

De Rode Schoentjes

Eerste opdrachten voor data analyse met SQL en R.

Kees Velthuijs

februari 2025

Contents

1	Over dit document	3
1.1	Waarom data analyse?	3
1.2	Hoe te starten?	3
1.2.1	RStudio	3
1.2.2	Notebooks en data upload	4
1.3	Voorkennis	5
1.4	Howto's	5
1.5	Cheat sheets	5
1.5.1	R en nuttige packages	5
1.5.2	Markdown	5
1.5.3	SQL	6
1.6	Andere bronnen	6
1.6.1	R en nuttige packages	6
1.6.2	SQL	6
2	Voorbereiding	7
2.1	Benodigde bibliotheken	7
2.2	Inlezen van de bestanden	7
2.3	De data	8
2.3.1	Filiaal	8
2.3.2	Categorie	8
2.3.3	Budget	9
2.3.4	Kassa	10
2.4	Datum in de tabel kassa	10
2.5	Datum tabel	11

3	De eerste opdrachten	13
3.1	Het SELECT statement	13
3.2	De FROM clause	13
3.3	De SELECT clause	13
3.3.1	Kolomfiltering	13
3.3.2	Kolomalias	14
3.4	De WHERE clause	14
3.4.1	Filialen in London	15
3.4.2	BETWEEN clause	15
3.4.3	IN clause	16
3.4.4	LIKE clause	17
3.5	DISTINCT	18
3.6	Samenvatten	18
3.6.1	COUNT(*)	19
3.6.2	COUNT(DISTINCT)	19
3.6.3	SUM	20
3.6.4	MIN en MAX	20
3.7	ORDER BY clause	20
3.8	GROUP BY clause	20
3.9	HAVING clause	20
4	Meerdere tabellen	21

1 Over dit document

Dit PDF-document bevat allerlei materialen die bedoeld zijn om wegwijs te worden in het startsemester van Fontys ICT en dan specifiek het profiel ICT & Business. De focus van deze materialen ligt op het gebied van data analyse, tranformatie en visualisatie met de talen R en SQL.

Let op:

Dit document is op dit moment in de fase ontwerp. Het kan zijn dat het tijdens het semester er onderdelen gewijzigd worden en/of er aan worden toegevoegd. Mocht u als lezer verbeteringen zien, neem dan contact op met de schrijver.

Alvast bedankt!

1.1 Waarom data analyse?

[TODO: uitleg van de positie van data analyse binnen ICT & Business]

1.2 Hoe te starten?

Dit document is een werkboek dat je meer inzicht geeft in data analyse. Het gaat over de schoenwinkels van De Rode Schoentjes. Zij verkopen schoenen, onderhoudsmiddelen en accessoires in grote steden in Europa.

Het eerste hoofdstuk is een inleiding. Daarin wordt uitgelegd hoe het document is opgebouwd en wat het doel is van het document.

Hoofdstuk 2 is in feite al het begin van je eerste eigen data analyse. De gegevens worden klaargezet zodat je aan de slag kunt. Het is voor de start nog niet nodig dat je de inhoud van de code helemaal begrijpt. Als het goed is wordt dat later wel duidelijk.

Jouw werk zit in hoofdstuk 3 en volgende. Dan ga je aan de gang met SQL en R om de data analyse uit te voeren.

Zorg dat je toegang hebt tot Rstudio. In de volgende paragraaf vind je de informatie om je eigen omgeving in te richten. In de paragraaf Notebook wordt uitgelegd hoe je de

1.2.1 RStudio

Voor de data analyse gebruiken we R. Dat is een taal die speciaal voor data analyse is ontwikkeld. In R maak je gebruik van zogenaamde Notebooks. Dat zijn tekstbestanden die je met programma's als Notepad of Visual Studio Code kunt openen.

Het is slim om een ontwikkelomgeving (IDE) te gebruiken. Daarmee kun je niet alleen de tekst van het notebook veranderen, maar ook meteen de code uitvoeren.

Het is goed om te weten dat je verschillende keuzes hebt voor je ontwikkelomgeving:

- **RStudio Cloud gebruiken.**

Gebruik je incidenteel RStudio en heb je weinig behoefte om te installeren, wil je vooral even proberen: gebruik dan RStudio Cloud. Deze is gratis te gebruiken en geschikt voor proefstuderen en incidenteel gebruik. Je maakt hier zelf een account aan. Zie voor meer info de guide.

