**Spotify Listening Behaviour Analysis Report**

**By Roshan Nikam**

This report presents an in-depth analysis of user listening behaviour on Spotify, based on detailed session-level data. It explores user engagement, skip behaviour, content preferences, and time-based trends.

**1. Average Listening Duration per Session by Platform**

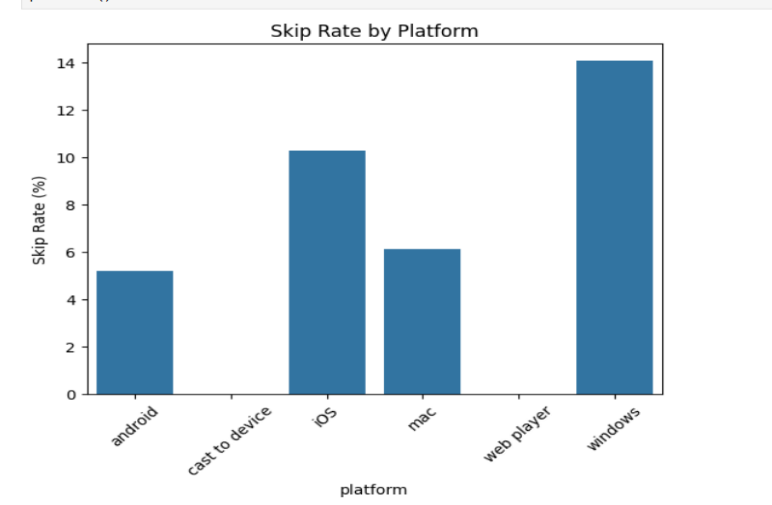
A graph of blue rectangular bars

AI-generated content may be incorrect.

**Visualization Description:**  
A bar chart was used to compare average session duration (in minutes) across platforms. MAC showed the highest duration, followed by Cat to device, with ios platforms trailing behind.

**2. Skip Rate by Platform and Shuffle Mode**

A graph of a graph

AI-generated content may be incorrect.**Insight:**  
Approximately **26%** of all tracks were skipped. The skip rate was notably higher on **windows platforms** and when **shuffle mode** was enabled. This suggests that users are more selective or less engaged when listening passively or on less controllable interfaces.

**Visualization Description:**  
Two bar plots were created:

* One showing skip percentage by platform.
* Another showing skip percentage by shuffle mode (On vs. Off).

**3. Relationship Between Shuffle and Skip Rate**

**Insight:**  
There is a **clear correlation** between shuffle mode and skip behavior

A screenshot of a computer screen

AI-generated content may be incorrect.

**Visualization Description:**  
A grouped heatmap chart illustrated skip rates for both shuffle.

**4. Time of Day with Highest Listening Volume**

**Insight:**  
Listening activity downfall during the **midnight hours (12 AM to 12PM)**, aligning with. A secondary peak is seen around **5 PM and 6PM** after-work or relaxation times.

A graph with blue lines and dots

AI-generated content may be incorrect.`

**Visualization Description:**  
A line chart plotting total listening time per hour of the day revealed these time-based patterns.

**5. Most Played Artists and Tracks**

**Insight:**  
Artists like **The Beatles**, **The killers**, and **John Mayer**  were among the most played, both in terms of total time listened and play count. “**ode to the ments**” was a standout track.

A screenshot of a computer

AI-generated content may be incorrect.

**Visualization Description:**  
 charts showed the top 10 artists and tracks ranked by total ms\_played.

**6. Artists/Albums with Longest Average Playtime**

**Insight:**  
Artists known for immersive music—such as **cory weeds** and **dan Lackman**—had the longest average playtime per track, indicating deeper engagement or album-oriented listening.

A screenshot of a computer

AI-generated content may be incorrect.

