DATA QUALITY REPORT

TAYLOR SWIFT SPOTIFY

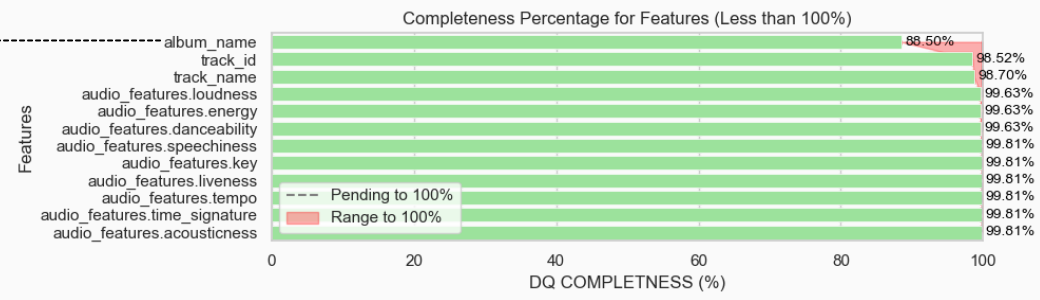
Context

The dataset comprises information about Taylor Swift's albums, including details of the audio track features.
It consists of 539 records and 27 features, including:

- 11 object types.
- 10 float64 types.
- 6 int64 types.

Completeness

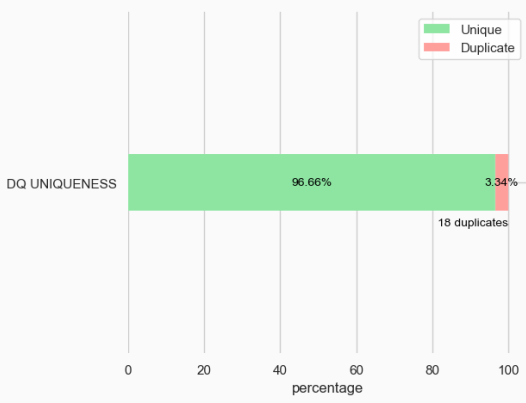
Approximately 44% of the features, as of 12/27, have at least one blank record.



The field with the highest number of blank records is "album_name" with 62 blank records (88.5% complete).

Uniqueness

The dataset contains **18** duplicate records **(3.3%)**



Timeliness

The information available doesn't specify the extent to which data accurately depicts reality at a given time.

Nevertheless, having timely information is crucial, which depends on setting up service level agreements between technical and business resources.

Validity

Type

1.explicit:
object --> bool:
Five values contain an integer different from the type expressed in the documentation (bolean).
2. track_id, track_name, album_total_tracks:
object --> integer:
Define the object type format, different from documentation (integer).
3. audio_features.key, audio_features.time_signature:
float64 --> integer:
Decimal Type is different from what is referred to in the documentation (integer).
4. audio_features.instrumentalness:
object --> float64:
A value contains an "x" which converts it to an object type, different of documentation (float).
5. audio_features.id, artist_id, artist_name, album_id, album_name, album_release_date:
object --> string:
Define the object type format, different from documentation (string).

Format

1.explicit:
Five values contain an integer different from the type expressed in the documentation (bolean).
2.audio_features.instrumentalness:
It contains 1 value with an "x" which does not correctly represent an exponential value.
3.album_total_tracks:
It contains 15 records where 30 is represented with a cardinal number, different from the documentation that indicates an integer for this field.

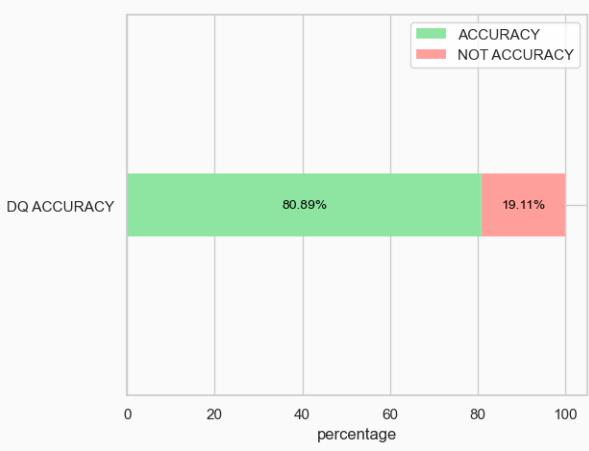
Range

1.duration_ms:
It contains 2 values out of range (negative), the track time could not be a negative number.
2.track_popularity:
It contains 7 values out of range (6 negatives and 1 more than 100), the value must be between 0 and 100.
3.audio_features.acousticness:
It contains 5 values out of range (2 negatives and 3 more than max value), the value must be between 0 and 1.
4.album_release_date:
It contains 15 records out of range, the year 1989 was when the artist was born, it can not correspond to a year of release.

Accuracy

Of the 539 records, only 436 **(81%)** represent correct values and in the correct representation, that is, they are accuracy (without determining the type of data that can be corrected most easily).

However it is important to check against the data source to determine if there are more values that are not exact.



Consistency

To ensure data consistency, it's essential to cross-check against another database or source.

This comparison helps to verify each field's existence, ensuring that they match in value and format.

Conclusion

The "taylor_swift_spotify" dataset includes blank records and a few instances of duplicate values and some validity issues. It's crucial to address these concerns to obtain accurate information, ensure operational efficiency, and promote confidence in decision-making.
Also, it is essential to consider various factors such as the usability, timing, flexibility, confidence and value of the data