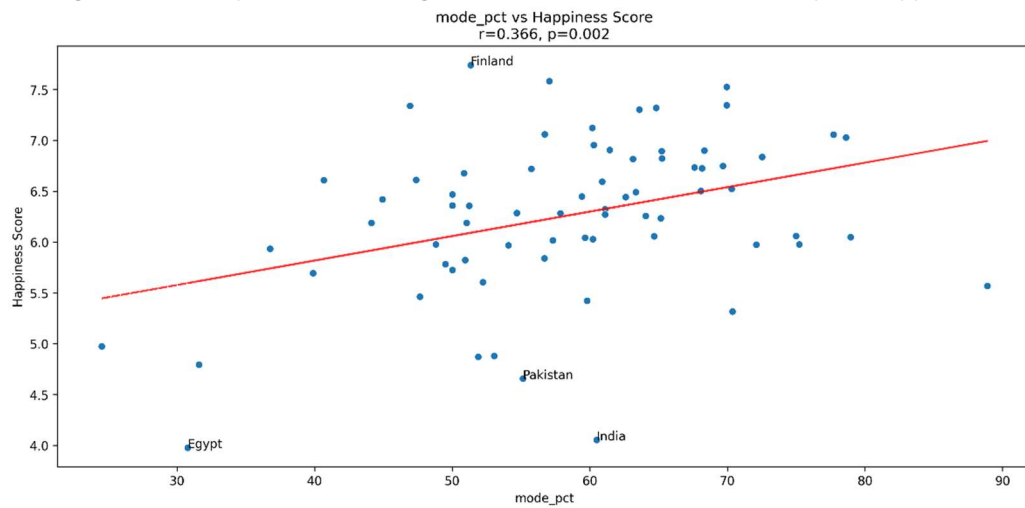

Figure 2. Our analysis also revealed a positive correlation with songs in the major key.
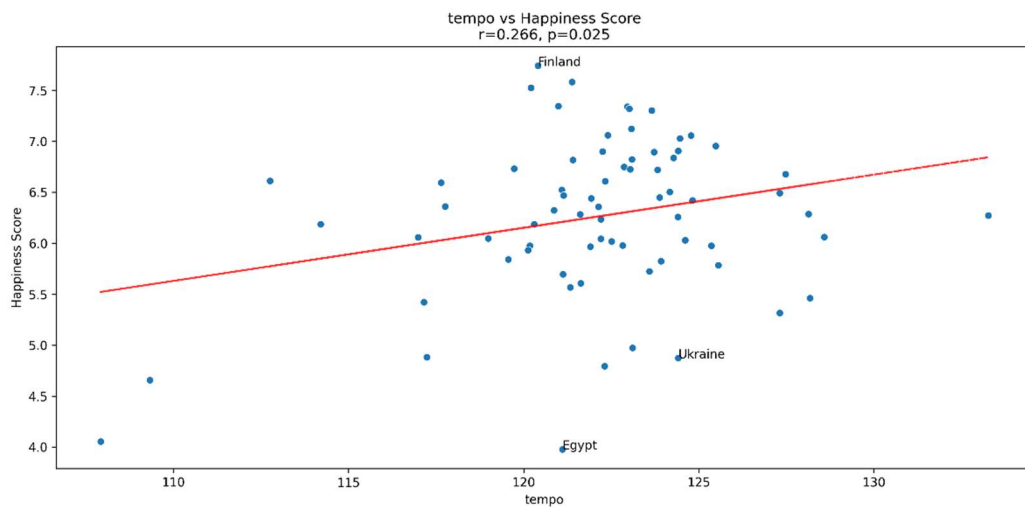

