

## Milepæl 1 – Rapport – Prosjektgruppe 13

### Intro

Rapporten omhandler nettsiden vi skal lage som formidler forskjellig informasjon om musikk. Dette gjelder for artister, albumer, meta-data for ulike sanger, top hits i spotify og antall banneord per sang. Nettsiden skal vise interessante funn rundt musikk, særlig rundt hvilke trekk som går igjen blant populære sanger og hvilke særevner de har. Vi ser også for oss at man skal kunne få anbefalinger basert på meta-data i sangene, og man skal kunne søke etter sanger og album.

### Datasett

#### Nøkkel-Verdi database: 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, og to 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 har 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.

På nettsiden har vi tenkt til å bruke dette datasettet som et oppslagsverk av artister, og siden id-en passer bra som nøkkel, passer den bra til Nøkkel-Verdi database.

Datasett: <https://www.kaggle.com/pieca111/music-artists-popularity>

#### Dokument database: Top Spotify Songs From 2010 To 2019

Dette datasettet tar for seg en rekke interessante data. Det viser de mest populære sangene for hvert år fra 2010 til 2019, hvilken hoved-sjanger sangen tilhører, bandet/artisten. I tillegg inneholder den data som beats per minute (bpm), 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 hjelpelike for å anbefale lignende musikk/artister, men også for å skape spillelister som henger sammen. Da kan man for eksempel lage «rolige», «energiske», og «danse» lister, eller finne ut hva slags type musikk som var populært et spesifikt år, alt 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.

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

#### Kolonnedatabase: Profanity In Music

Dette datasettet ser på hvor mange ord som er i en sangtekst og hvor mange av disse ordene som er banneord. Den inneholder kolonner for 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?

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

### Grafdatabase: 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.

Sammen med de andre datasettene er det mulig å gi en oversikt over hvilke produsenter som ofte vinner, hvilke artister/band som stikker av med flest premier hvert år, og hvordan det å vinne en Grammy-premie påvirker opptreden til sangen.

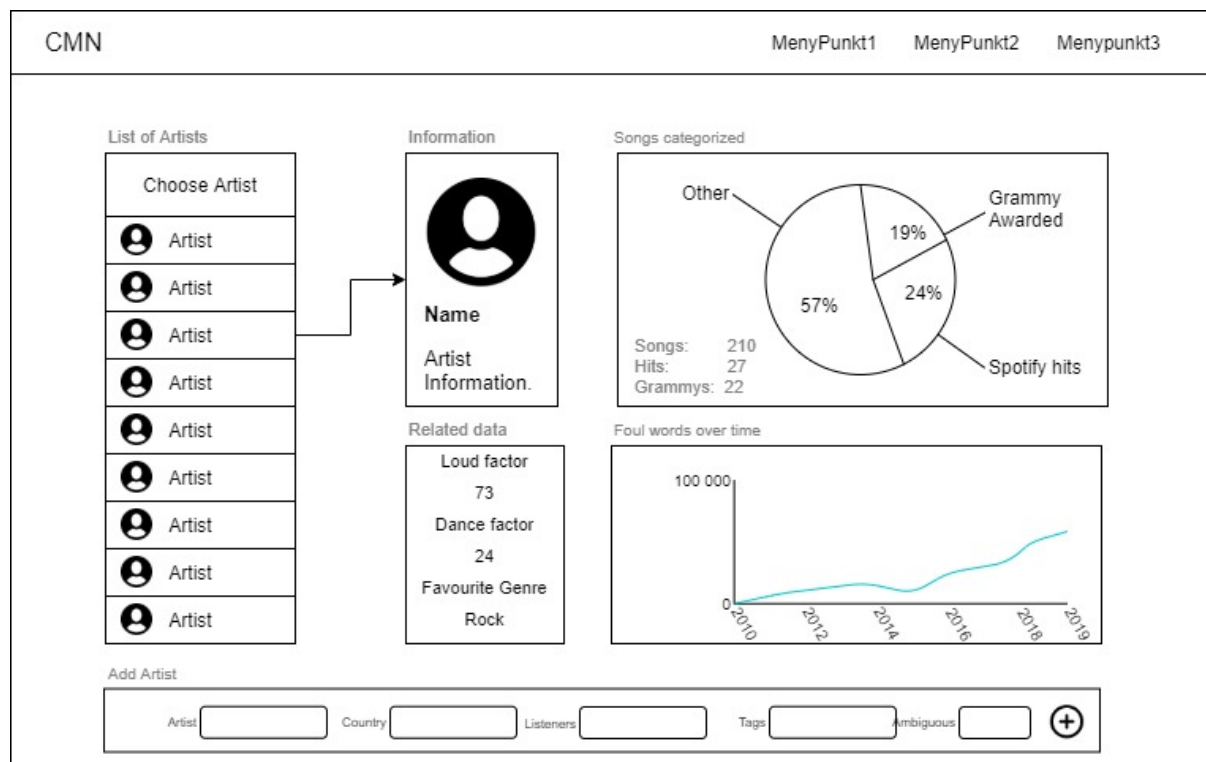
Vi tenkte at grafdatabase passet dette datasettet siden vi kan vise forhold mellom premiene og sangene, albumer og artister.

