Spotify is one of the top streaming platforms with contents such as podcasts, audiobooks, and, mainly, music. The platform requires users to have a premium subscription in order to access and download millions of songs on multiple devices, without Wi-Fi. The main source of revenue for Spotify is the user subscriptions, so it is very important to maintain and increase user engagement and longevity for the business to grow in profit. It would benefit Spotify to know about upcoming hit songs in order to market to its users before the songs are released. This can help maintain high user engagement as they are constantly aware of new and popular music present on the platform. By using characteristics of past popular songs, Spotify can predict popularity for future songs and use them for efficient marketing in order to maintain and increase user engagement.

To begin, we need to collect quality and sufficient data, which can be found in this CSV file provided by Kaggle: https://www.kaggle.com/datasets/zeesolver/spotfy . Additionally, we need to identify which parts of the data will be useful to the problem we are trying to solve. This can be done through data wrangling. Once this step is complete, we can continue to the next step of exploratory data analysis where we can identify relationships between characteristics and see which of these characteristics affects the popularity of a song. Finally, we can develop a machine learning model that predicts the popularity of future songs using a scale that will measure popularity from 1 to 10.

The deliverables in this project include a GitHub repository containing all the Jupyter notebooks I will use for each data analysis step of the project, a slide deck explaining the important findings from the data set, and a project report explaining the overall project.