Eventuele bestanden moet je altijd eerst uploaden naar de cloud server!

****Let op**:**

Deze versie is slechts 15 uur per maand te gebruiken volgens de huidige voorwaarden, dat zal te weinig zijn om alle opdrachten in dit document uit te voeren!

- **RStudio server van FHICT gebruiken.**

Heb je behoefte aan wat meer mogelijkheden en loop je tegen de grenzen aan van de cloud versie? Dan hebben we binnen FHICT een eigen RStudio server. Dit kan handig zijn als je lokaal vast loopt, geen nieuw account op RStudio Cloud aan wil maken, of je eigen installatie niet om zeep wil helpen :). Ga naar de FHICT Rstudio server en log in met je **I account (nummer zonder @fontysict.nl)**. Het werkt vergelijkbaar met RStudio Cloud.

Eventuele bestanden moet je dus altijd eerst uploaden (naar de FHICT server)!

- **Installeer R & RStudio lokaal op je eigen machine.**

Dit is aan te raden als je zelf controle wilt over welke libraries je wilt kunnen installeren, je lokale bestanden makkelijk wilt kunnen gebruiken en als je met diverse databronnen wilt kunnen koppelen.

Ga je de verdieping ICT & Business doen, kies dan dit.

Je moet hier **2 stappen** voor doen, in deze volgorde:

1. Installeer de programmeertaal R.

Dit is de zogenaamde 'interpreter' die R code kan uitvoeren. Via deze site kom je bij de download pagina uit. Check vervolgens welk besturingssysteem je nodig hebt en pak de meest recente installer (.exe voor Windows of .pkg voor Mac)

2. Installeer de ontwikkelomgeving Rstudio.

Dit is de user interface waarin je handig je Notebooks (werkbestanden) kan bewerken, net zoals Visual Studio. Je kunt Rstudio hier vinden. Let op dat je de Free/Open Source versie pakt voor jouw besturingssysteem.

Let op: Je zult als je zelf installeert ook de juiste libraries moeten installeren! Zie deze HowTo.

Kom je er niet uit? Er zijn voldoende tutorials te vinden.

1.2.2 Notebooks en data upload

In R gebruiken we eigenlijk altijd zogenaamde Notebooks. Dat zijn .Rmd (RMarkdown) bestanden waar je uitleg, code (chunks) & output samen kunt voegen. Dit zorgt o.a. voor herhaalbaarheid en leesbaarheid van je stappen, een hele verbetering t.o.v. data analyse in MS Excel. Rstudio opent standaard niet met een Notebook. Dat moet je zelf uploaden of aanmaken als je aan de slag gaat.

Download uit canvas het bestand Rode Schoentjes - Data Analysis.rmd en het databestand Rode Schoentjes - Data.zip. Volg de instructies hieronder voor het openen van het notebook en beschikbaar krijgen van de data.

- **Online versie:**

Heb je de online versie van R-studio, kies rechtsonder in het scherm voor de tab *Files*. Klik op de knop *Upload*. Browse naar de map "downloads" en selecteer het Rmd-bestand. Als je op "ok" klikt, wordt het document ingeladen en zie je het tussen de bestanden staan. Klik erop en je dan wordt het zichtbaar in het scherm linksboven. Voor het ophalen van de data klik je nog een keer op de knop *Upload*. Nu open je het zip bestand Rode Schoentjes - Data.zip. R-studio zal het zip-bestand uitpakken en de databestanden los in het bestandsoverzicht zetten.

- **Lokale installatie:**

Als je R studio op je pc geïnstalleerd hebt, open je het bestand via het menu *File/Open*. Ga naar de map met downloads en open het Rmd-bestand. De databestanden zijn in dit geval beschikbaar in de map Downloads op je pc. Vraag hulp aan je docent om de bestanden op de goede plaats neer te zetten. Mogelijk moet de code van het notebook worden aangepast om de data succesvol in te lezen.

