På dette nettstedet samler vi oversikt over artister og band, musikken de har utgitt, hva slags sjangere de spiller, hva som har vært populært, hva slags terningkast albumene har fått, og hvor mye banning sangene inneholder.

Popularitet/Streaming – Key-value pair database

Dataset: https://www.kaggle.com/pieca111/music-artists-popularity

I musikk popularitets datasettet er det ferdigstilte id-er som kan brukes til å referere til artister, opphavsland, tagger om sjanger, språk, osv, og 2 forskjellige kanaler for avspilling (Valgt avspilling og anbefalt avspilling). I tillegg er det en boolsk verdi som sjekker om det er flere artister som deler samme kanal. Datasettet har også metadata fra to forskjellige opphav, en fra last.fm og en fra musicbrainz, begge med en oppføring for nasjonalitet og tagger med samme format. Dette gjør det lett å sjekke de forskjellige tag-ene og landene opp mot hverandre for å unngå dobbel utskrift eller mulige feil.

Topp spotify 2010 til 2019 – Document database

Dataset: <https://www.kaggle.com/leonardopena/top-spotify-songs-from-20102019-by-year>

Dette datasettet tar for seg en rekke interessante data. Det viser de mest populære sangene for hvert år fra 2010 til 2019, hvilken hovedsjanger sangen tilhører, bandet/artisten, men også annen data som beats per minute, hvor energisk sangen er, hvor dansbar den er, hvor høylytt den er, og hvor sannsynlig innspilling skjedde «live».

Det som gjør dette datasettet spesielt interessant er disse metadataene som kan være ekstremt behjelpelige for å anbefale lignende musikk/artister, men også for å skape spillelister som henger sammen. Det går an å da for eksempel lage «rolige», «energiske», og «danse» lister basert på enkle tallverdier.

Her føler vi det er logisk å legge dataen i dokument database hvor vi kan legge sangene/artistene etter år, og ha nøkler på sanger da de kun kan være oppført én gang.

Musikk popularitet over tid – graph

Dataset: <https://www.kaggle.com/jkrowling/profanity-in-music>

Dette settet ser på hvor mange ord som er i en sangtekst, og hvor mange av disse ordene som er banne ord. Settet artist, album(ikke alle) og låt, som gjør det lettere å lete igjennom artistene for å finne sanger med mye banning. Det kan også være gøy å se om det er noen sammenheng mellom hvor mye banning det er i en sang, og hvor populær den blir. Er det en magisk sone? Hvem banner mest?

Grammys – column

Dataset: <https://www.kaggle.com/unanimad/grammy-awards>

Grammy premier kan av og til virke som at de går til litt tilfeldige artister, så vi tenkte å vise her om sammenkobling mellom sangpopularitet og pris stemmer. Her kan vi også finne koblinger mellom artister som ikke nødvendigvis henger sammen, og det var en oversikt over alle medvirkende på sangene. Dette kan også gi oversikt over hvilke produsenter som ofte vinner, og hvilke artister/band som stikker av med flest premier hvert år. Vi kan også bruke datasettet for å se om det å vinne en Grammy påvirker videre lytting på sangene.