Figure 3. There is a moderately positive correlation between tempo and happiness.
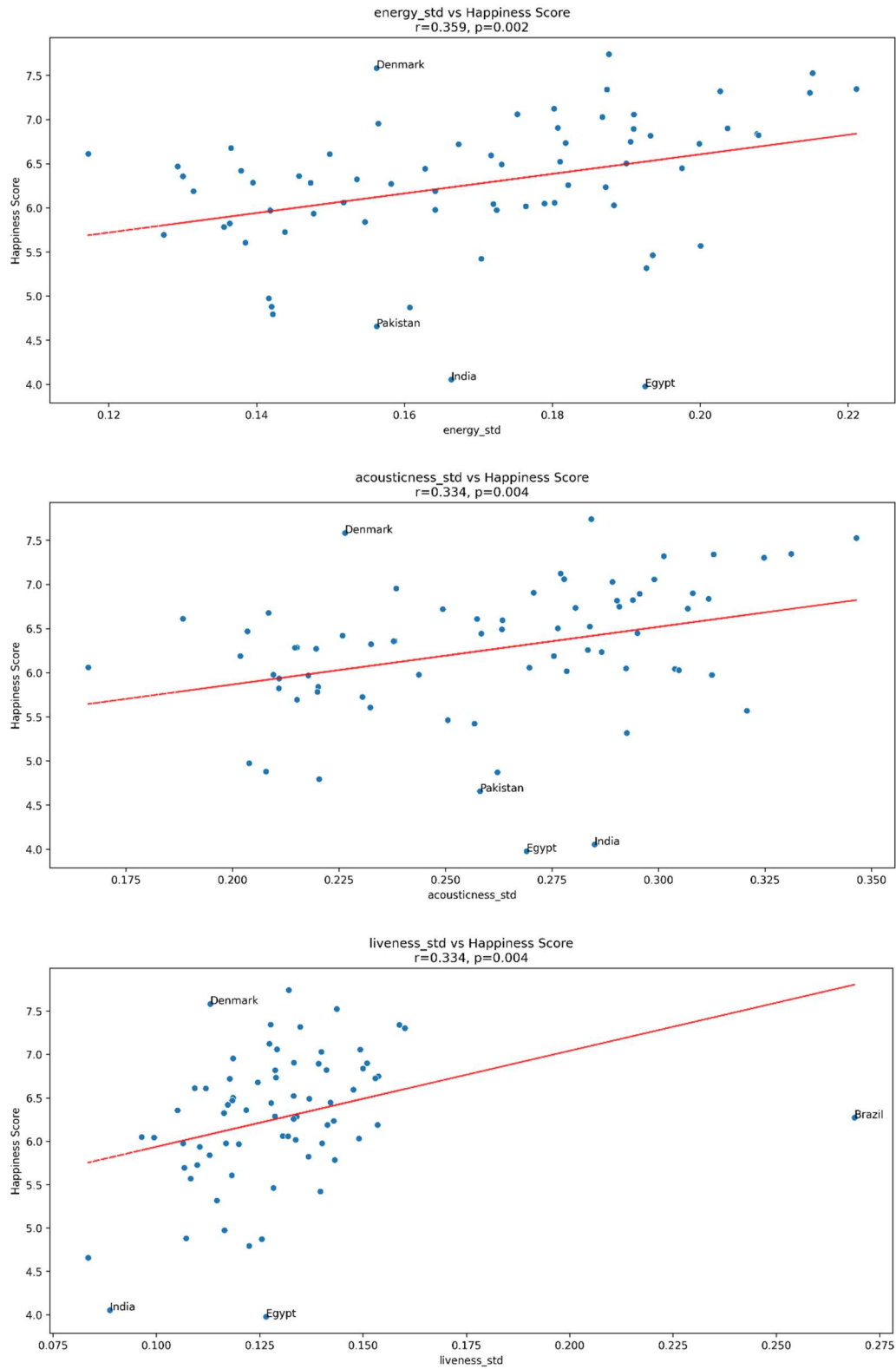