1.3 Voorkennis

Voordat je met de materialen op deze site aan de slag gaat, raden we ten eerste aan om een of meerdere kennismakingscursussen te doen op het gebied van R (Tidyverse en GGPlot) en/of SQL (tenzij je daar al ervaring mee hebt). Deze kun je vinden via o.a.:

- Datacamp (Vraag je docent voor studententoegang)

1.4 Howto's

Onze colleg Koen Suilen heeft een aantal veel voorkomende problemen op een rijtje gezet om je op weg te helpen. De links hieronder verwijzen naar zijn archief op github.io waar hij die toelicht:

- Gebruiken en installeren van R & RStudio Bevat ook nuttige tips als je de FHICT RStudio server gebruikt. Installeren hoeft niet tijdens de oriëntatie (mag wel).
- Gebruiken en installeren van libraries/packages in RStudio Bevat ook nuttige tips als je de FHICT RStudio server gebruikt. Installeren hoeft niet tijdens de oriëntatie (mag wel).
- Veranderen working directories & file management Voor de oriëntatie staat hier hoe je bestanden naar de RStudio server kunt uploaden, de rest is meer voor de verdieping.
- Data importeren (CSV/XLS/Database)
- Mock data genereren
- Exporteren naar HTML of PDF

1.5 Cheat sheets

Hieronder vind je links naar een aantal handige cheat sheets die je kunt gebruiken als reminder als je de opdrachten aan het uitvoeren bent.

- RStudio IDE over gebruik van de RStudio ontwikkel omgeving (door RStudio)

1.5.1 R en nuttige packages

- Tidyverse voor beginners (door Datacamp)
- Data transformatie met dplyr voor filteren, groeperen, sorteren, samenvatten, etc. met dplyr (door Rstudio)
- Data visualisatie met ggplot2 (door RStudio)
- Lubridate voor het werken met datums en tijden (door RStudio)
- R short reference card met van van alles een beetje (door Tom Short)

1.5.2 Markdown

- R Markdown met op de eerste pagina dingen over hoe en wat in de RStudio IDE en op de tweede pagina syntax en voorbeelden (door RStudio)
- Markdown guide oke, niet echt een cheatsheet maar wel een handig overzicht met wat nuttige do's en don'ts

1.5.3 SQL

- SQL cheat sheet cheat sheet die heel veel verder gaat dan wat nodig is voor het startsemester (door sqltutorial.org)

1.6 Andere bronnen

Uiteraard heeft Datacamp allerlei cursussen die je kunt volgen. Zie Canvas voor de links naar de cursussen en hoofdstukken of vraag je docent.

1.6.1 R en nuttige packages

- R seek zoekmachine die specifiek zoekt op domeinen waarvan bekend is dat er content over R op staat (daar waar je via Bing, Google of een andere algemene zoekmachine nog wel eens op hele andere programmeertalen uit komt)
- R for datascience boek over R geschreven in R Markdown (door Hadley Wickham en Garrett Golemund, met bijdragen van vele anderen voor de online versie)
- ggplot2 essentials een website waar ze veel verschillende soorten grafieken laten zien, maar ook tips om je grafieken leesbaarder te maken met legenda, titels, etc.

1.6.2 SQL

- SQL leren een hele basic website waarin SQL uitgelegd wordt in het Nederlands (door Donny Semeleer)
- FHICT SQL tool tool waarin je SQL statements uit kunt proberen als je een Fontys ICT i-account hebt.

2 Voorbereiding

2.1 Benodigde bibliotheken

Laad de nodige bibliotheken voor de verwerking van de data:

- tidy
- tidyverse
- dplyr
- ggplot2
- readxl
- RSQLite
- lubridate
- tsibble

2.2 Inlezen van de bestanden

De data bestaat uit verkoopgegevens van de winkels van De Rode Schoentjes in Nederland, België, Duitsland, Denemarken, Frankrijk en het Verenigd Koninkrijk. De data bestaat uit vier Excel sheets die je hebt gedownload in één zip-bestand en hebt ingelezen in R-studio. In je r-studio app zou je die in het scherm rechtsonder moeten kunnen zien als je de tab **Files** hebt openstaan.

Bestand	Omschrijving
filiaal.xlsx	Locatie van de filialen
categorie.xlsx	De naam van de omzetgroepen van het assortiment
budget.xlsx	De verkoopprognose per filiaal per week
kassa.xlsx	De verkoop per categorie per kassa per filiaal en dag

2.3 De data

2.3.1 Filiaal

De table Filiaal bevat de gegevens van alle winkels van De Rode Schoentjes. De winkels zijn gevestigd in verschillende Europese landen, zoals Nederland, België, Frankrijk, Duitsland, Denemarken en het Verenigd Koninkrijk.

