- **Intro:**

Spotify is a Swedish audio streaming and media services provider with over 365 million monthly active users, including 165 million paying subscribers, as of June 2021. Spotify provides access to over 70 million songs, 2.2 million podcasts and 4 billion playlists. This platform is used to collect a dataset for training a classification model to be able to predict whether a song was a hit or not a hit based on different song attributes such as danceability, loudness, duration, etc.

- **Question/need:**

Many music artists and producers wonder what makes songs hits or not, what is the perfect formula for a hit song which will be on billboard and top 100 charts week after week and month after month. How long should a song be, or how danceable or loud should it be, and which genres usually get the most love from worldwide listeners to contain the most hits. By doing this project, these questions will be answered to help guide producers and songwriters understand the criteria of hit songs.

- **Data Description:**

Data is collected using an open access Spotify dataset from Kaggle containing more than 20k rows and +10 feature columns with different attributes for a song such as danceability, loudness, duration, etc. This dataset obtained from Kaggle contains the rank of songs on the 100 top hits billboard and extra information where we can determine if they were hits or not. Another dataset is acquired using a Spotify API which was used to collect data from different Spotify lists and merged with the previous dataset via song id, song title, and artist to have in the end the songs that were hits labeled 1 and the songs that weren't not labelled 0.

- **Tools:**
    - Spotify API
    - SpotiPy library
    - Sci-kit Learn
    - Visualization libraries (Seaborn, Matplotlib)
    - Pandas, NumPy

- **MVP Goal:**

The goal MVP for this project should clearly communicate the answer for some of these questions using comprehensive visualizations as well as a preliminary assessment of the classification baseline model performance.