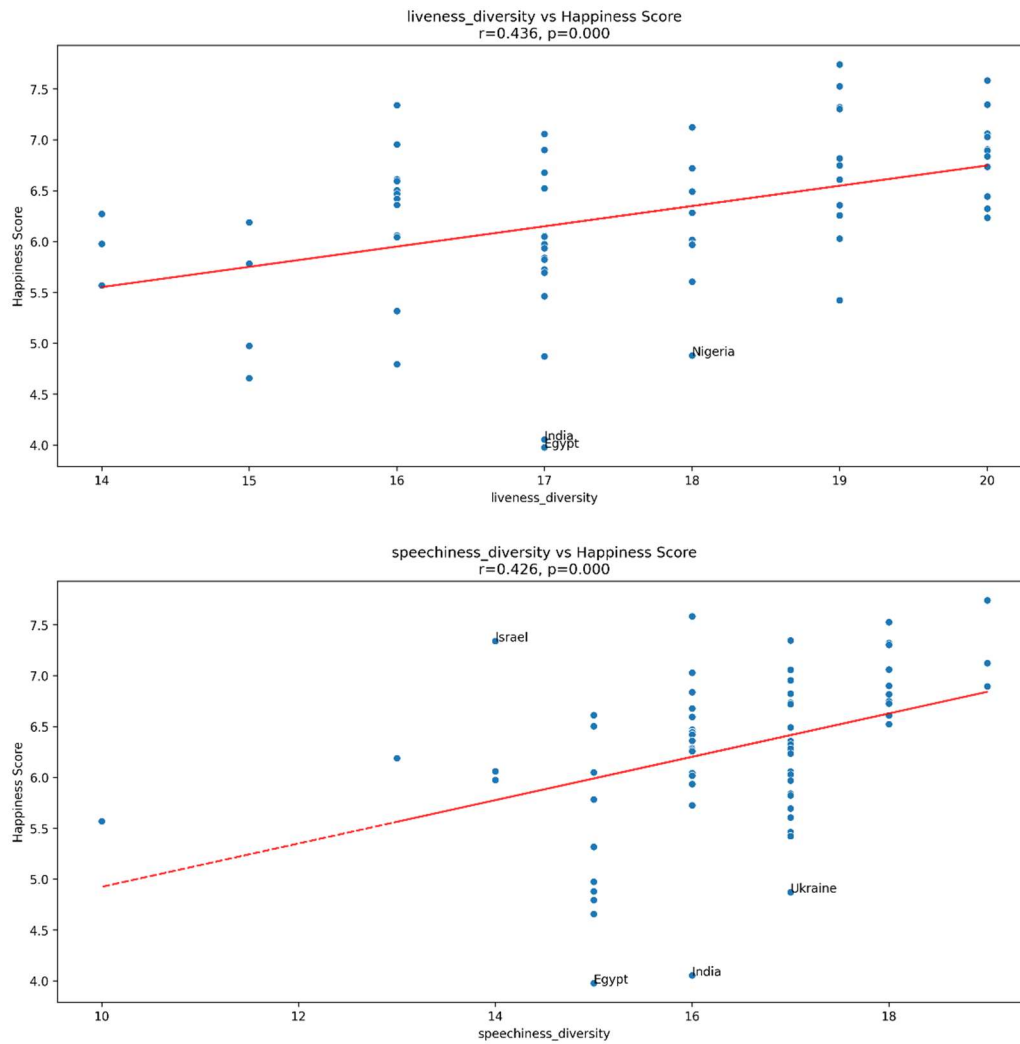Figure 4. Strong positive correlations exist between happiness and variability in energy, liveness, and acousticness.