In de tabel zijn de volgende gegevens aanwezig:

Kolom	Betekenis
ID	Een unieke code voor de locatie.
Locatie	De naam voor de locatie waar de winkel is.
Adres	Straatnaam en huisnummer van de winkel.
Plaats	Plaats waar de winkel gevestigd is.
Land	Landcode van het land waar de winkel gevestigd is.
Manager	De naam van de manager van het filiaal
Email	Het e-mailadres voor berichten naar het filiaal

Hierbij een overzicht van hoe enkele locaties zijn opgeslagen in de tabel "filiaal".

Table 3: Displaying records 1 - 10

ID	Locatie	Adres	Plaats	Land	Manager	Email
1	Hoog Catharijne	Godebaldkwartier 54	Utrecht	NL	Tieneke van Dam	t.vandam@derodeschoentjes.nl
2	Winkelcentrum Kronenburg	Kronenburgpassage 1	Amsterdam	NL	Jolanda Versteijnen-De Laat	j.versteijnen-delaat@derodeschoentjes.nl
3	Kalverstraat	Kalverstraat 92	Amsterdam	NL	Inge de Jong-Broerse	inge.dejong@derodeschoentjes.nl
4	Heuvel Galerie	Heuvel Galerie 122	Eindhoven	NL	Johan de Boer	j.deboer@derodeschoentjes.nl
5	Herestraat	Herestraat 23	Groningen	NL	Babette Waardenburg	babs.waardenburg@derodeschoentjes.nl
6	Winkelcentrum Zuidplein	Zuidplein 68	Rotterdam	NL	Ben Klaassen	ben.klaassen@derodeschoentjes.nl
7	Arena	Arena 16	Den Bosch	NL	Zara de Winter-Schellens	zara.dewinter@derodeschoentjes.nl
8	De Barones	Barones 30	Breda	NL	Lukas Nguyen	lukas.nguyen@derodeschoentjes.nl
9	Grote Staat	Grote Staat 45	Maastricht	NL	Aisha Mahmoud-Peters	a.mahmoud-peters@derodeschoentjes.nl
10	Designer Outlet Roermond	Stadsweide 2	Roermond	NL	Bram Willems	b.willems@derodeschoentjes.nl

2.3.2 Categorie

Alle producten in de winkel zijn ingedeeld in tien categorieën. Alle winkels voeren in principe alle producten in dezelfde categorieën. Op die manier zijn de verkopen van de winkels makkelijker met elkaar te vergelijken.

Aangezien er tien categorieën zijn, krijg je hieronder een overzicht de hele tabel “categorie”.

Table 4: 9 records

Kolom	Betekenis
ID	Een unieke code voor de categorie.
Omschrijving	Een omschrijving voor de categorie.

ID	OMSCHRIJVING
1	Damesschoenen
2	Herenschoenen
3	Kinderschoenen
4	Schoenonderhoud
5	Accessoires
6	Sportschoenen
7	Seizoensproducten
8	Werkschoenen
9	Orthopedische schoenen

2.3.3 Budget

Aan het begin van elk kwartaal wordt op het hoofdkantoor een prognose gemaakt van de omzet in de komende maanden. Dit budget is de doelstelling voor de filialen en bevat de verwachte omzet per week voor de filialen.

De tabel bevat de volgende gegevens:

Kolom	Betekenis
ID	Een unieke code voor een budgetregel.
<i>FiliaalID</i>	Een verwijzing naar het filiaalnummer in de tabel Filiaal .
Jaar	Het nummer van het jaar waarvoor het budget geldt.
Week	Het nummer van de week waarvoor het budget geldt.
Budget	De omzetdoelstelling voor het filiaal in de aangegeven periode.

Om een indruk te krijgen van de gegevens in de tabel budget zijn hieronder een aantal regels uit de tabel “budget” afgedrukt:

Table 7: Displaying records 1 - 10

ID	FiliaalID	Jaar	Week	Budget
1	1	2025	1	14700
2	2	2025	1	18100
3	3	2025	1	18900
4	4	2025	1	17700
5	5	2025	1	15900
6	6	2025	1	18900
7	7	2025	1	18900
8	8	2025	1	15900

ID	FiliaalID	Jaar	Week	Budget
9	9	2025	1	15700
10	10	2025	1	15200

2.3.4 Kassa

Alle verkopen in het filiaal worden via de kassa afgerekend. Aan het einde van de dag worden de kassa's afgesloten. Bij dat proces wordt de omzet per categorie opgeslagen. Deze informatie wordt verzameld op het hoofdkantoor. Daardoor ontstaat een tabel met de volgende kolommen:

Kolom	Betekenis
ID	Een unieke code voor een omzetregel in de kassa.
<i>FiliaalID</i>	De verwijzing naar het filiaalnummer in de tabel Filiaal .
KassaID	Een uniek nummer voor een kassa binnen het filiaal.
<i>CategorieID</i>	Het nummer van een omzetcategorie in de tabel categorie .
Bedrag	De omzet voor deze categorie voor deze kassa in het filiaal.

