Data Lifecycle Scenario: Music Streaming Analytics

You are given a short scenario to analyze.

Step 1: Individual work Draw the lifecycle diagram, indicating:

- the sources of data,
- potential issues with quality, trustworthiness, or usability,
- the expected type(s) of data (structured / semi-structured / unstructured)

Step 2: Group Discussion: Compare your individual diagrams and notes with group members. Reflect on each stage of the lifecycle with your team and answer the guiding questions provided, completing the **lifecycle-step boxes** together. Prepare a brief summary (keywords or 1–2 sentences) to present.

Time guideline: 5 minutes individually, 10 minutes in groups.

Scenario Description: Music Streaming Analytics

A music streaming platform collects detailed user interactions, including song plays, skips, likes, playlist edits, and search queries. Users' demographic information, device types, and listening contexts are also recorded. The platform aims to understand listening habits, improve personalized recommendations, create curated playlists, and tailor promotions. Data comes from multiple sources and formats and must be cleaned, integrated, and analyzed efficiently. Real-time trend detection is also valuable for highlighting emerging artists or popular songs.

Scenario Analysis (individual work): Draw the lifecycle diagram, indicating:

- Source(s) of the data (human/machine, internal/external)
- Data type (structured / semi-structured / unstructured)
- Any obvious quality, reliability, or trustworthiness issues

Lifecycle diagram

Data Lifecycle Reflection (group work): For each stage, reflect on what it could involve in this scenario. Consider:

- How can duplicate or incomplete logs be handled?
- What privacy considerations exist when combining demographics and listening habits?
- How can analysis improve personalized playlists or recommendations?
- Could trends be detected in near real-time, and how?

Collect

Clean / Preprocess

Store

Analyze

Communicate / Report