Capstone Project I - Proposal

Objective and Goal:

The goal of this project is to use the audio features of music tracks to predict the songs' popularity among listeners. Being able to predict a song or album's popularity before it is released has many applications in music business. As an example, a company named "The Music Fund", based in San Francisco, CA, takes advantage of models developed using data to predict popularity of a newly produced album by an artist. If the model predicts high probability of popularity, the company acts as a label and sponsors the musician. This is a purely data-oriented decision without the need for the sponsor company to know the artist or even like their music work.

Dataset:

The dataset to be used for this project is called "Spotify Audio Features" and is made available via the official Spotify web API. I accessed the dataset from the Kaggle competitions website(https://www.kaggle.com/tomigelo/spotify-audio-features). This is a dataset of more than 116,000 songs (one row per song) with 17 variables. The variables include Artist name, Track ID, Track name, Popularity (an integer between 0 and 100 that is calculated based on the number of streams a song had) as well as the following 13 audio features that characterize each song: Acousticness, danceability, duration_ms, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, time_signature and valence. For more details on the meaning of the audio features visit:

https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/

Methodology:

After doing some exploratory data analyses, the formal statistical analyses for this problem will be a regression since the outcome variable (i.e. Popularity) is a range of integers. However, if time allows, I might also turn this into a classification problem by grouping the outcome into the categories 0-10, 11-20, 21-30, …, 91-100 and trying to predict the category.

Final Deliverable:

The final deliverable will be the code, a slide deck and a 10-12 page report.