**How to Make a Hit on Spotify: Insights from Data Analysis**

In today’s digital music landscape, understanding what makes a song popular on streaming platforms like Spotify is crucial for artists, producers, and record labels. Our analysis of **170,000 songs released between 1921 and 2020** uncovers the key factors that drive success on Spotify.

### **What Defines Popularity on Spotify?**

Spotify's **popularity index (0-100)** measures a song’s success based on several factors, calculating based on:

* **Streams**
* **Engagement**
* **Recency**
* **Playlist Inclusion**

For our study, we defined a **hit song** as having a popularity score of **50 or higher**—roughly the **top 25% of all songs** in the dataset.

### **The Evolution of Popularity Over Time**

One key trend in our analysis was the steady rise in song popularity over recent years. This can be attributed to **Spotify’s launch in the U.S. in 2011**, which transformed the way people consume music. However, older songs tend to have lower popularity scores, suggesting that recency is a significant driver of success.

### **Which Features Matter Most for Success?**

Our study examined **10 key audio features**, such as **energy, danceability, loudness, speechiness, and valence**, to determine their impact on popularity.

* **Energy**: Higher-energy songs tend to perform better.
* **Loudness**: Loudness correlates strongly with energy (0.78), making it an important predictor.
* **Danceability**: While not as influential as energy, it still plays a role.
* **Acousticness and Instrumentalness (correlation: 0.33)**: Popular songs are more likely to have higher energy and fewer acoustic elements.

### **Building a Predictive Model for Popularity**

To predict a song’s popularity, we tested several machine learning models:

1. **Linear Regression (R² = 0.759, MAE = 7.98)**

MODEL 1: PREDICTING POPULARITY BASED ON SONG FEATURES

This model, which relies on characteristics like acousticness, tempo, and speechiness, provides a reasonable level of predictive power. The **R-squared value of 0.759** indicates that approximately **75.9% of the variation in song popularity** can be explained by these features, showing a strong relationship between a song’s attributes and its success. The **Mean Absolute Error (MAE) of 7.98** suggests that, on average, the model’s predictions deviate by about 8 popularity points. This level of accuracy makes it a useful tool for understanding the impact of song features on popularity, though there is still room for refinement.

MODEL 2: PREDICTING POPULARITY BASED ON GENRE, ARTIRTS, RELEASED YEAR, AND NUMBER OF SONGS

To further refine predictions, we built a **genre estimation,** assigning a genre to each song based on its audio features and release year.

#### **Example: Identifying EDM/House Songs**

* Released **after 1980**
* **Tempo** (BPM) falls within **120-160**
* **Danceability** is higher than **0.65 (65%)**

The **Genre + Artists + Num Songs** model performs significantly better. This could partially explain that factors like who the artist is and what genre the song belongs to have a major influence on popularity. Listeners often gravitate toward familiar artists and genres they already love, making these aspects more impactful than individual song characteristics.

While features like **acousticness, tempo, and speechiness** contribute to the overall listening experience, they don’t dictate a song’s success as much as an artist’s reputation or the genre’s

From a statistical perspective, using these features to the model increased its **R-squared value to 0.805** and lowered the **Mean Absolute Error (MAE) to 6.84**, making it a more reliable predictor than models based solely on song features. This means the model explains over **80% of the variation in song popularity**, demonstrating that an artist’s established presence and genre affiliation play a crucial role in determining success.

1. **Random Forest Regressor (R² = 0.809, MAE = 6.74)**

provided a more accurate prediction of song popularity by capturing complex, non-linear relationships between audio features. Unlike traditional linear models, this approach leverages multiple decision trees, each analyzing different subsets of data, and averages their predictions to reduce errors and overfitting. With an **R-squared value of 0.809**, the model explains **80.9% of the variation in song popularity**, making it a strong predictor. The **Mean Absolute Error (MAE) of 6.74** indicates improved precision, with lower deviation in predictions compared to the linear regression model. The most influential factors in determining popularity included **acousticness, speechiness, loudness, and duration**, suggesting that both the texture of a song and its structural elements play a significant role in listener engagement.

1. **Logistic Regression**

The **Logistic Regression model** achieved approximately **78% accuracy** in predicting whether a song would be popular or not. This classification model identified **loudness, valence, and danceability** as the most influential factors in determining success. Loudness reflects the song’s intensity, valence measures its emotional positivity, and danceability indicates how rhythmically engaging it is. Together, these features help define a song’s overall appeal, making this model useful for distinguishing between hits and less popular tracks.

### **Key Takeaways for Artists and Record Labels**

* **Emotion and energy matter**: Listeners gravitate toward high-energy, emotionally powerful songs.
* **New music has an edge**: Fresh tracks have a higher chance of success due to Spotify’s algorithmic preferences.
* **Genre plays a role**: Pop remains dominant, while older genres struggle to maintain relevance in streaming culture.

### **Final Thoughts**

The music industry is increasingly data-driven. By leveraging audio features, machine learning, and historical trends, artists and record labels can optimize their strategies for making a hit song. Whether you’re an aspiring musician or an executive scouting talent, understanding these data-backed insights can help you make better decisions in the world of streaming.