Transforming Tabular Data to Natural Language

by Xyrel De Mesa

Most Streamed Spotify Songs 2023

Data Card | Code (175) | Discussion (27) | Suggestions (0)

track_name Name of the song	artist(s)_name Name of the artist(s) of the song	# artist_count Number of artists contributing to the song	# released_year Year when the song was released
943 unique values	Taylor Swift 4% The Weeknd 2% Other (897) 94%	 1 8	 1930 2023
Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	2	2023
LALA	Myke Towers	1	2023
vampire	Olivia Rodrigo	1	2023
Cruel Summer	Taylor Swift	1	2019
WHERE SHE GOES	Bad Bunny	1	2023
Sprinter	Dave, Central Cee	2	2023
Ella Baila Sola	Eslabon Armado, Peso Pluma	2	2023
Columbia	Quevedo	1	2023
fukumean	Gunna	1	2023
La Bebe - Remix	Peso Pluma, Yng	2	2023

spotify_streams.pdf
Page 2 of 327

spotify_streams.pdf

"1 artist(s) collaborated on the track 'Cruel Summer' by Taylor Swift, which was released on 8/23/2019. The song has been featured in 7858 Spotify playlists and reached 100 on the Spotify charts, accumulating 800840817 streams. Additionally, it has appeared in 116 Apple playlists and 207 on the Apple charts, while also being included in 125 Deezer playlists and reaching 12 on the Deezer charts. The track is characterized by a tempo of 170 BPM, a danceability of 55%, a valence of 58%, and an energy level of 72%. It has an acousticness of 11%, an instrumentalness of 0%, a liveness of 11%, and a speechiness of 15%. The song is in the key of A and is classified as Major. You can find the cover art [here](https://i.scdn.co/image/ab67616d0000b273e787cffec20aa2a396a61647)."

"1 artist(s) collaborated on the track 'WHERE SHE GOES' by Bad Bunny, which was released on 5/18/2023. The song has been featured in 3133 Spotify playlists and reached 50 on the Spotify charts, accumulating 303236322 streams. Additionally, it has appeared in 84 Apple playlists and 133 on the Apple charts, while also being included in 87 Deezer playlists and reaching 15 on the Deezer charts. The track is characterized by a tempo of 144 BPM, a danceability of 65%, a valence of 23%, and an energy level of 80%. It has an acousticness of 14%, an instrumentalness of 63%, a liveness of 11%, and a speechiness of 6%. The song is in the key of A and is classified as Minor. You can find the cover art [here](https://i.scdn.co/image/ab67616d0000b273ab5c9cd818ad6ed3e9b79cd1)."

Introduction to Natural Language Generation (NLG)

1 Automated Text Creation

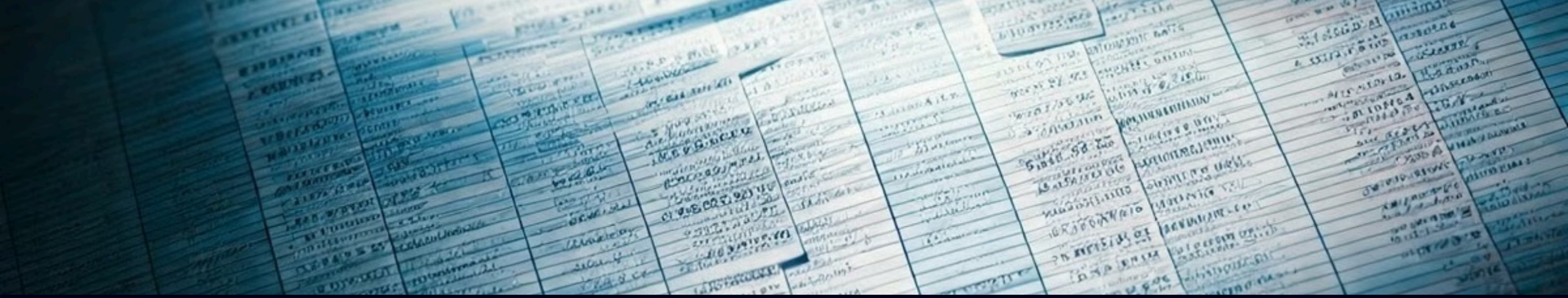
NLG systems convert tabular data into human-readable narratives and summaries.

2 Contextual Understanding

NLG leverages language models to generate text that is tailored to the audience and use case.

3 Scalable Personalization

NLG enables the creation of dynamic, customized content at scale.