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

## Skisser Og Dataformidling

### Artister

Skissen under illustrerer hvordan informasjon relatert til ulike artister kan framstilles med datasettene. Her er det tenkt at når man velger en artist ut ifra listen, vil vinduene til høyre oppdateres med ulik informasjon som benytter data fra alle datasettene.



**Songs Categorized** vil gi en prosentvis oversikt over antallet sanger som har fått en Grammy Award, antallet sanger som ble en hit i spotify datasettet, og de resterende «vanlige» tilfellene. Utrekningen her gjennomføres ved å sjekke om sangen finnes i enten Grammy Awards- eller Spotify Hits-

datasettene, og dermed markere sangene ettersom. Da vil man stå igjen med en liste sanger og hvilken kategori de tilhører (Grammy Awarded, Spotify Hits og Other), og man kan vise de på en Pie Chart.

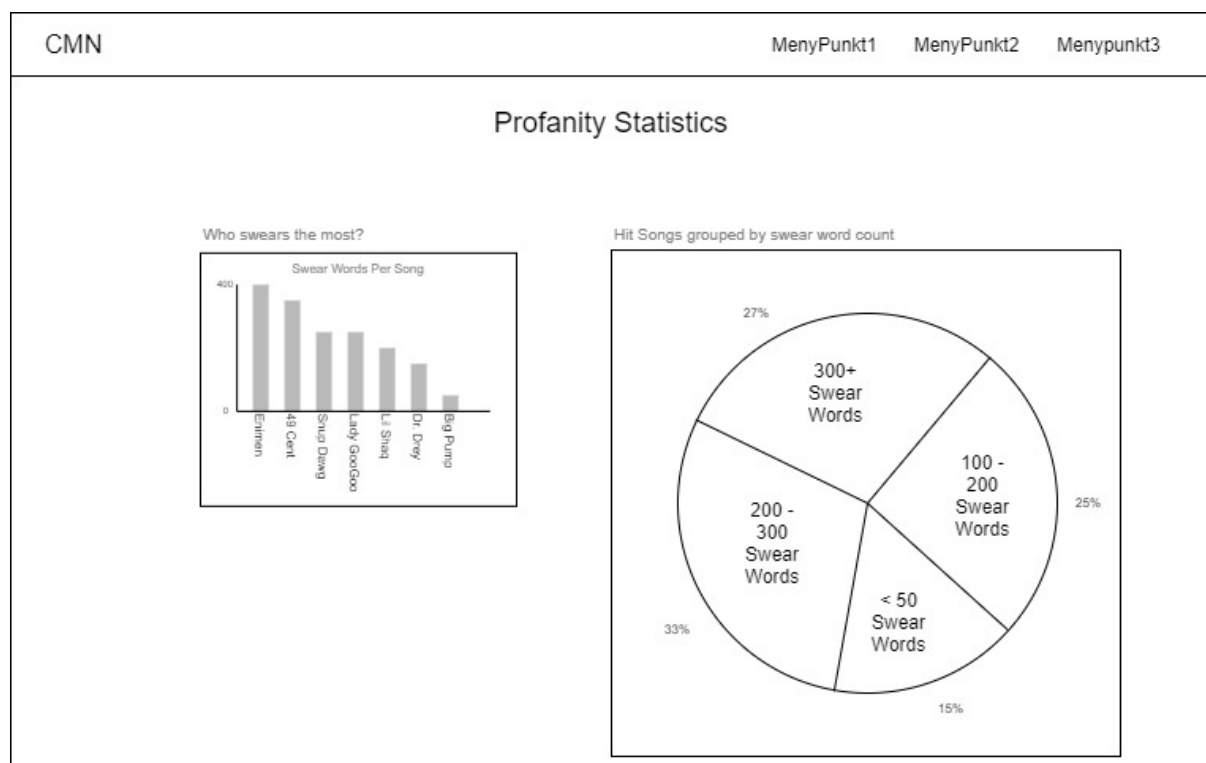
**Foul words over time** viser hvor utviklingen av antall bannord brukt av artisten hvert år. Måten dette beregnes, er ved å hente ut banneord per sang fra «Profanity In Music» datasettet, deretter legge sammen antall banneord basert på årstall (når sangene ble lansert). Da vil man ende opp med antall banneord per år, som kan for eksempel framstilles med et linjediagram.

