IN2090 – Innleveringsoppgave 4
SQL 2

Publisert dato: 24.10.2024.

Innleveringsfrist: 08.11.2024 23:59.

Filer som skal leveres: 1: `innlevering4.sql`.

Antall forsøk: 2.

Les gjennom hele teksten nøye før du begynner å løse oppgavene. Oppgaven skal leveres som en ren tekstfil med filendelsen `.sql`. Bruk SQL-kommentarer for ren tekst og kommentarer.

HUFFs avdeling for Oversikt og Gruppering av Hovedtrekk ved Underbevisste Følelser hos Folk (HUFFOGHUFF, eller enda kortere HOFF) har som oppgave å holde styr på hva folk tenker, eller kommer til å tenke gitt spesifikke hendelser. De innser at deres nåværende data ikke er gode nok til dette formålet, og har kommet frem til at dette gjøres best ved å analysere folks og folkegruppers/lands film og TV-vaner.

Denne oppgaven skal derfor løses med filmdatabasen. Informasjon om hvordan du kobler deg til denne, finnes i dokumentene *Databaser for IN2090*[1] og databasen dere skal bruke er *Filmdatabasen*[2] lenket fra semestersidene.

I hver oppgave skal du skrive en `SELECT`-spørring som løser oppgavene. Ta med oppgangsnumrene som kommentarer, f.eks.:

```
-- Oppgave 0

SELECT count(*) FROM film;

-- Kommentarer på én linje startes med to bindestreker
/* Om du vil skrive kommentarer over flere linjer,
kan de skrives som denne kommentaren */
```

Antall rader spørringene skal returnere står i parentes. Ulike tolkninger kan føre til andre resultater. Det er uansett lurt å levere svar på oppgaver der spørringen din ikke gir forventet antall rader, det kan hende du var veldig nærme!

[1] `https://www.uio.no/studier/emner/matnat/ifi/IN2090/h24/programvare/postgres.pdf`

[2] `https://www.uio.no/studier/emner/matnat/ifi/IN2090/h24/programvare/filmdatabasen.pdf`

Oppgave 1 – Oppvarming

HOFF ønsker først å få litt oversikt over datagrunnlaget og hvordan de ulike tabellene henger sammen. Skriv derfor en spørring som finner navn på alle skuespillere og rollen de spiller i filmen med tittelen `'Star Wars'`. (108 rader)

Oppgave 2 – Land

For å starte på dette nye prosjektet ønsker HOFF å vite hvor mange filmer hvert land lager. Dette sier jo noe om hvor filminteresserte folkene i de ulike landene er, og dermed hvor godt all annen informasjon de finner ut i denne analysen kan anvendes på innbyggerne i det gitte landet.

Skriv derfor en spørring som finner antall filmer som er laget i hver land. Sorter resultatet fra høyest til lavest antall. (190 rader)

Oppgave 3 – Spilletider

HOFF ser av resultatet på forrige spørring at det er store variasjoner mellom land, og ønsker å undersøke dette litt nærmere. De lurer nå på om det kan være variasjoner mellom land på hvor lenge filmene varer, kanskje det kan gi oss noe informasjon om hvor tålmodige innbyggerne i landet er.

I tabellen `runningtime` listes det opp filmer sammen med deres spilletider i hht. ulike parametre. F.eks. kan ulike land ha ulike versjoner av samme film, og derfor ulik lengde. Spillelengdene er dessverre i tekst, men en del av dem er skrevet med kun tall og gir spilletiden i minutter. Vi kan finne ut hvilke rader som kun har tall i `time`-kolonnen med følgende uttrykk:

```
time ~ '^\d+$'
```

Skriv en spørring som finner gjennomsnittlig kjøretid for filmer per land (hvor `country` ikke er `NULL`), men hvor vi kun ønsker å ta med de landene hvor vi har minst 200 kjøretider (i minutter som beskrevet over). (44 rader)

Oppgave 4 – Komplekse mennesker

En av de ansatte i HOFF fremholder at komplekse mennesker gjerne ser på komplekse filmer, og å ha en liste over de mest komplekse filmene ville vært nyttig for videre analyse. En kompleks film er, i henhold til denne ansatte, en film som har mange sjangre.

