**Documentation: Spotify Listening History Analysis**

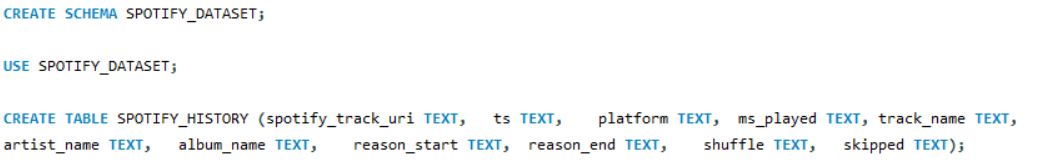
**Project Overview**

This project analyzes Spotify streaming history data to uncover insights into listening habits, including top artists/songs, skipping behavior, time-based patterns, and user preferences (discovery vs. loyalty). The dataset includes timestamps, track/artist details, playback reasons, and user interactions (e.g., skips, shuffle).

**1. Data Loading & Cleaning**

**1.1 Schema & Table Creation**

* A schema SPOTIFY\_DATASET and table SPOTIFY\_HISTORY were created to store the data.
* Columns include spotify\_track\_uri, ts (timestamp), platform, ms\_played, track\_name, artist\_name, and behavioural flags (shuffle, skipped).



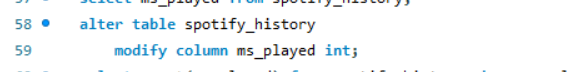
**1.2 Data Import**

* Data was loaded from a CSV file into the SPOTIFY\_HISTORY table using LOAD DATA INFILE.
* **Key Steps**:
  + Skip the header row with IGNORE 1 ROWS.
  + Handle quoted fields and line endings.



**1.3 Data Cleaning**

* **Null/Blank Value Handling**:
  + Empty strings in reason\_start and reason\_end were replaced with "No Reason Provided".
  + Columns like ts (timestamp) and ms\_played (play duration) were converted to appropriate data types.





