

# Informe Calidad de Datos

Grupo R5 - Dirección de Analítica

# Reporte de anomalías

1. Fuente de datos
2. Consideraciones de anomalías
3. Estadísticas (Métricas)
4. Ejemplos de anomalías

# 1. Fuente de datos

Fuente: API Spotify

Datos: Discografía Taylor Swift

La información analizada en este informe ha sido recopilada utilizando la interfaz de programación de aplicaciones (API) de Spotify, una plataforma líder en streaming de música. Específicamente, los datos examinados se centran en la artista Taylor Swift, cuya presencia en la industria musical ha sido un fenómeno notable.

La API de Spotify proporcionó acceso a una variedad de atributos detallados relacionados con las canciones, álbumes y artistas disponibles en su extenso catálogo. En este análisis, nos sumergimos en los datos asociados con Taylor Swift, explorando aspectos que van desde las características musicales de sus canciones hasta la popularidad de sus álbumes.

Definición de dato:

En el contexto de este informe, un "dato" se refiere a la información específica obtenida mediante el cruce entre una fila y una columna en el conjunto de datos. Cada celda en esta matriz de datos representa un valor singular que proporciona detalles sobre aspectos particulares de la discografía de Taylor Swift. Estos valores abarcan desde atributos detallados de canciones, álbumes y artistas hasta métricas relacionadas con la popularidad, características musicales y otra información relevante recopilada de la API de Spotify.

## 2. Consideraciones de Anomalías

- 2.01 Id de canción nulo.
- 2.02 Filas duplicadas.
- 2.03 Valores nulos.
- 2.04 Formato incorrecto en nombres de canciones según la convención de nombramiento en inglés.
- 2.05 Caracteres mal codificados en nombres de canciones.
- 2.06 Datos no booleanos en la columna 'explicit'.
- 2.07 Datos no numéricos en la columna 'album\_total\_tracks'.
- 2.08 Formato diferente a fecha en la columna 'album\_release\_date'.
- 2.09 Formato no numérico en la columna 'audio\_features.instrumentalness'.
- 2.10 Datos no convertibles a numéricos en la columna 'audio\_features.instrumentalness'.
- 2.11 Valores fuera del rango [0,1] en la columna 'audio\_features.danceability'.
- 2.12 Valores fuera del rango [0,1] en la columna 'audio\_features.energy'.
- 2.13 Valores fuera del rango [0,1] en la columna 'audio\_features.liveness'.
- 2.14 Valores fuera del rango [3,7] en la columna 'audio\_features.time\_signature'.
- 2.15 Valores fuera del rango [-1,11] en la columna 'audio\_features.key'.
- 2.16 Valores fuera del rango [-60,0] en la columna 'audio\_features.loudness'.
- 2.17 Valores fuera del rango [0,100] en la columna 'track\_popularity'.
- 2.18 Valores fuera del rango [0,100] en la columna 'artist\_popularity'.
- 2.19 Valores fuera del rango [82000, 10000000] en la columna 'duration\_ms'.
- 2.20 Valores fuera del rango [2006, 2024] en la columna 'album\_release\_date'.

### 3. Estadísticas (Métricas)

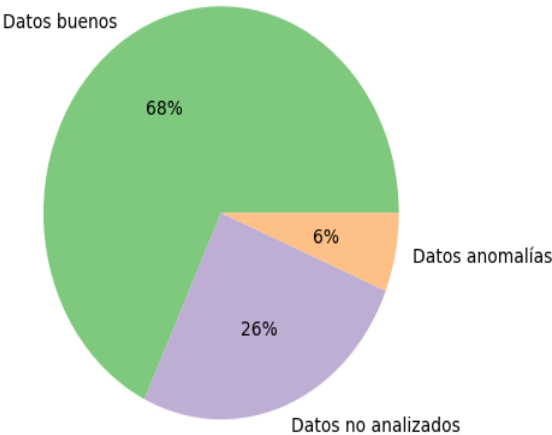
#### Estadísticas de los datos

Número de variables	27
Cantidad de observaciones	575
Total de datos	15525
Celdas vacías	89
Celdas vacías (%)	0.57
Filas duplicadas	54
Filas duplicadas (%)	9.39