Så skriv en SQL-spørring som finner de 10 kinofilmene med flest sjangre (kinofilmer har `'C'` i `filmitem.filmtype`-kolonnen). Dersom det er filmer med like mange sjangre, velg dem som kommer alfabetisk først. (10 rader)

Oppgave 5 – Land og filmvaner

HOFF innser at de blir nødt til å ytterligere tilpasse analysene sine etter hvilket land personen er født og oppvokst i, og trenger derfor mer informasjon om hvert lands overordnede filmpreferanser.

Skriv en spørring som for hvert land finner totalt antall filmer laget i det landet, gjennomsnittlig rating for disse filmene, og sjangeren som er vanligst for landets filmer. (170-190 rader, avhengig av antagelser)

Hint: Her er det lurt å forsøke å dele spørringen opp i flere delspørringer, hvor hver delspørring finner hver sin ting om landene. Atså, skriv en spørring som finner antall filmer, en som finner gjenomsnittlig rating, og en som finner den mest populære sjangeren, og så forsøk å kombiner disse resultatene til slutt.

Oppgave 6 – Vennskap

En svært viktig gruppe med følelser er de som fører til varig vennskap. En av de ansatte i HOFF mener av erfaring at kun virkelig gode venner kan jobbe sammen over lang tid, og at dette jo også må gjelde innen film og gjelder like godt for land som mennesker. Den ansatte ønsker derfor å finne hvilke land som har sammarbeidet mest om å lage filmer, ettersom dette da sier noe om vennskapet mellom landene.

Skriv derfor en SQL-spørring som finner navnene på alle par av land som har laget mer enn 150 filmer sammen (i henhold til `filmcountry`-tabellen). Merk: Hvert par av land skal bare forekomme én gang i resultatet. (40 rader)

Oppgave 7 – Mot

Et av HOFFs formål med denne analysen er å vite hvor godt folk ville taklet skumle fenomener og farefulle hendelser.

- *Hva avgjør om folk takler slike fenomener og hendelser da?*, spør lederen for prosjektet.
- *Hvorvidt de tørr å se skumle filmer*, forelår en ansatt.
- *Ja, grøssere!*, sier en annen.
- *Ja, og filmer fra Romania!*, påpeker en tredje. Stillhet, blanke ansikter og forundring fyller rommet.
- *Det er jo der vampyrer og varulver kommer fra...*, legger han til. De andre nikker anerkjennende.
- *Men dette blir vel veldig mange filmer?*, sier en.

– *Ja, la oss begrense oss til kun de skumleste, altså de som skjer i mørket!*, fastslår lederen.

Skriv derfor en spørring som finner navn og produksjonsår på alle filmer som har en tittel som inneholder minst ett av ordene `'Dark'` og `'Night'`, og som er grøsser (har sjanger `horror`) eller er fra Romaina (eller evt. begge). Resultatet skal ikke inneholde duplikater. (457 rader)

Merk: Det kan finnes filmer fra `'Romania'` som ikke har noen sjanger (altså ikke forekommer i `filmgenre`-tabellen) og `'horror'`-filmer som ikke har noe land (altså ikke er oppført i `filmcountry`).

Oppgave 8 – Lunsj

Etter en lang formiddag med mye analyse er det på tide med lunsj for de HOFF-ansatte. En av de ansatte lurer på om ikke du kan finne en film som pause-underholdning. For øyeblikket er derimot den ansatte litt lei av folk og deres følelser, og ønsker derfor å finne en film med lite folk. Filmen skal heller ikke være for gammel da.