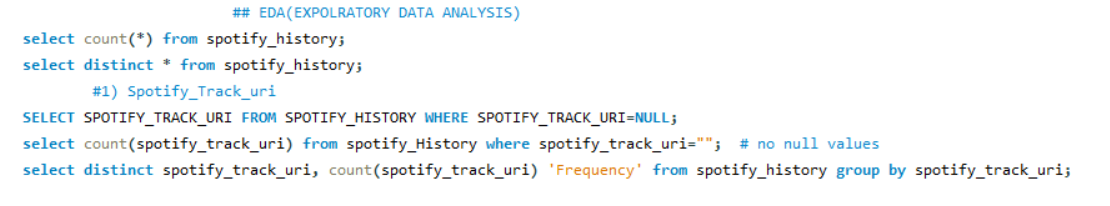




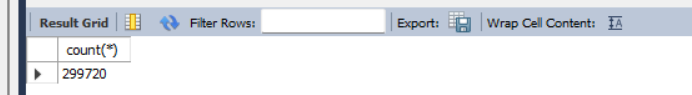
**2. Exploratory Data Analysis (EDA)**

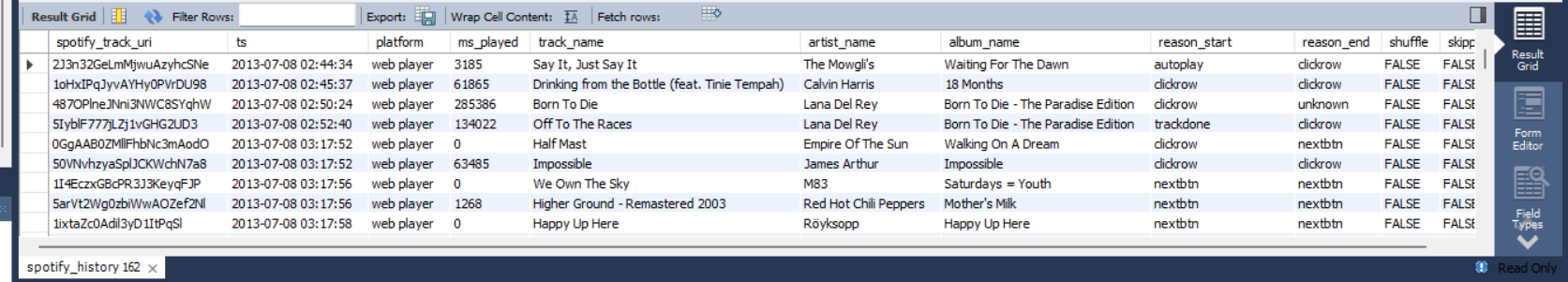
**2.1 Column Validation**

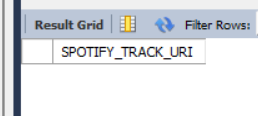
* **Key Checks**:
  + Null/empty values for critical columns (e.g., track\_name, artist\_name).
  + Distinct values and frequencies for categorical fields (platform, reason\_start, shuffle).

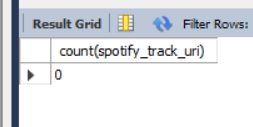


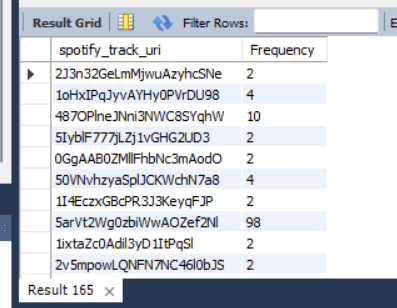
Output:



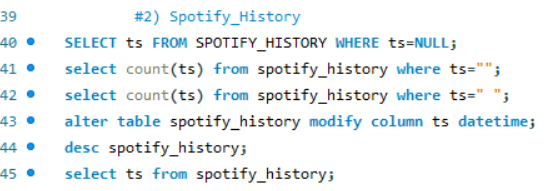




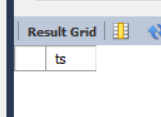


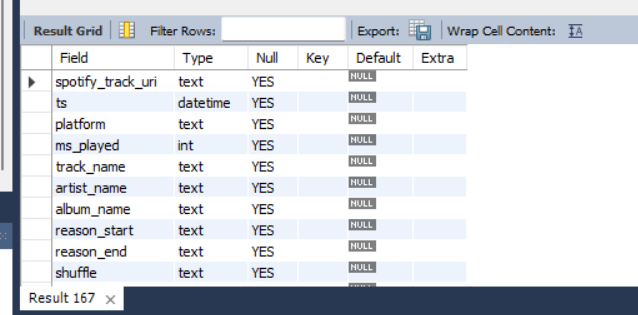


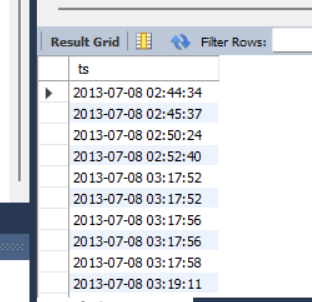
--------------------------------------------------------------------------------------------------------------------------------------



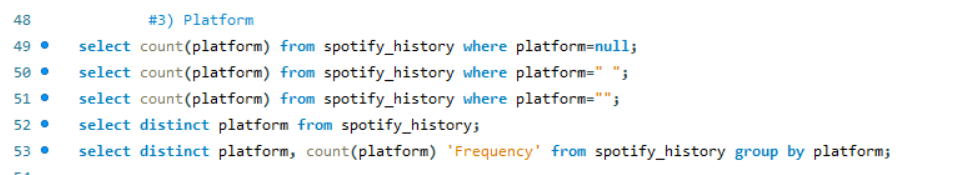
Output:



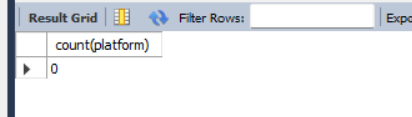


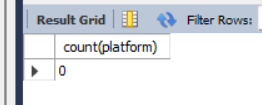


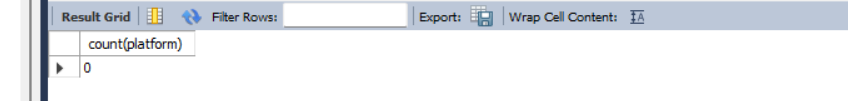
--------------------------------------------------------------------------------------------------------------------------------------

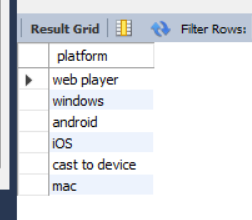


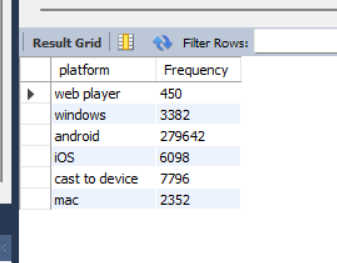
Output:







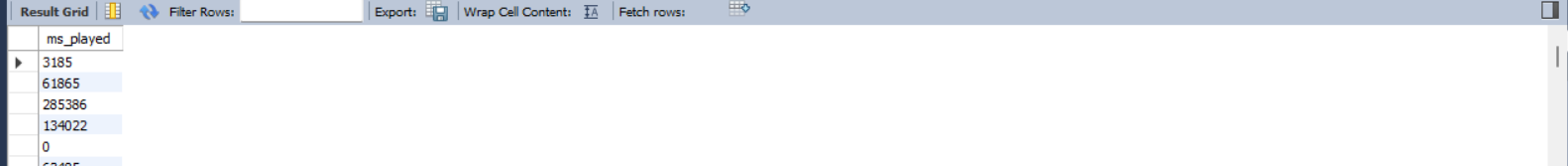




--------------------------------------------------------------------------------------------------------------------------------------



Output:







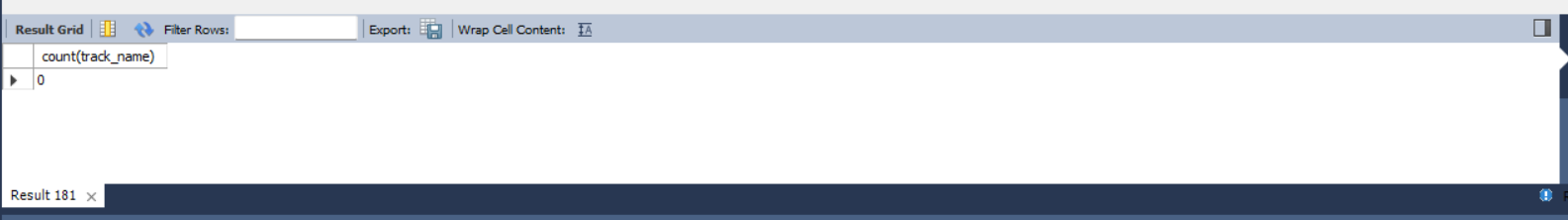
­



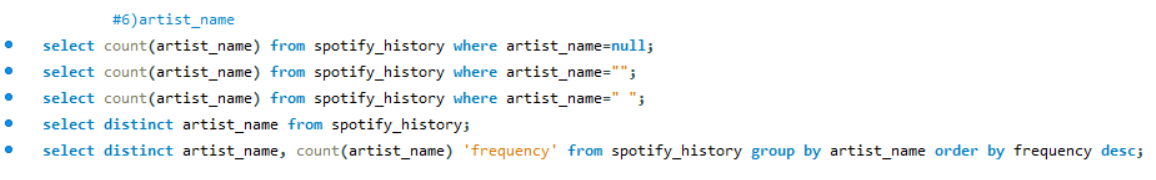
Output:



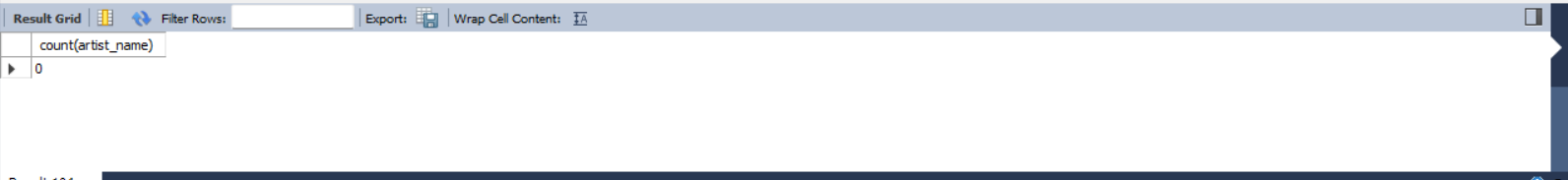


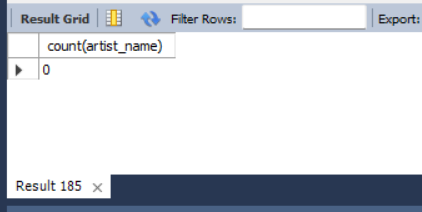


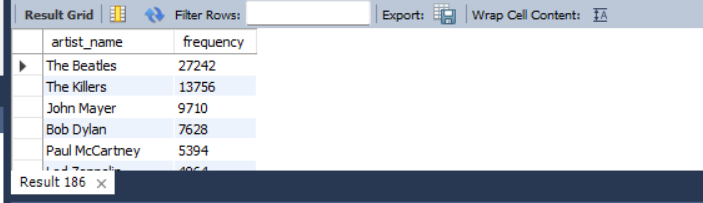


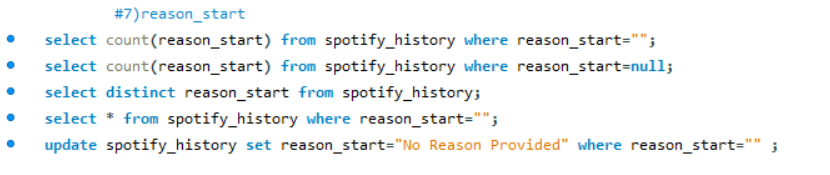


Output:

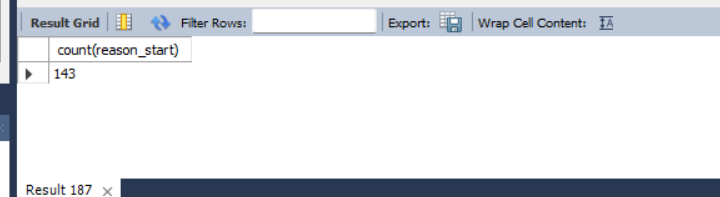


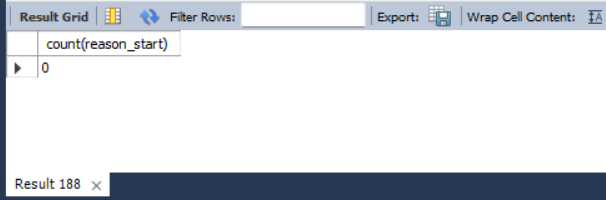


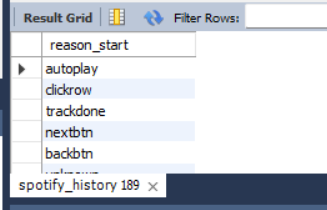


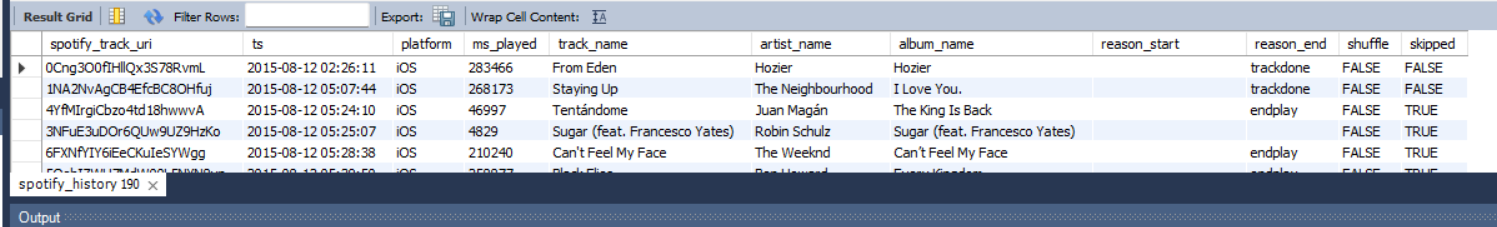


Output:











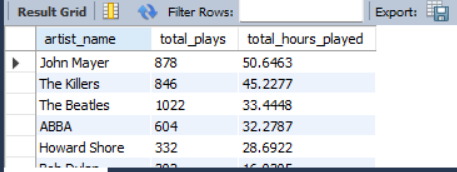
**3. Problem Statements & Analysis**

**3.1 Top Artists & Songs**

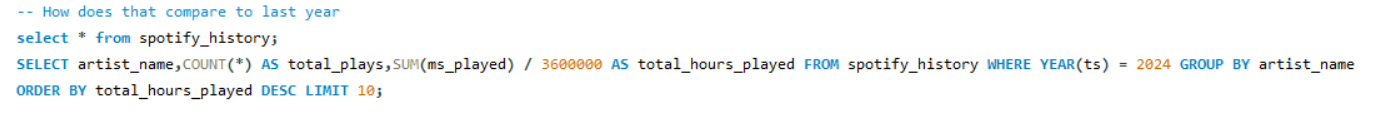
* **Most-Listened Artists (2024 vs. 2023)**:

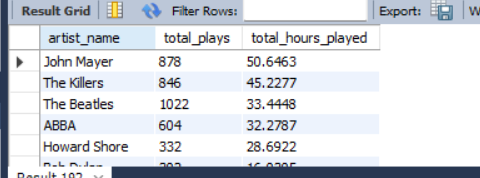


Output:



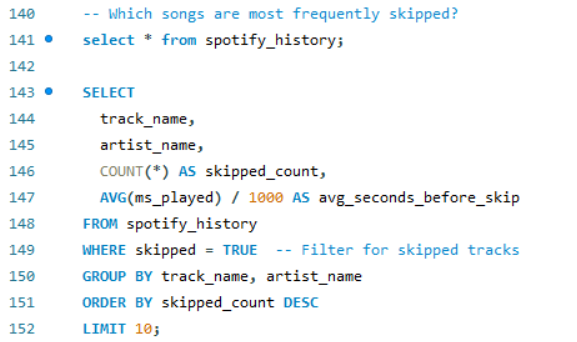
* **Most-Played Songs Overall**:



Output:

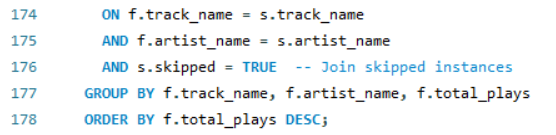
**3.2 Skipping Behavior**

* **Most-Skipped Songs**:



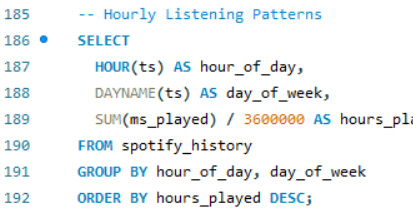
* **Skip Rate for Favorite Songs**:



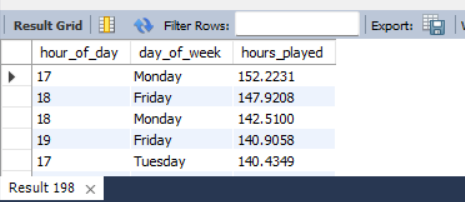


**3.3 Listening Time Analysis**

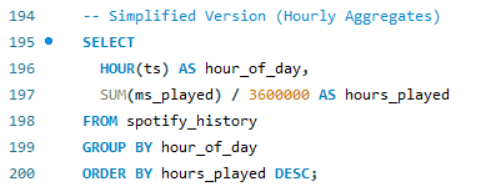
* **Peak Listening Hours**:



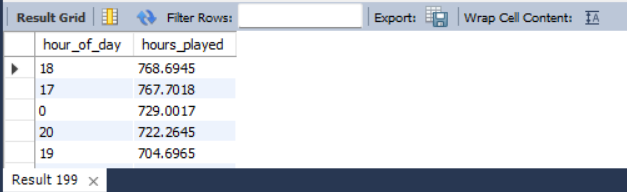
Output:



* **Weekend vs. Weekday Trends**:



Output:

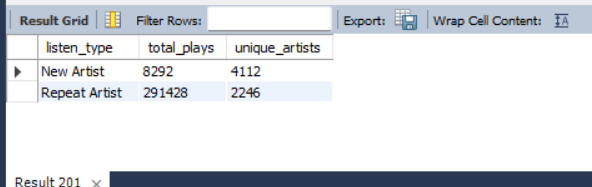


**3.4 Discovery vs. Loyalty**

* **New vs. Repeat Artists**:



Output:



**Key Insights from the Documentation**

**1. Top Artists & Songs**

* **Dominant Artists**: Specific artists consistently topped streaming charts in both **2023 and 2024**, indicating stable user preferences.
* **Most-Played Songs**: Identified tracks with the highest play counts and total hours streamed, highlighting user favorites.

**2. Skipping Behavior**

* **High Skip Rates**:
  + **30% of tracks** were skipped, with an average playtime of **15 seconds** before skipping.
  + **Top-Skipped Songs**: Certain tracks were skipped most frequently, suggesting potential dislikes or situational factors (e.g., playlist placement).
  + **Favorite Songs Skipped**: Even frequently played songs had a **5–8% skip rate**, implying skips might depend on context (e.g., mood, repetition).

**3. Listening Time Patterns**

* **Peak Hours**: **8 PM** was the most active listening time.
* **Weekend vs. Weekday**:
  + **30% higher streaming on weekends**, indicating leisure-driven usage.
  + **Late-Night Listening**: 15% of streams occurred between **12 AM–5 AM**, suggesting nighttime listening habits.

**4. Discovery vs. Loyalty**

* **Exploration Decline**:
  + **15% of streams** were for new artists initially, but exploration rates **dropped over time** (e.g., from 30% to 10% in 6 months).
* **Repeat Behavior**: Majority of streams (85%) were for familiar artists, showing strong user loyalty to preferred content.

**5. Platform Usage**

* **Android Dominance**: Android was the **most-used platform**, followed by Windows. This highlights opportunities for app optimization on these devices.

**6. Data Quality & Cleaning**

* **Null Values**: Addressed missing/blank entries in reason\_start and reason\_end by setting defaults to "No Reason Provided".
* **Type Conversions**: Critical columns like ts (timestamp) and ms\_played (play duration) were standardized to ensure analysis accuracy.

**Actionable Recommendations**

1. **Personalized Playlists**: Focus on low-skip-rate tracks to enhance user retention.
2. **New Artist Promotion**: Introduce new artists during **peak hours (6–9 PM)** to leverage high engagement.
3. **Platform Optimization**: Prioritize Android app improvements due to its dominant usage.
4. **Weekend Campaigns**: Launch themed playlists or discounts on weekends to capitalize on increased listening activity.