Til slutt regnes **Related data** ved å finne gjennomsnittet av loudness-, danceability- og top genre-kolonnene fra *Top Spotify Songs datasettet*. Da kan man finne ut av hvor dansbar og høylitt artisten er, samt hvilken sjanger de liker best.

Neders på siden kan man legge til en ny artist. På feltene for Country og Ambiguous kan tekstfeltene byttes ut med nedtrekks-lister. I første omgang vil ikke den nylig tillagte artisten ha noen oppføring i de andre datasettene

## Statistikk om banneord

Dette er en side hvor man får se effekten av banneord i musikk.



**Who swears the most** tar for seg de ulike artistene og rangerer de fra mest til minst banneord per sang (i gjennomsnitt). Dette kan man beregne ved å legge sammen antall banneord fra alle sanger til en artist og finne gjennomsnittet (dele på antall sanger). Deretter sammenligner man med andre artister. Dette kan vises fram med en kolonne-tabell.

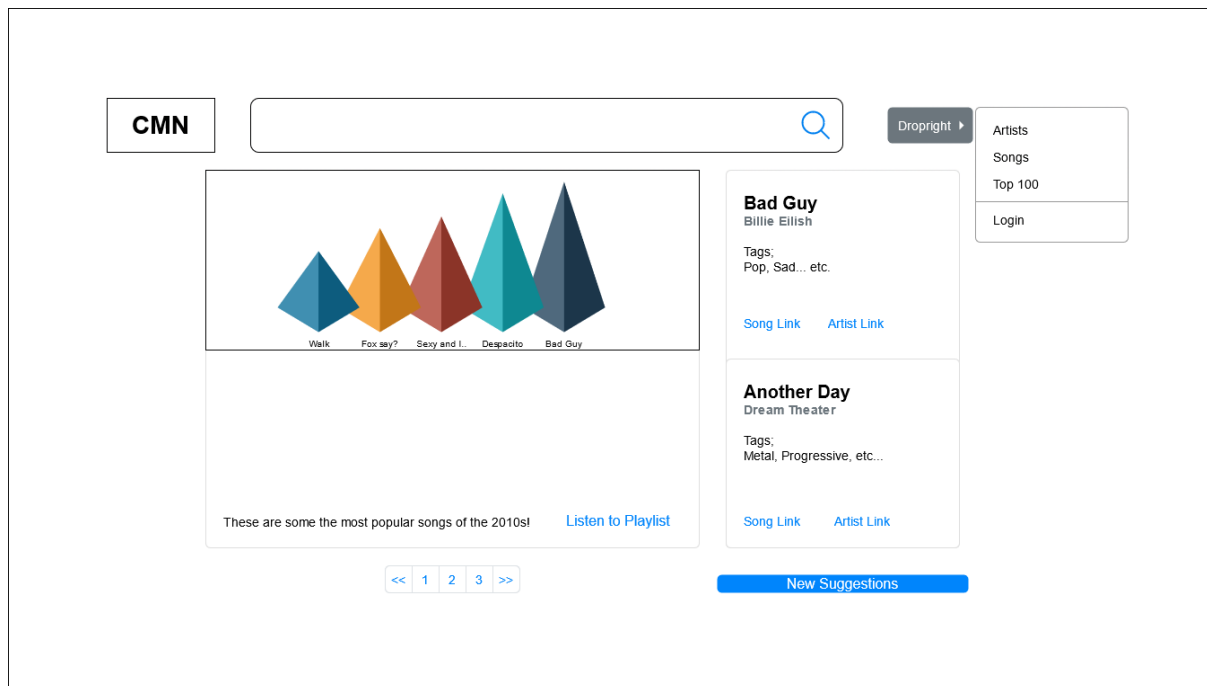
**Hit Songs grouped by swear words** viser relasjonen mellom antallet banneord og prosentvis antallet hit-sanger av vilkårlig artist. Dette kan være interessant for å se om antallet bannord har noe å si på opptreden til sangen. Denne kan regnes ut ved å først finne sangene fra *Profanity in Music*

datasettet og Spotify datasettet for å se hvilke sanger som er en hit. Deretter gruppere sanger på antall banneord. Det kan vises fram med en pie chart.

