

De vragen zijn ingedeeld in drie categorieën:

- Categorie A     Vragen die met een SQL-query op te lossen zijn en waar je alleen de vraag en het antwoord (of de antwoorden) hoeft te laten zien
- Categorie B     Vragen waarvoor een visualisatie nodig is
- Categorie C     Vragen waarvoor een R-model nodig is

Zie ook de projecthandleiding voor richtlijnen over hoeveel vragen van welke categorie je mag/moet kiezen.

## **Vragen Categorie A (SQL Algemeen)**

- A1. Welke acteur (m/v) die in Friesland is geboren heeft het meeste succes gehad. Denk na over hoe je succes definieert.
- A2. Welke acteur (m/v) heeft de meeste dubbelrollen?
- A3. Welke acteur (m/v) heeft de langste filmcarrière? (dus geen series)
- A4. Welke Nederlandse film heeft het meeste opgebracht in de bioscoop?
- A5. Welke film heeft het langst geduurd om op te nemen?
- A6. Welke tv-serie is het duurst geweest om op te nemen? Bonus: houd hierbij rekening met de inflatie.
- A7. Welke film is het duurst geweest om op te nemen? Bonus: houd hierbij rekening met de inflatie.
- A8. Welke film heeft het meest opgebracht in een week/maand/jaar/altijd
- A9. Welke films spelen (gedeeltelijk) in New York?
- A10. Welke films spelen in meer dan 1 land?
- A11. Welke schrijvers spelen in hun eigen films en welke films zijn dat?
- A12. Welke regisseurs treden op als cameo in tenminste 1 van hun films?
- A13. In welke films speelde Joop Braakhekke?
- A14. Welke regisseur heeft de meeste films met Jim Carrey in de hoofdrol geregisseerd?

- A15. Welke acteur of actrice speelt het meest in de slechtst gewaardeerde films?
- A16. In welk jaar tussen 1990 en nu zijn de meeste films met het woord 'beer' in de titel geproduceerd. En wat is het meest voorkomende genre?
- A17. Zijn er tussen 2010 en nu films uitgekomen waar de voornaam van een van de acteurs/actrices voorkomt in de naam van de film?
- A18. Wat is de kortste film met een waardering van 8.5 of hoger?
- A19. Welke film heeft de hoogste score met de minste stemmen?
- A20. Welke 3 acteurs/actrices hebben samen in de meeste films gespeeld?
- A21. Welke song is het vaakst gebruikt in de soundtrack van een film? Geef een lijstje van de films.
- A22. Zelfde vraag als 33, maar nu voor tv-series.
- A23. Hoeveel films heeft Woody Allen gemaakt? In hoeveel daarvan speelde hij zelf mee, en zijn er ook films waarin hij wel speelde maar niet regisseerde?
- A24. Geef een overzicht van personen die meer dan 1 functie vervulden bij een film. Dwz èn schrijver èn regisseur waren, of èn acteur èn producer, etc.

## Vragen Categorie B (visualisaties)

- B1. Maak een graaf (visueel), indien twee acteurs met elkaar in dezelfde film/serie hebben gespeeld maak je een verbinding (edge) tussen de twee acteurs. Visualiseer dit.
- B2. Voor bonus-punten: maak van vraag B1 een (dynamische) network-graph.
- B3. Geef het aantal films dat in een land gemaakt is weer in de tijd. Dwz maak een grafiek waarin op de x-as het jaar staat en op de y-as het aantal gemaakte films.
- B4. Voor bonuspunten: maak in vraag B3 het land variabel.
- B5. Maak een visuele presentatie van de in de loop van de tijd veranderende populariteit van onderwerpen (genre) van films. Denk daarbij aan horror, science-fiction, romance, geweld etc. Zoek zelf naar een goede manier om dit te presenteren.

- B6. Sommige films hebben een getal in hun titel (bijvoorbeeld: The Sixth Sense, Eight Days a week, Se7en). Maak een visuele weergave van het aantal malen dat een bepaald getal in een filmtitel voorkomt. Bonus: kijk of de Wet van Benford geldt.
- B7. Maak een kaart (b.v. google maps / openstreetview) met geboorteplaatsen van acteurs. Zodat op de kaart te zien is wie waar geboren is.
- B8. Maak een kaart (b.v. google maps / openstreetview) met landen waar een flim speelt. Zodat op de kaart te zien is waar de films spelen.
- B9. Maak een word-cloud met alle acteurs uit een land naar keuze. Dwz als je België kiest moet er een word-cloud met Belgische acteurs/actrices komen, met de acteur die in de meeste films gespeeld heeft het grootst. Kies je Nederland dan krijg je de Nederlandse acteurs etc.
- B10. Maak een visuele weergave van de kosten voor een film gerelateerd aan het aantal acteurs (m/v). Kijk ook of er verschil is tussen beroemde acteurs en onbekendere acteurs.

## Vragen Categorie C (R-modellen)

- C1. TV-series worden per seizoen gemaakt. Indien succesvol worden ze verlengd. Zoek uit welke factoren een rol spelen bij het bepalen hoeveel seizoenen van een tv-serie worden gemaakt. Maak hiervoor een lineair regressiemodel. Maak ook een grafische weergave hiervan.
- C2. Op IMDB staat onder meer een lijst van MPAA-ratings (de Amerikaanse versie van de Nederlandse Kijkwijzer). Maak een model waaruit je kan afleiden in welke categorie een film valt op grond van de motivatie in de MPAA-ratings lijst. Valideer het model.
- C3. Maak een model waaruit je de MPAA-categorie (zie ook vraag C2) afleidt uit de tekst van de plot (NB: hier wordt niet een grafische plot bedoeld, maar de plot van de film). Valideer het model.
- C4. Er wordt vaak gezegd dat naarmate actrices ouder zijn zij moeilijker een rol krijgen. Ga (met R) na of dat zo is. Je zou bijvoorbeeld een lineair regressiemodel kunnen maken met leeftijd en geslacht als onafhankelijke variabelen en aantal films per jaar als afhankelijke variabele.

C5. Ga na met R of er een verband is tussen de kosten van een film en de lengte ervan. Analyseer dit op een vergelijkbare manier als in vraag C4.

C6. Maak met logistic regression en/of cart-trees een model waaruit blijkt welke factoren leiden tot een succesvolle film. Maak eerst een goede beschrijving van wat succes is (en wat dus de afhankelijke variabele is) en bedenk een aantal onafhankelijke variabelen die je kunt gebruiken om een model te maken. Valideer ook het model.

C7. Er wordt wel eens gezegd dat Europese films meer 'cultureel' zijn dan Amerikaanse. Maak een model in R waarmee je kunt nagaan of Franse films minder vaak van het genre geweld en/of horror zijn dan Amerikaanse.