Hieronder krijg je een indruk van de gegevens die in de tabel “kassa” worden opgeslagen:

Table 9: Displaying records 1 - 10

ID	FiliaalID	KassaID	Datum	CategorieID	Bedrag
1	1	1	1735430400	1	746.19
2	1	1	1735430400	2	430.94
3	1	1	1735430400	3	528.11
4	1	1	1735430400	4	262.67
5	1	1	1735430400	5	490.51
6	1	1	1735430400	6	391.97
7	1	1	1735430400	7	135.39
8	1	1	1735430400	8	65.19
9	1	1	1735430400	9	53.02
10	2	1	1735430400	1	1027.04

2.4 Datum in de tabel kassa

De datums worden in SQLite bij de import vanuit Excel opgeslagen als nummers. Daarbij wordt het 'unixepoch' berekening toegepast. Om van dit numerieke gegeven te vertalen naar een datum in de vorm van jaar, maand en jaar is een aparte functie nodig. In het onderstaande voorbeeld is aan de data in van de eerste 10 regels van de tabel kassa de datum toegevoegd.

```
SELECT *
  , date(datum, 'unixepoch') as 'berekenende datum'
FROM Kassa
```

Table 10: Displaying records 1 - 10

ID	FiliaalID	KassaID	Datum	CategorieID	Bedrag	berekende datum
1	1	1	1735430400	1	746.19	2024-12-29
2	1	1	1735430400	2	430.94	2024-12-29
3	1	1	1735430400	3	528.11	2024-12-29
4	1	1	1735430400	4	262.67	2024-12-29
5	1	1	1735430400	5	490.51	2024-12-29
6	1	1	1735430400	6	391.97	2024-12-29
7	1	1	1735430400	7	135.39	2024-12-29
8	1	1	1735430400	8	65.19	2024-12-29
9	1	1	1735430400	9	53.02	2024-12-29
10	2	1	1735430400	1	1027.04	2024-12-29

2.5 Datum tabel

De tabel kassa heeft alleen de gegevens van de dagen waarop er in de winkel verkocht is. Als de winkel gesloten is, is er geen verkoop. Die datum komt in de tabel kassa dan niet voor. Zo kun je nooit zien wanneer er niets verkocht is.

In Europa houden we de maandag aan als eerste dag van de week. Dat in tegenstelling tot de Verenigde Staten. Dat lijkt misschien niet belangrijk, maar wel om de eerste week van het jaar te bepalen. Week 1 is de eerste week waarin er vier dagen in het nieuwe jaar vallen. Als Nieuwjaar op een donderdag valt is dat in de Verenigde Staten week 53, maar in Europa week 1.

Daarom is het handig om een tabel te hebben met alle datums, het weeknummer, de maand en het jaar. Die moet je niet in Excel maken, maar laten genereren uit code. Hierbij het script:

Dit is hoe de gegenereerde datumtabel er uitziet:

Table 11: Displaying records 1 - 10

Datum	Wdag	Wd	WdNm	WkNr	JrWk	MndNr	Mnd	MndNm	Mdag	Qrt	JrNr	Jdag
2024-12-29	1	zo	zondag	52	24/52	12	dec	december	29	24Q4	2024	364
2024-12-30	2	ma	maandag	1	25/01	12	dec	december	30	24Q4	2024	365
2024-12-31	3	di	dinsdag	1	25/01	12	dec	december	31	24Q4	2024	366
2025-01-01	4	wo	woensdag	1	25/01	1	jan	januari	1	25Q1	2025	1
2025-01-02	5	do	donderdag	1	25/01	1	jan	januari	2	25Q1	2025	2
2025-01-03	6	vr	vrijdag	1	25/01	1	jan	januari	3	25Q1	2025	3
2025-01-04	7	za	zaterdag	1	25/01	1	jan	januari	4	25Q1	2025	4
2025-01-05	1	zo	zondag	1	25/01	1	jan	januari	5	25Q1	2025	5
2025-01-06	2	ma	maandag	2	25/02	1	jan	januari	6	25Q1	2025	6

