Proposal: Capstone Project Two
Predicting track popularity on Spotify

Springboard Data Science Career Track
Leo Evancie

With around 345 million users and 155 million subscribers worldwide, Spotify holds an estimated 32-35% market share for online music subscription services[1]. Anyone who makes music, from indie garage bands to legacy record labels, stands to benefit from optimizing their content for this platform. Similarly, Spotify stands to benefit from predicting the popularity of new music. Conveniently, Spotify provides a web API and a Python library (Spotipy), through which we can glean relevant data and insights.

Spotify Web API Non-User Data

Data Group	Methods	Data Endpoints
Albums	Get an Album or Several Albums Get an Album's Tracks	Album Name, Album Type, Artist(s) Info, Track Info, Images, Release Date, Total Tracks, Popularity, Preview URL, Spotify URI, Spotify ID
Artists	Get an Artist or Several Artists Get an Artist's Albums Get an Artist's Related Artists Get an Artist's Top Tracks	Artist Name, Genres, Popularity Score, Followers, Images, Spotify URI, Spotify ID
Browse	Get Available Genre Seeds Get a Browse Category OR Categories Get a Category's Playlists Get a List of Featured Playlists Get a List of New Releases Get Recommendations Based on Seeds	Genre Seeds, Search Seeds, Category's Playlists, New Release Track Information
Tracks	Get a Track or Several Tracks Get Audio Analysis for a Track Get Audio Features for Several Tracks Get Audio Features for a Track	Album Info, Artist Info, Available Markets, Popularity, Trak Number, Explicit, ISRC, Spotify URI, Spotify ID Danceability, Energy, Key, Loudness, Mode, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo, Duration, Time Signature

https://medium.com/@maxtingle/getting-started-with-spotifys-api-spotipy-197c3dc6353b

Since Spotify's data include popularity ratings, we can model popularity as a function of any number of variables, e.g., genre, song length, number of tracks on an album. What's more, Spotify encodes typically subjective musical qualities like "danceability" and "instrumentalness" numerically, so we can also explore whether such features improve our model.

Applying the data science method to the Spotify database could yield insights into how musical acts can set themselves apart on one of the most-used streaming platforms. This analysis would be useful for independent artists, established acts, and record labels. In a famously capricious industry where so much of success is determined outside of the creator's control, it is invaluable to understand how a few key decisions may be associated with better streaming popularity (i.e., more revenue).

On the platform side, a popularity-prediction model would improve decisions about which music to highlight for listeners on the "New" and "Discover" pages, especially for newer users about whom Spotify has not yet gleaned enough data to tailor recommendations. If Spotify can accurately gauge whether newly-hosted music will be popular, it can preferentially serve those tracks, boosting engagement.

[1]"Spotify Revenue and Usage Statistics (2021)", by Mansoor Iqbal, *BusinessofApps*. 04/02/2021.
https://www.businessofapps.com/data/spotify-statistics/