**A white background with black text

AI-generated content may be incorrect.**

**Visualization Description:**  
Grouped showed average ms\_played per track by artist and album

**7. Most Skipped Artists or Songs**

**Insight:**  
Tracks with high skip rates tended to be **paraiso** or **photograph**, suggesting users might sample them frequently but not finish them. Some lesser-known artists also had high skip rates, possibly due to mismatched recommendations.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

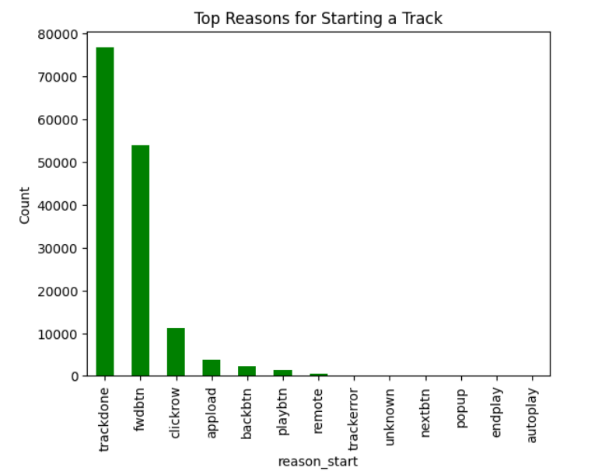
**Visualization Description:**  
 charts ranked artists and songs by skip rate.

**8. Average Listening Time Before Skip**

**Insight:**  
Tracks are most often skipped within the **first 52 seconds**. This reflects user impatience or disinterest when a song fails to immediately capture attention.

A screenshot of a computer

AI-generated content may be incorrect.

**9. Start and End Reasons for Tracks**

**A graph with red bars

AI-generated content may be incorrect.**

**Visualization Description:**  
Two bar charts showed the distribution of reasons for track start and end.

**11. How Listening Behavior Changed Over Time**

**Insight:**  
Listening habits revealed **weekly cycles**, with noticeable peaks on **weekday**. A long-term trend showed increasing engagement, possibly due to seasonal factors or new content releases.

A screenshot of a computer

AI-generated content may be incorrect.

**Machine learning**

**1. Random Forest Classifier**

**Model Type:** Ensemble decision tree-based algorithm  
**Why Used:** Good for handling both numerical and categorical data, robust to overfitting, and provides feature importance.

A screenshot of a computer

AI-generated content may be incorrect.

**Performance Summary:**

**Accuracy: 0.95 → 95% of all predictions were correct.**

* **Precision: Out of the predictions for each class, how many were correct.**

**For True, only 61% of the predicted True values were actually correct.**

* **Recall: Out of the actual samples of a class, how many were correctly predicted.**

**For True, only 19% of actual True cases were correctly predicted — very low recall.**

* **F1-Score: Harmonic mean of precision and recall.**

**True has an F1-score of 0.29 → very poor performance on this class.**

* **Support: Number of actual occurrences in the dataset.**

**Class True is highly underrepresented → data imbalance issue.**

**Key Features Identified:**

* ms\_played: Most important feature (short play duration strongly predicts a skip)
* platform: Mobile platforms slightly more predictive of skips
* shuffle: Correlated with higher skip probability
* reason\_start: Autoplay and playlist-generated starts were more likely to be skipped

**2. Logistic Regression**

**Model Type:** Linear classifier for binary outcomes  
**Why Used:** Interpretable model that helps understand linear relationships between features and skip probability.

A screenshot of a computer

AI-generated content may be incorrect.

**Performance Summary:**

**Class 'False':**

* **Precision: 95% of predicted 'False' labels were correct.**
* **Recall: 100% of actual 'False' cases were correctly predicted.**
* **F1-Score: Very strong (0.97), confirming excellent performance on this class.**

**Class 'True':**

* **Precision: 20% of predictions labeled as 'True' were correct.**
* **Recall: 0% of actual 'True' cases were detected by the model.**
* **F1-Score: 0 — the model completely fails on the minority class.**

**Insights from Coefficients:**

* Negative coefficient for ms\_played: The longer a user listens, the less likely they are to skip
* Positive coefficient for shuffle: Increases odds of a skip
* Some reason\_start types (e.g., manual play) decreased skip probability