Datum	Wdag	Wd	WdNm	WkNr	JrWk	MndNr	Mnd	MndNm	Mdag	Qrt	JrNr	Jdag
2025-01-07	3	di	dinsdag	2	25/02	1	jan	januari	7	25Q1	2025	7

3 De eerste opdrachten

Aanvangssituatie:

Je hebt de ontwikkelomgeving van Rstudio open staan. Het document Rode Schoentjes - Data Analyse.rmd is geopend in het scherm rechtsboven. Rechtsonder, in het overzicht met **Files**, zijn de Excel bestanden budget.xlsx, categorie.xlsx, filiaal.xlsx en kassa.xlsx zichtbaar.

Mocht dat niet het geval zijn, kijk dan in hoofdstuk 1 na hoe deze situatie gecreëerd kan worden of neem contact op met uw docent.

Bij deze eerste opdrachten wordt telkens maar gebruik gemaakt van één tabel waaruit de gegevens worden gehaald. Op basis daarvan worden de methodes voor selecteren en samenvatten van gegevens uitgelegd. Aan de hand van de opdrachten doe je ervaring op in het toepassen van de onderdelen.

3.1 Het SELECT statement

Met SQL kun je gegevens uit een database ophalen. Daarbij wordt een vast template gebruikt. De meeste databases, MS SQL, MYSQL, Progress, Oracle, Postgres, ondersteunen een vorm van SQL. Dat is handig, want zo kun je altijd zien wat er in een tabel zit.

In basis heeft het SELECT statement heeft een template met componenten in een vaste volgorde. We beginnen eenvoudig en zullen elke keer een onderdeel aan het template toevoegen.

3.2 De FROM clause

Het eerste onderdeel van het SELECT statement is de FROM clause. Achter FROM staat de naam van de tabel die gebruikt wordt. Leer jezelf aan dat je achter FROM altijd maar één tabelnaam vermeldt. Later zul je leren hoe je desondanks gegevens uit meer tabellen kunt opvragen.

Laten we beginnen met een overzicht zien van alle productcategorieën.

Table 12: 9 records

ID	OMSCHRIJVING
1	Damesschoenen
2	Herenschoenen
3	Kinderschoenen
4	Schoenonderhoud
5	Accessoires
6	Sportschoenen
7	Seizoensproducten
8	Werkschoenen
9	Orthopedische schoenen

3.3 De SELECT clause

3.3.1 Kolomfiltering

Achter SELECT staan de namen van kolommen die je in de uitvoer wilt zien. Als je alle kolommen wilt zien gebruik je een sterretje.

Het gebruik van de `SELECT` clause is een vorm van *kolomfiltering* of *verticale filtering*. Door daar iets in te voeren, zijn niet alle kolommen zichtbaar.

Vervang het sterretje (*) na `SELECT` zo dat van de filialen alleen het id, de naam van de manager en het e-mailadres van de manager wordt getoond.

Table 13: Displaying records 1 - 10

ID	Manager	Email
1	Tieneke van Dam	t.vandam@derodeschoentjes.nl
2	Jolanda Versteijnen-De Laat	j.versteijnen-delaat@derodeschoentjes.nl
3	Inge de Jong-Broerse	inge.dejong@derodeschoentjes.nl
4	Johan de Boer	j.deboer@derodeschoentjes.nl
5	Babette Waardenburg	babs.waardenburg@derodeschoentjes.nl
6	Ben Klaassen	ben.klaassen@derodeschoentjes.nl
7	Zara de Winter-Schellens	zara.dewinter@derodeschoentjes.nl
8	Lukas Nguyen	lukas.nguyen@derodeschoentjes.nl
9	Aisha Mahmoud-Peters	a.mahmoud-peters@derodeschoentjes.nl
10	Bram Willems	b.willems@derodeschoentjes.nl

3.3.2 Kolomaliases

Een kolomnaam `ID` is natuurlijk niet zo mooi. We kunnen de kolomnaam vervangen door een zogenaamde *alias*. Achter de naam van de kolom zet je dan `AS` met een nieuwe naam: de alias.

Hieronder een voorbeeld voor de tabel `categorie`. De kolomnaam `ID` is vervangen door `Code`.