Figure 5. Liveness diversity and speechiness diversity showed the strongest positive correlations with happiness.
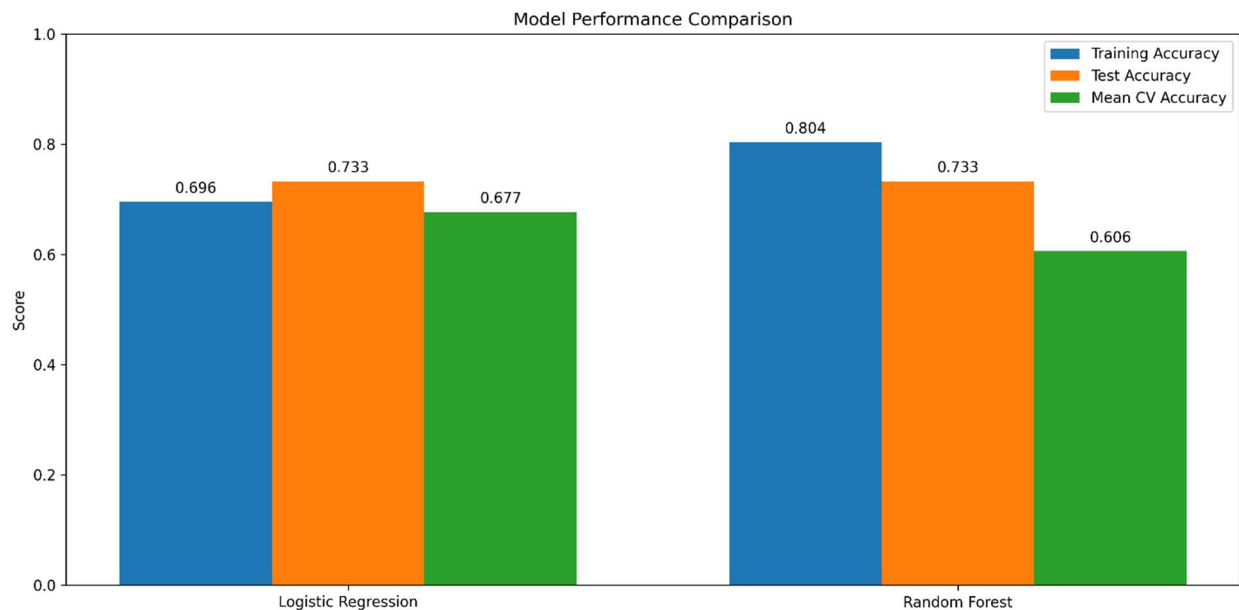
6 ALGORITHMS & MACHINE LEARNING

