<https://www.kaggle.com/datasets/muhmores/spotify-top-100-songs-of-20152019>

I dette projekt har vi tænkt os at lave nogle forskellige statiske udregninger ud fra datasettet 'Spotify Top 100 Songs of 2010-2019’. Vi vil bruge pandas og numpy til dataudtrækning og datamanipulation fra vores dataset, samt vil vi bruge selenium til at webscrape sangtekster fra https://genius.com/

Vi regner bl.a. med at bruge genre, bpm, pop, artist type, dur, val og dnce som vores feature vektor. Vi vil svare på følgende spørgsmål vha. disse oplysninger:

1. Hvad er den mest optimale bpm indenfor dance pop, modern alternative rock og acoustic pop
   1. Vi vil tage top-5 mest populære sange inde for hver genre for at finde den gennemsnitlige bpm for hver genre
   2. Her vil vi bruge pandas dataframe
2. Vise en graf over de mest populære genre
   1. Her vil vi bruge matplotlib til at lave et barplot
3. Har årstal udgivet og val et sammenhæng
   1. Her vil vi lave linær regression for at se om sammenhængen er positiv eller negativ
4. Hvad er top-3 genre inden for de forskellige artist types
   1. Her vil vi bruge pandas dataframe
5. Hvilken artist har haft flest populære sange
   1. Her gør vi brug af get\_dummies
   2. Her vil vi bruge pandas dataframe
6. Hvilken genre vil blive mest populær i 2020
   1. Her gør vi brug af 3D graf hvor x=år, y=popularitet og z=genre
   2. Her vil vi bruge matplotlib til at lave en 3D graf
7. Sentiment analysis på top-5 mest populære sange
   1. Webscaper sangteksterne