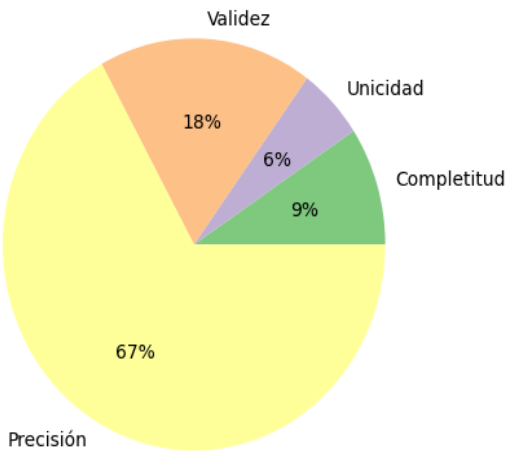
#### Tipos de variables

Numérico	16
Texto	11
Fecha	0

Resumen de análisis



Tipos de anomalías

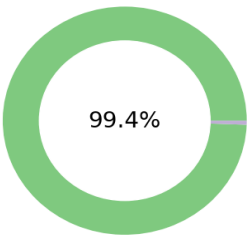


#### Puntuación global

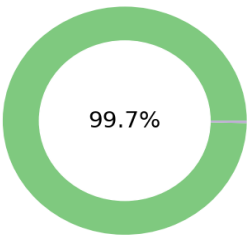
Calidad de los datos: 93.8

Puntuación del total de datos, comparada con cada una de las categorías de anomalías.

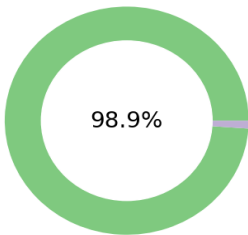
Completitud



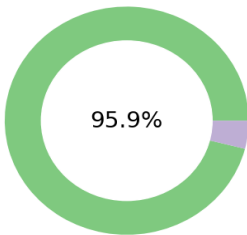
Unicidad



Validez



Precisión



Análisis de valores únicos

Columnas de texto

explicit	4
track_id	512
track_name	331
audio_features.instrumentalness	240
audio_features.id	519
artist_id	1
artist_name	1
album_id	26
album_name	24
album_release_date	23
album_total_tracks	17