We tested two classification approaches:

1. **Random Forest**:

   - Test Accuracy: 73.3%
   - CV Accuracy: 62.4% (±31.5%)
   - Best Parameters: max_depth=2, min_samples_leaf=2

2. **Logistic Regression**:

   - Test Accuracy: 73.3%
   - CV Accuracy: 67.7% (±15.3%)
   - Best Parameters: C=0.001, balanced class weights

WINNER: Logistic Regression

While both models achieved the same test accuracy, Logistic Regression showed more stable cross-validation performance and better interpretability of feature relationships.

7 RESULTS

The analysis revealed several key patterns:

1. **Classification Performance**:

   - Both models achieved identical test accuracy (73.3%) in classifying countries as above or below median happiness.
   - Both models were better at identifying happier countries (88% precision) than less happy ones.
   - Logistic Regression showed more stable performance across cross-validation folds.

2. **Key Indicators**:

   - Musical diversity emerged as the strongest predictor. Countries with greater diversity in speech-like music elements tended to be happier. This could reflect greater cultural openness or diversity in musical expression.
   - The proportion of major key songs showed positive influence. Happier countries tended to listen to more songs in major keys.
   - Danceability showed an unexpected negative relationship. Surprisingly, countries with less danceable music tended to be happier.

Feature Coefficients in Logistic Regression
(+ indicates association with higher happiness)
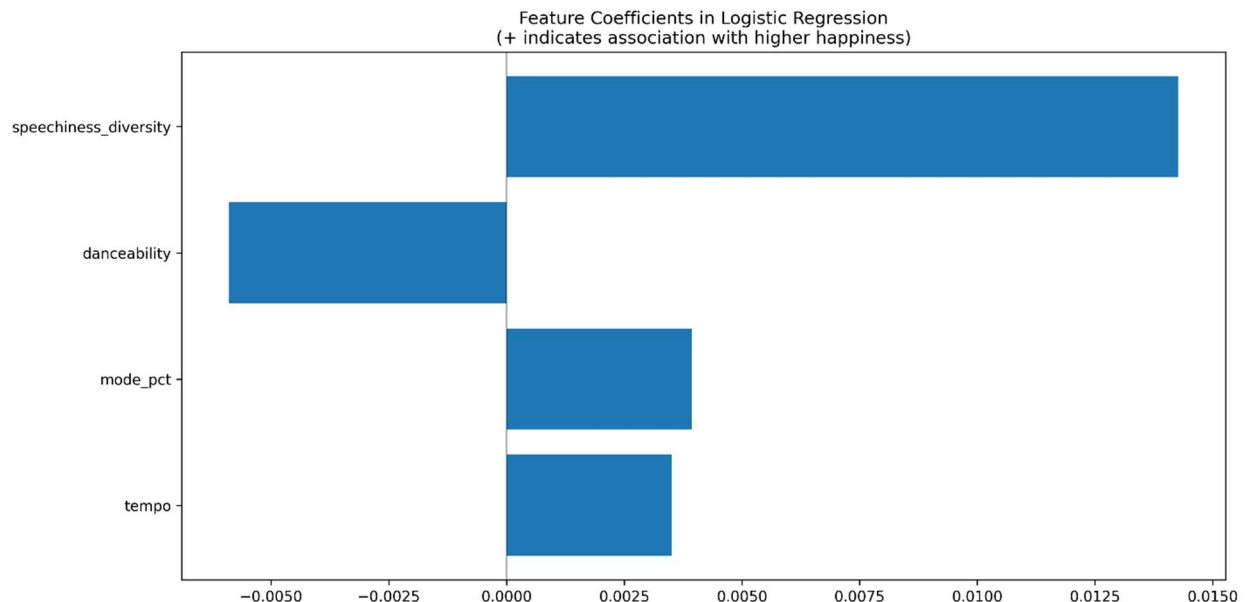
8 FUTURE IMPROVEMENTS

1. **Data Expansion**:

   - **Gather longitudinal data.** Future research would benefit from gathering longitudinal data; that is, collecting data across multiple years to track changes over time. To align with the World Happiness Report's data collection period, we limited our Spotify dataset to include only popular tracks from October through December 2023.
   - **Include genre-specific analysis.** Currently, the Spotify API can only access genre for musical artists; it cannot access genre for individual tracks. Therefore, our dataset lacked genre data for each track. In the future, it would be informative to perform genre-specific analysis to see whether musical genre has any salience in a country's happiness metrics.
   - **Include more countries in our analysis.** The United Nations recognizes 193 countries in the world. Of those 193 countries, 180 have access to Spotify. Our dataset was limited to only 73 countries; two of which, unfortunately, were not included in our World Happiness dataset. Future analysis would benefit from including the full breadth of countries representing Spotify's userbase.

2. **Methodological Improvements:**

   - **Consider multi-class classification.** Rather than simply dividing countries into two groups (above or below median happiness), we could divide them into multiple groups, such as happiness quartiles (happiest 25%, second happiest 25%, etc.).
   - **Incorporate cultural and economic controls.** To add depth to our analysis, we could include cultural dimensions (such as individualism vs. collectivism), religious or linguistic diversity metrics, education levels, and urban vs. rural population ratios. By controlling for these factors, we could better understand whether the relationship between music and happiness exists independently of these other influences.

References

Croom, A. M. (2012). Music, neuroscience, and the psychology of well-being: a précis. *Frontiers in Psychology*, *2*, 393. https://doi.org/10.3389/fpsyg.2011.00393

Helliwell, J. F., Layard, R., Sachs, J. D., Aknin, L. B., & Wang, S. (2024). *World Happiness Report 2024*. Sustainable Development Solutions Network.

Spotify. (2024). Spotify for Developers: Web API Reference. https://developer.spotify.com/documentation/web-api/reference/