Het is ook mogelijk om in de alias een spatie op te nemen. Daarvoor plaats je de alias tussen blokhaken `[]`. In het onderstaande voorbeeld is dat gedaan met de "Categorie omschrijving".

```
SELECT id AS Code
       , omschrijving as [Categorie omschrijving]
FROM categorie
```

Table 14: 9 records

Code	Categorie omschrijving
1	Damesschoenen
2	Herenschoenen
3	Kinderschoenen
4	Schoenonderhoud
5	Accessoires
6	Sportschoenen
7	Seizoensproducten
8	Werkschoenen
9	Orthopedische schoenen

3.4 De WHERE clause

Met de `WHERE` clause geef je een voorwaarde waaraan de regel moet voldoen. Daarmee beperk je het aantal regels dat wordt getoond. Om die reden wordt dit ook wel *regelfiltering* of *horizontale filtering* genoemd.

3.4.1 Filialen in London

Selecteer uit de tabel `filiaal` de locatie en het adres van de locaties in Londen.

Table 15: 2 records

Locatie	Adres
Oxford Street	252 Oxford Street
Westfield London	Ariel Way

Selecteer uit de tabel `filiaal` de namen van de locaties en managers van filialen in België en Frankrijk.

```
SELECT locatie  
      , manager  
FROM filiaal  
WHERE Land = 'BE' OR Land = 'FR'
```

Table 16: Displaying records 1 - 10

Locatie	Manager
Meir Shopping	Thomas De Smet
Nieuwstraat Brussel	Nathalie Dujardin
Shopping Gent Zuid	Karim El Haddadi
Wijnegem Shopping Center	Sophie Dubois
Waasland Shopping	Luca Moretti
Mediacité Shopping	Fatima Ben Saïd-Bakkali
Ring Shopping Kortrijk	Elise Fontaine
Les Grands Prés	Julie Verhoeven-Lambert
Les Quatre Temps	Sofia Almeida-Rousseau
La Part-Dieu	Jean-Baptiste Chevalier

3.4.2 BETWEEN clause

In de tabel `kassa` staat de werkelijke omzet van een kassa in een filiaal. Selecteer het filiaalnummer, kassanummer, categorie, de datum en de omzet. Doe dat voor de categorieën seizoensproducten (7) en orthopedische schoenen (9). Laat alleen de regels zien voor de filialen in Denemarken (filiaal 27 tot en met 30) van 8 tot 9 april 2025.

Om de datum uit `kassa` te vertalen naar een “echte” datum gebruik je de tekst `date(datum, 'unixepoch')`.

Table 17: 8 records

FiliaalID	KassaID	CategorieID	Datum	Bedrag
27	1	7	2025-04-08	163.77
29	1	7	2025-04-08	91.54
29	1	9	2025-04-08	128.43
30	1	9	2025-04-08	67.01
27	1	9	2025-04-09	192.24
28	1	7	2025-04-09	100.09

FiliaalID	KassaID	CategorieID	Datum	Bedrag
29	1	7	2025-04-09	164.92
30	1	9	2025-04-09	56.02

In de **WHERE** clause heb je voor het selecteren van de categorieën **OR** nodig, maar voor de datums en de filialen **AND**. Dat maakt het opstellen van een goede **WHERE** clause best ingewikkeld.

Door het gebruik van **BETWEEN** wordt het eenvoudiger. Zie hieronder de overzichtelijk. Vooral als je nu de categorie zou uitbreiden

Table 18: 6 records

FiliaalID	KassaID	CategorieID	Datum	Bedrag
29	1	7	2025-06-06	180.97
27	1	7	2025-06-07	52.16
28	1	7	2025-06-07	166.13
28	1	9	2025-06-07	36.69
30	1	7	2025-06-07	125.94
30	1	9	2025-06-07	111.30

Voor het bepalen van de omzet is het belangrijk om te weten welke weeknummers de omzet van april 2025 bepalen. Kun jij dat opvragen uit de tabel **datum**? De naam van de maand staat in de kolom **mondnm**.

Table 19: 5 records

WkNr
14
15
16
17
18

3.4.3 IN clause

Selecteer uit de tabel **Budget** de gewenste omzet in week 1 en 5 van filialen in Brabant. Dat zijn de filialen 4, 7 en 8. Laat alle kolommen van de tabel **Budget** zien.