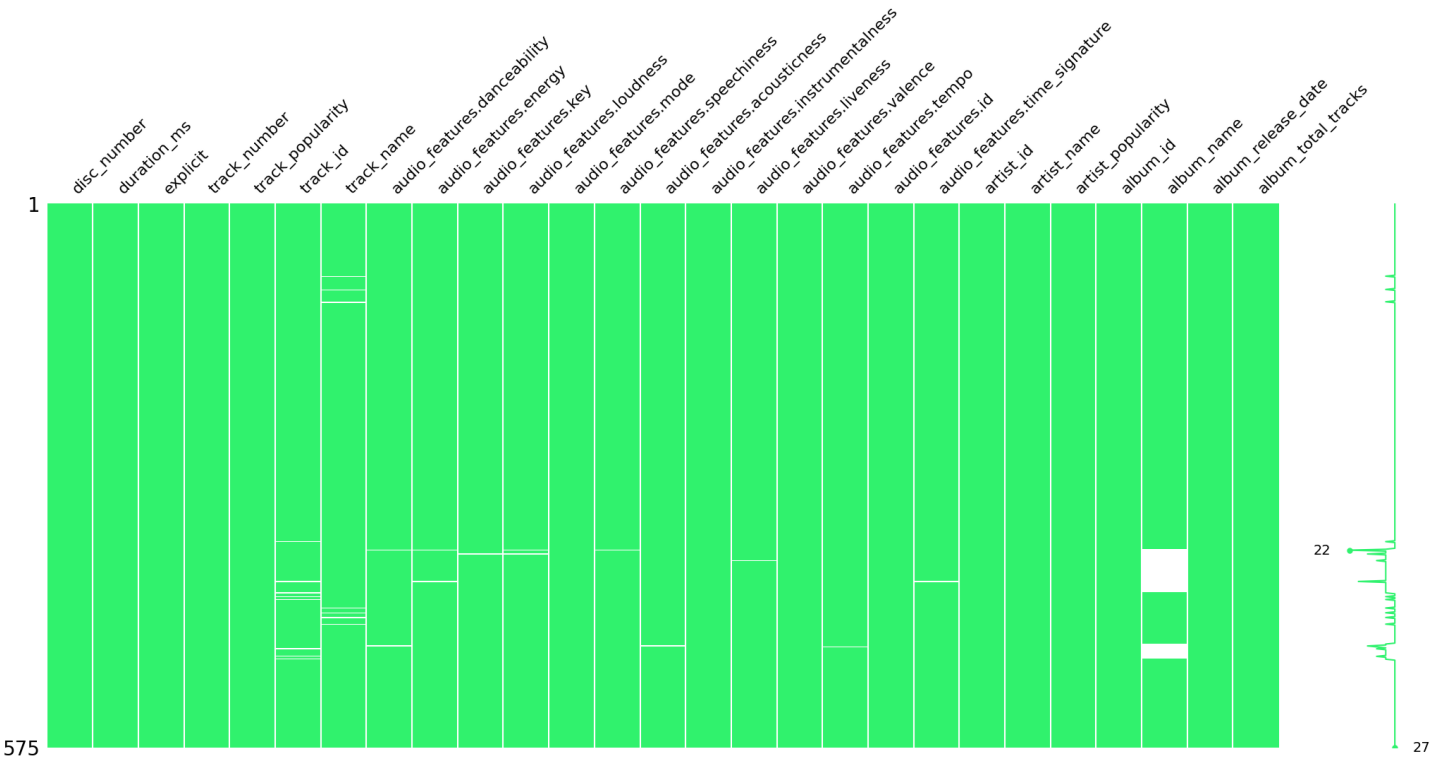
Columnas numéricas

disc_number	2
duration_ms	364
track_number	46
track_popularity	73
audio_features.danceability	267
audio_features.energy	348
audio_features.key	12
audio_features.loudness	448
audio_features.mode	2
audio_features.speechiness	292
audio_features.acousticness	401
audio_features.liveness	271
audio_features.valence	326
audio_features.tempo	450
audio_features.time_signature	3
artist_popularity	1

Columnas de fecha

Sin valores

Matriz de nulidad



## Ejemplos de anomalías

### Ejemplo 1:

Anomalías en la columna 'album\_release\_date':

- La columna 'album\_release\_date' contiene fechas de lanzamiento de álbumes, donde el formato de fecha es YYYY-MM-DD. Pero idealmente esta columna debería ser de tipo datetime, y no de tipo object.

album\_release\_date\_type\_is\_datetime = False

- Otra anomalía en esta columna es que hay fechas de lanzamiento de álbumes que son mayores al año actual y al año en que Taylor Swift lanzó su primer álbum.

year\_anomalies = ['2027-05-26', '1989-10-24']

### Ejemplo 2:

Anomalías en la columna 'audio\_features.instrumentalness':

- La columna 'audio\_features.instrumentalness' contiene valores que indican si una canción no contiene voces. Esta columna debería ser de tipo numérico, y no de tipo object.

instrumentalness\_type\_is\_numeric = False

- Otra anomalía en esta columna es que hay valores que no pueden ser convertidos a numéricos.

instrumentalness\_type\_conver\_anomalies = ['7.28x-06']