Spotify Data Field Dictionary & Technical Explanation

Project Title: Spotify Listening Insights – Data Dictionary & Domain Documentation

1. Introduction

This document provides a complete and professional **data dictionary** for the Spotify dataset used in the Listening Insights & Engagement Analytics Dashboard. It defines each field, its purpose, format, and analytical relevance. This document ensures clarity during data modeling, transformation, and dashboard development.

---

2. Data Dictionary Overview

The Spotify dataset contains multiple fields representing track-level listening behavior, playback metadata, and user interaction patterns. Each attribute described below plays a crucial role in behavioral analytics, listening trend identification, and engagement measurement.

---

3. Field-Level Descriptions

1. spotify_track_uri

Description: A globally unique identifier for each track in Spotify's catalog.
Format: `spotify:track:<base-62 string>`
Example: `spotify:track:3n3Ppam7vgaVa1iaRUc9Lp`
Purpose: - Links the event log to official Spotify metadata. - Used for cross-referencing track details such as album, duration, popularity, etc. - Essential for joining external metadata sources.

---

2. ts (Timestamp)

Description: Exact UTC timestamp marking when the track stopped playing.
Format: ISO 8601 (e.g., `2024-02-07T14:30:45Z`)
Purpose: - Helps analyze listening patterns across time. - Enables calculation of session times and daily/weekly engagement patterns. - Useful in building heatmaps for listening hours.

---

3. platform

Description: The device or platform used for streaming the track.
Possible Values: - desktop (Windows/Mac app) - mobile (iOS/Android app) - web (Spotify Web Player) - smart_speaker (Alexa, Google Home, etc.)

Purpose: - Helps understand listening device preference. - Can be used to segment engagement (mobile users vs desktop users). - Useful for device-based performance optimization.

---

4. ms_played

Description: Total duration (in milliseconds) for which the track was listened before stopping or skipping.
Format: Integer (e.g., 215000 ms = 3 minutes 35 seconds)
Purpose: - Core measurement for user engagement. - Helps differentiate between completed plays vs partial plays. - Used to compute total listening time and revenue contribution (Spotify relies on duration-based payouts).

---

5. track_name

Description: The title of the song that was played.
Example: "Shape of You"
Purpose: - Essential for identifying most played tracks. - Used for generating Top Tracks charts. - Supports label-based filtering in dashboards.

---

6. artist_name

Description: The name of the performing artist.
Example: "Ed Sheeran"
Purpose: - Central for Artist Engagement Analysis. - Helps calculate unique artists played per year. - Used to generate artist ranking and YoY comparisons.

---

7. album_name

Description: The name of the album associated with the track.
Example: "÷ (Divide)"
Purpose: - Critical for Album Insights module. - Helps identify user preferences at album level. - Used to compute album-level listening frequency.

---

8. reason_start

Description: Describes how or why the track started playing.
Possible Values: - trackdone (previous song finished) - clickrow (user manually selected) - backbtn (user pressed previous) - fwdbtn (user pressed next) - playbtn (user pressed play) - autoplay (Spotify auto-queued the next song)

Purpose: - Helps understand listening intent vs algorithm-driven plays. - Useful for analyzing user-initiated vs auto-play behavior. - Supports engagement and interaction pattern analysis.

---

9. reason_end

Description: Describes why the track stopped playing.
Possible Values: - trackdone (track finished normally) - endplay (user paused/stopped playback) - fwdbtn (user skipped forward) - backbtn (user went to previous track) - logout (session ended/disconnected)

Purpose: - Crucial for analyzing skip rates. - Helps measure user retention per track. - Provides insights into track performance and drop-off behavior.

---

10. shuffle

Description: Indicates whether shuffle mode was enabled.
Values: TRUE / FALSE
Purpose: - Helps understand how often users rely on shuffle mode. - Useful for analyzing listening randomness. - Can be used to build shuffle vs non-shuffle behavior comparisons.

---

11. skipped

Description: Indicates whether the user skipped the song before completion.
Values: TRUE / FALSE
Purpose: - Key metric for identifying track engagement. - High skip rates can indicate low user interest. - Useful for building skip-rate dashboards and track performance KPIs.

---

4. Summary

This data dictionary forms the foundation for accurate data modeling and dashboard creation in the Spotify Listening Insights project. It ensures a clear understanding of data attributes, enabling efficient ETL processing, visualisation design, and analytical interpretation.