SPOTIFY AND DECIBEL ANALYZER REPORT

January 2025

Purpose of the Project

This project explores the relationship between music listening habits and environmental sound intensity. By analyzing Spotify listening data alongside decibel levels captured by the Apple Health app, we aim to uncover patterns in music preferences and how they align with acoustic environments. The goal is to understand whether external sound intensity affects the frequency and choice of music.

Goals of the Project

The primary objective of the project is to analyze the relationship between:

1. **Listening Frequency**: How often specific artists are played on Spotify.
2. **Environmental Sound Levels**: Measured in decibels (dB) using Apple Health.

We aim to evaluate whether these factors are independent or show any significant dependency, leveraging statistical tests and visualization techniques.

Hypothesis

The analysis tests the following hypotheses:

1. **Null Hypothesis (H_0)**: Listening frequency and decibel levels are independent.
2. **Alternative Hypothesis (H_1)**: Listening frequency and decibel levels are dependent.

Two statistical tests, **Spearman Correlation** and **Chi-Square Test**, were employed to ensure the robustness of the results. Both tests yielded consistent outcomes, indicating no significant dependency between the two variables.

Methodology

Data Collection

1. **Spotify Data**:
   o Extracted streaming history, including artist names, listening timestamps, and playback duration.
2. **Apple Health Decibel Data**:
   o Collected environmental sound levels and synchronized them with Spotify listening timestamps.

Data Processing

1. Merged datasets based on timestamps to associate artists with the decibel levels during playback.
2. Filtered data to ensure only valid listening sessions were analyzed (e.g., removed zero-frequency entries).

Statistical Testing

1. **Spearman Correlation**:
   o Evaluated the monotonic relationship between decibel levels and listening frequency.
   o Result: Weak or no correlation was observed, failing to reject H_0.
2. **Chi-Square Test of Independence**:
   o Decibel levels and frequencies were binned into categories.
   o Result: No significant dependency, confirming the outcome of the Spearman test.

Visualization

- Scatter plots, bar charts, and regression lines were used to visualize trends between listening frequency and decibel levels.
- Charts for the top 10 loudest and quietest artists were created, prioritizing those with higher frequencies.

Findings

1. **Top Artists by Decibel Levels**:
   o Identified the 10 loudest and quietest artists based on average decibels.
   o Quietest artists were selected by prioritizing those with the highest frequency at the lowest decibel levels.
2. **Statistical Results**:
   o Both Spearman Correlation and Chi-Square Test confirmed no dependency between decibel levels and listening frequency.
3. **Visual Patterns**:
   o Regression lines showed weak relationships between variables, supporting the hypothesis of independence.

Limitations

1. **Dataset Size**:
   - Relies on personal listening and Apple Health data, which may limit generalizability.
2. **Binning Choices**:
   - Results may vary with different binning strategies for decibel and frequency ranges.

Conclusion

The project successfully integrated Spotify and Apple Health data to analyze listening habits in varying sound environments. Despite the rigorous statistical analysis, no significant dependency was observed between listening frequency and decibel levels, reinforcing the null hypothesis. The findings provide a framework for further exploration into music and environmental acoustics.

Acknowledgments