Data Strategy & Requirements Document: Spotify Analysis

BI Analyst: Tanzil Ali Rizvi
Client/Sponsor: Spotify Analytics Team

Purpose:

Translate stakeholder needs into data requirements for a Spotify performance analytics dashboard that delivers insights into streaming trends, song performance, and artist impact.

Data Sources:

- Spotify Web API (Primary): Tracks, albums, artists, genres, markets, and popularity data.
- Internal performance logs: Historical stream counts and playlist metrics.
- External sources: Billboard or third-party charts for comparative benchmarking.

Required Data Fields:

- Track: track_name, track_id, artist_id, album_id, duration_ms, popularity, is_explicit
- Artist: artist_name, followers, genres
- Album: album_name, album_type, release_date, total_tracks
- Stream Metrics: region, total_streams, date_added, chart_position

Data Cleaning & Transformation Rules:

- Normalize text (lowercase names, remove special characters)
- Convert duration_ms -> minutes (rounded to 2 decimals)
- Categorize album_type as "Single" or "Album"
- Remove duplicates by track_id and artist_id
- Filter out records missing popularity or stream counts
- Create Separate columns for Dates (dd, mm, yyyy) and other tasks like converting to the required data types (eg: Date from Text to Date Datatype) by using formulas in Google Sheets

Metric Definitions:

- Total Streams: SUM(stream_count) grouped by artist/track
- Average Popularity: MEAN(popularity)
- Explicit Song %: COUNT(explicit="Yes") / COUNT(all songs)
- Artist Reach: DISTINCT count of artists per region

- Stream Growth %: (Current Week Previous Week) / Previous Week * 100

Success Criteria:

- Dashboard refreshes automatically from live Spotify data
- All KPIs defined consistently with business logic
- Clear, interactive visuals for non-technical stakeholders

Compliance & Privacy:

- No personally identifiable listener data used.
- Data collected solely from Spotify's public APIs and aggregated performance datasets.