## Forside

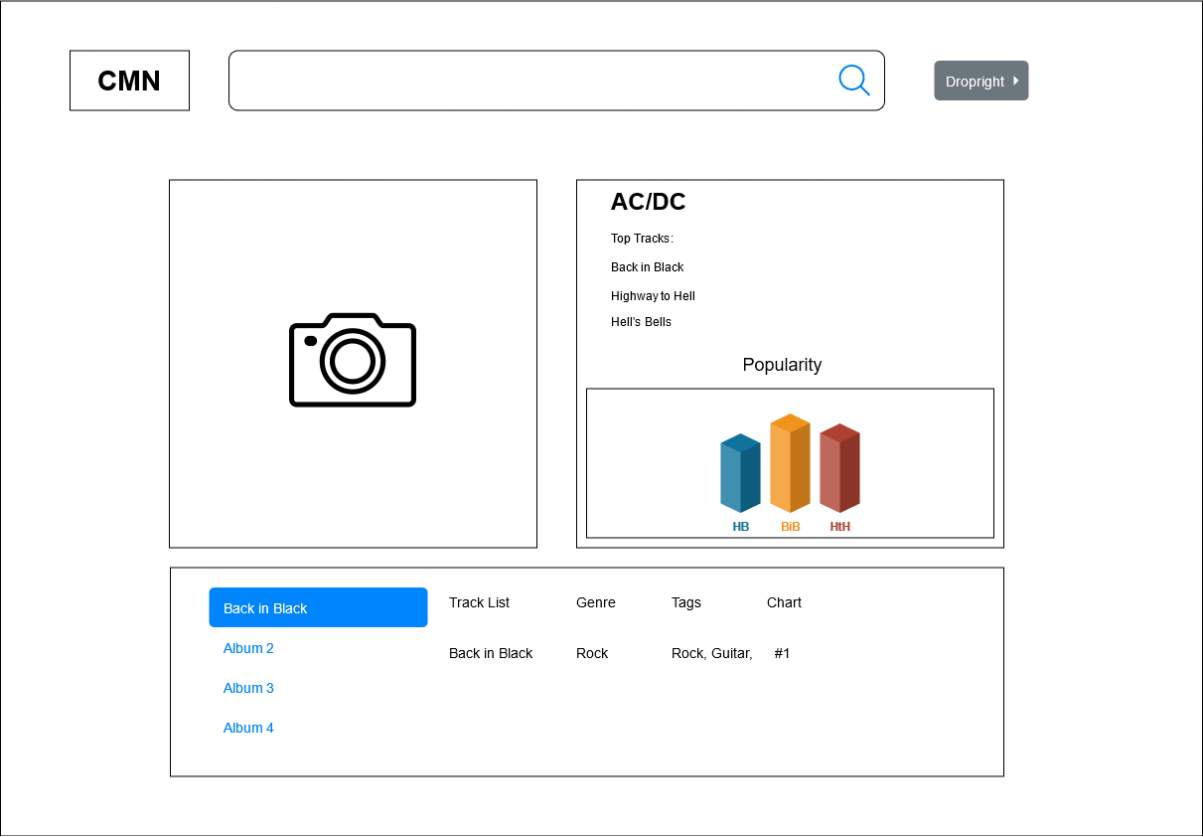
Her er en skisse som viser hvordan forsiden kan se ut. Det vil være et søkefelt der man kan søke etter artister og sanger. Under søkefeltet vises det statistikk som bruker kan klikke seg inn på, blant annet for å komme til artist-siden.

Vi har tanker om å vise anbefalte sanger basert på meta-dataen beskrevet i Spotify datasettet.



Inspisere album

Her vises det hvordan det kan se ut når man klikker seg inn på et spesifikt album. Dataen viser hvilke sanger albumet består av og relatert data.



## Kategorier

Side som viser forskjellige kategorier. Disse kan brukes til å vise hit-sanger for gitte år.