Skriv derfor en spørring som finner tittel og antall deltakere til alle filmer laget i 2010 eller senere, og som kun har 2 eller færre deltakere (i henhold til `filmparticipation`-tabellen, uansett rolle). (28 eller 31 rader).

Merk: Dataene i filmdatabasen er litt mangelfulle, så det er ikke gitt at filmen faktisk har få deltakere, bare at de ikke finnes i databasen.

Oppgave 9 – Introspeksjon

I lunsjen ender de ansatte opp med å diskutere egne filmvaner, og innser at samtlige kun ser på Sci-Fi- og horror-filmer. De blir derfor litt usikre på om det i det hele tatt finnes filmer som hverken er Sci-Fi eller horror.

Skriv derfor en spørring som finner antall filmer som hverken har sjanger `'Sci-Fi'` eller `'Horror'`. (1 rad)

Oppgave 10 – Kompetanseheving (vanskelig)

De ansatte i HOFF innser nå at de har relativt liten innsikt innen film, og at dette kanskje kan påvirke kvaliteten til analysene fra dette prosjektet. Etter et krisemøte blir det laget en ny handlingsplan hvor en hovedprioritering er kunnskapsinnhenting og kompetansebygging innen film. Planen er nå å se et utvalg av de mest interessante filmene, for å dekke bredden innen film.

De definerer en *interessant film* som en kinofilm som både har en `rank` høyere eller lik `8` og har mer enn `1000` `votes`. Dette blir dermot altfor mange filmer å se, så de må gjøre et utvalg av disse.

Etter en opphetet diskusjon ender de opp med følgende kompromiss til utvalg av de interessante filmene:

- De 10 som har høyest rank og votes (altså om de har lik rank, så velg de som har høyest votes).
- De filmene som Harrison Ford er skuespiller i.
- De filmene som har sjanger `'Comedy'` eller `'Romance'`.

Du får derfor i oppgave å lage en spørring som finner tittel på de utvalgte filmene beskrevet over. (170 rader)

Hint: Her er det lurt å bryte spørringen opp i delspørringer i en `WITH`*-klausul. Start med å først finne* `filmid` *til alle* interessante filmer*, og så skriv delspørringer som gjør uttrekk fra disse i henhold til listen over (altså skriv en delspørring som finner* `filmid` *til de 10 høyest rangerte interessante filmene, så en annen delspørring som finner* `filmid` *til de interessante filmene med Harrison Ford som skuespiller, og så en delspørring som finner* `filmid` *til de interessante filmene som har sjanger* `'Comedy'` *eller* `'Romance'`*) og så kombiner alle disse* `filmid`*'ene med passende mengdeoperasjon, og til slutt hent ut tittel på filmene.*

Levering

Innleverinsoppgaver leveres i Devilry[3]. Sørg for at du er registrert i systemet ved å logge inn og se at `innlevering4` er tilgjengelig som en oppgave for IN2090. *Sjekk dette før du begynner å løse oppgavene!*. Dersom du ikke er registrert, send en mail til `leifhka@ifi.uio.no`.

Etter innlevering vil du få tilbakemelding på innleveringen din, i likhet med en oblig. Selvom innleveringsoppgaven ikke er obligatorisk, er den et svært nyttig læringstilbud og en viktig del av undervisningen i kurset. Fokuser på å lære mest mulig og les tilbakemeldingene du får fra retterne nøye. Merk også at det ikke blir publisert løsningsforslag på innleveringsoppgavene, så å levere dem inn og deretter få tilbakemelding er eneste måte å få kjennskap til hvor god egen besvarelse er.

Du kan levere så mange ganger du vil, men det er kun den siste leveringen du vil få tilbakemelding på. Dersom du ønsker å levere et andre forsøk med tilhørende tilbakemelding, vennligst oppgi dette som en kommentar når du leverer, slik at retterne kan åpne for et andre forsøk etter å ha gitt tilbakemelding.

Lykke til!

[3] `https://devilry.ifi.uio.no/`