

Report

Charting the Music Scene: An Analysis of the 2020s

A Data Analysis Report by Jasmine Pearson

The objective of this project is to examine music trends in the current societal climate by analyzing the top-ranking tracks, albums, and artists across a variety of genres. A Spotify API was coded to retrieve the top 50 tracks from each of 2020 to 2024 in six main genres. The scope of this analysis explores relationships between track, artist, and album popularity as measures of success, artist follower counts, track sub-genres, album record labels, and release dates. An application for this project would be for an emerging artist to grasp the landscape to mimic decisions with proven success, and highlights areas to focus efforts within the industry.

Initially, a pre-made data set available online was going to be used. Attention shifted when there appeared to be no datasets obtainable that fitted the desired criteria. From there construction of a Spotify API was researched then coded using Python in a Jupyter Notebook. Significant challenges arose when attempting to retrieve track audio features, such as key, time signature, and danceability; Spotify revoked access to these features at the end of 2024. The data was originally compiled into three csv files with the intention of using table joins in Looker Studio. However, that did not appear to be possible, so a single csv file was used instead. A large amount of data was missing for a number of hip-hop tracks, so those records were excluded from the final table. Looker Studio was used to visualize the data since it is free to use and does not require a Windows computer.

The focus of the top tracks' dashboard is the time series chart displaying the number of tracks released during each epoch week compared to their combined average popularity. In Spotify terms, popularity is a relative ranking in the range of zero to 100 and is the best measure available for evaluating success on the platform. Generally, the average popularity has a direct relationship with the number of tracks released, as in weeks 3, 30, and 49 most noticeably. However, some weeks show an inverse relationship between the two factors, such as weeks 14, 24, and 52. The chart can be used in conjunction with the 'release day' pie chart, where Friday holds the overwhelming majority, to predict the days of the year to release music for the best chances of success. When the tree map is used to filter the dashboard, it can be observed that the three most common sub-genres straddle multiple main genres. The first and third most common sub-genres (red dirt and r&b) have performed relatively well

with their most common track popularities falling within the 71-80 range. R&b has a tighter distribution in popularity rankings than red dirt.

Through the bubble chart, it can be learned that there is a gradual upward trend with weak correlation of artist popularity compared to track popularity. This trend shows that the popularity of an artist does help the popularity of the track to an extent, but there are many cases where a track can be successful irrespective of the artist who released it. The bubble size shows the number of tracks by the artist within the dataset. There is a rough trend where the more popular an artist is, the greater number of tracks appear in the dataset. This could be due to the resources that come with music career success, thus ability to release more music. The 'artist followers' chart exhibits a right skewed trend that resembles a negative exponential decay in which the majority of artists have up to 1.5 million followers, from the data available. In term of the industry, the chart exemplifies that there is plenty of potential for success for up-and-coming artists.

The album dashboard sheds light on the performance of a selection of record labels. The stacked column chart presents the number of albums released under the label and the maximum and minimum popularities of the albums. Despite the greatest number of albums being signed under Columbia, the greatest median popularity came from Chappell Roan PS/ Island. This graph serves as a good indicator for artists looking to be signed as to what success they might expect from each label and how popular of a choice the label is within the industry. The 'album tracks' pie chart is an extension of the 'album type' pie chart. Together they suggest the target length of an album, in terms of tracks rather than duration, to optimize chances of ranking highly in the charts. As seen in the other dashboards, too, the 'count vs popularity' chart serves as a general reference for the data sample being analyzed when filters are applied.

The key takeaway from this project was gaining familiarity with Looker Studio. Along the way, several valuable side objectives were accomplished, including learning how to build an API, constructing a data analysis report, and selecting appropriate chart types to effectively visualize insightful comparisons. If this project were to be implemented for an emerging artist, improvements that will be made include using BigQuery to join the three tables initially built and selecting tracks from one genre from a broader date range. Data will be pulled from the artist's Spotify to gain perspective on their current standings. Consequently, a more accurate picture can be illustrated that is tailored to the artist.

Appendix

Spotify API documentation: <https://developer.spotify.com/documentation/web-api>

Spotify API Python code:

<https://www.kaggle.com/code/jaspearson/spotify-api-tracks-albums-artists-for-5-years>

Dashboards: <https://lookerstudio.google.com/reporting/96df509c-7b81-47ad-a10d-10731d01dd9d>