Limitations of Tabular Datasets: Compatibility with LLMs

Since LLMs like GPT are built to understand and generate natural language, raw numerical data or tabular formats can be less efficient for extracting meaning and insights.

NLG Use Cases: Reports, Summaries, and Personalized Content

Automated Reporting

Generate comprehensive, data-driven reports to inform decision-making.

Summarization

Distill key insights and highlights from large datasets into concise summaries.

Personalized Content

Tailor content and narratives to individual user needs and preferences.

Customer Communications

Craft engaging, data-driven messages to enhance customer experiences.

NLG Process from Tabular Data

1

Context and Form Extraction

- What are the column names and their data types?
- What do the columns contain: statistical distribution, unique values, etc.?
- Which features are numerical? Categorical?

2

Template Generation

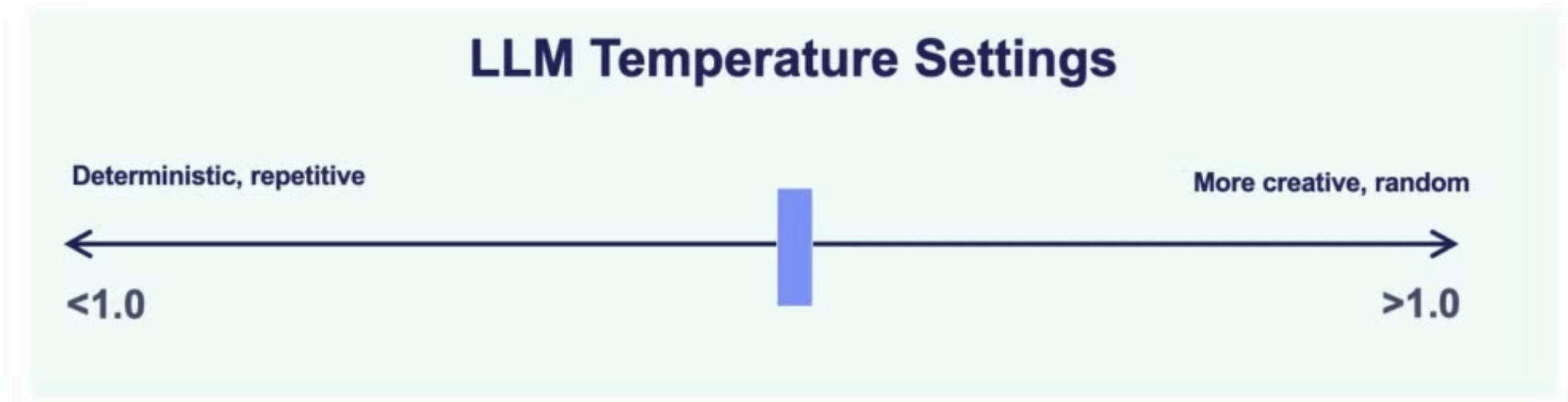
- What template can be used to generate the narrative of the dataset?
  - {song} by {artist} has a total of {stream_count} streams from {start_date} to {end_date}

3

Template Population

- Populate the template using the actual values of the dataset per row
  - Blinding Lights by The Weeknd has a total of 3.32B streams from Nov 29, 2019 to Jan 1, 2023

Additional Concepts: Temperature in LLMs



LOWER temp: If you need concise, factual, and direct answers (such as answering technical questions, providing definitions, or delivering structured information), a lower temperature ensures consistency and reduces randomness.

Use Cases:

- Customer support with factual responses
- Generating summaries of data
- Providing definitions or step-by-step instructions
- Writing technical content with minimal variation

HIGHER temp: If you want the model to come up with more varied, nuanced, or open-ended ideas, a higher temperature can encourage the model to explore different possibilities.

Use Cases:

- Storytelling or creative writing
- Brainstorming ideas (e.g., for marketing campaigns or innovation)
- Generating dialogue or character-driven content
- Fun, informal conversations or creative responses

Additional Concepts: Retry Decorator

A **retry decorator** is a function that automatically retries execution if it encounters certain types of errors (like network timeouts or server issues). It allows users to specify the number of retry attempts, the wait time between retries, and other conditions.

Why Use a Retry Function/Decorator for LLM APIs?

1. **Handle Network Failures**: Retry on transient connectivity issues or timeouts, improving reliability.
2. **Manage Rate Limiting**: Automatically retry when hitting rate limits (e.g., HTTP 429), often with exponential backoff.
3. **Deal with Temporary Outages**: Recover from temporary API downtimes or server issues without manual intervention.
4. **Format Error Handling**: Retry on specific format errors to ensure outputs are give as expected/desired.