Project Proposal: Analyzing "Most Streamed Spotify Songs 2023" Dataset

Introduction: This project involves analyzing the "Most Streamed Spotify Songs 2023" dataset obtained from Kaggle. The dataset includes key values related to the popularity of songs on the streaming platform, encompassing metrics such as streams, beats per minute (bpm), danceability, liveness, and valence.

Dataset Definition: The dataset comprises continuous numeric data, including:

* Streams: The number of times a song has been streamed.
* BPM: Beats per minute, indicating the tempo of the song.
* Danceability: A metric representing the suitability of a song for dancing.
* Liveness: The likelihood that a song is performed live.
* Valence: The musical positiveness of a song.

Objectives: The primary goal of this project is to identify the key factors that significantly impact a song's popularity on streaming services. Understanding these factors can influence a company's decision to sign a new artist or assess the viability of sponsorships or collaborations between musicians and businesses.

Hypotheses:

1. High Danceability Hypothesis:
   * Null Hypothesis (H0): There is no significant difference in popularity between songs with high danceability and those with low danceability.
   * Alternative Hypothesis (H1): Songs with high danceability are more popular than songs with low danceability.
2. Liveness Impact Hypothesis:
   * Null Hypothesis (H0): The level of liveness does not significantly affect the popularity of a song.
   * Alternative Hypothesis (H1): Songs with low liveness are more popular than songs with high liveness.

Stakeholders: This study is valuable for various stakeholders, including:

* Music Companies: Insights can aid in decision-making regarding artist signings and marketing strategies.
* Musicians: Understanding key factors for popularity can guide their creative processes.
* Businesses: Assessment of collaboration viability between musicians and businesses for sponsorships.

Methodology: The analysis will involve exploratory data analysis of the dataset, focusing on the identified variables. Two-sample t-tests will be conducted to assess the statistical significance of danceability and liveness in relation to song popularity. The entire analysis will be performed using Python, and the findings will be presented through a PowerPoint presentation.

Conclusion: By exploring and testing these hypotheses, this project aims to contribute valuable insights into the factors influencing song popularity on streaming services, providing actionable information for stakeholders in the music industry.