Table 20: 6 records

ID	FiliaalID	Jaar	Week	Budget
4	4	2025	1	17700
7	7	2025	1	18900
8	8	2025	1	15900
168	4	2025	5	19300
171	7	2025	5	19700
172	8	2025	5	16800

Je merkt dat het nog niet zo eenvoudig is. Als je verschillende waardes wilt filteren in een kolom kun je gebruik maken van de IN clause. De uitkomst blijft natuurlijk hetzelfde, maar het SELECT statement oogt daardoor een stuk overzichtelijker.

```
SELECT *
FROM Budget
WHERE Week IN (1, 5) AND FiliaalID IN (4, 7, 8)
```

Table 21: 6 records

ID	FiliaalID	Jaar	Week	Budget
4	4	2025	1	17700
7	7	2025	1	18900
8	8	2025	1	15900
168	4	2025	5	19300
171	7	2025	5	19700
172	8	2025	5	16800

Kun je met gebruik van de IN clause een overzicht maken van de locaties van filialen in Denemarken en Frankrijk?

Table 22: 8 records

Locatie
Strøget Kopenhagen
Bruun's Galleri
Fisketorvet Shopping
Rosengårdcentret
Les Quatre Temps
La Part-Dieu
Centre Commercial Cap 3000
Polygone Riviera

3.4.4 LIKE clause

In de vorige paragrafen hebben we aandacht gehad voor de filtering. Je ziet dat er categorieën zijn die verwijzen naar schoenen en andere productgroepen. Kun je een overzicht maken met alle categorieën waar in de omschrijving de tekst "schoenen" voorkomt?

```
SELECT *
FROM categorie
WHERE omschrijving like '%schoenen%'
```

Table 23: 6 records

ID	OMSCHRIJVING
1	Damesschoenen

ID	OMSCHRIJVING
2	Herenschoenen
3	Kinderschoenen
6	Sportschoenen
8	Werkschoenen
9	Orthopedische schoenen

3.5 DISTINCT

Bij het opvragen van gegevens krijg je in principe de kolomgegevens van alle regels die in de tabel gevonden zijn. Er is een mogelijkheid om alleen unieke gegevens te krijgen door achter **SELECT** de term **DISTINCT** toe te voegen.

Hierbij een voorbeeld om alle unieke w van alle filialen op te halen. Hoewel er 2 filialen in Londen zijn, komt de plaatsnaam Londen nu maar 1x voor.

```
SELECT DISTINCT Plaats
FROM filiaal
WHERE Land = 'UK'
```

Table 24: 6 records

Plaats
Londen
Birmingham
Manchester
Glasgow
Sheffield
Cardiff

3.6 Samenvatten

Tot nu toe hebben we de gegevens van kolommen opgevraagd. Het is ook mogelijk om kolomgegevens samen te vatten. Bij het samenvatten wordt een berekening uitgevoerd op de kolomwaarden die geselecteerd zijn uit de tabel achter **FROM** en gefilterd zijn met de **WHERE** clause. Om samen te vatten voegen we zogenaamde *aggreate phrases* toe achter **SELECT**.

Er zijn verschillende functies voor het samenvatten beschikbaar in SQL:

Functie	Omschrijving
COUNT()	Tel het aantal kolomwaarden.
SUM()	Bereken het totaal van de kolomwaarden.
MIN()	Bepaal de minimum van de kolomwaarden.
MAX()	Bepaal het maximum van de kolomwaarden.
AVG()	Bepaal het gemiddelde van de kolomwaarden.

3.6.1 COUNT(*)

Telt het aantal kolomwaarden. Kijk hieronder hoe het totaal aantal budgetregels van filiaal 10 wordt geteld.

```
SELECT COUNT(*)
FROM Budget
WHERE FiliaalID = 10
```

Table 26: 1 records

COUNT(*)
27

3.6.2 COUNT(DISTINCT)

Binnen de functie COUNT() kun je te tekst DISTINCT toevoegen. Daarmee worden alle unieke waarden geteld. Als een kolomwaarde meer dan 1x voorkomt, wordt die maar één keer geteld. Kijk hieronder naar het aantal kassa's dat omzet had op Eerste Paasdag (18 april 2025).

```
SELECT COUNT (DISTINCT FiliaalID)
FROM kassa
WHERE date(datum, 'unixepoch') = '2025-04-18'
```

Table 27: 1 records

COUNT (DISTINCT FiliaalID)
17

Ga nu zelf aan de slag en tel het aantal keren dat de omzet aan de kassa voor de productcategorie werkschoenen (8) minder was dan € 10.

Table 28: 1 records

COUNT(*)
173

Tel het aantal unieke productcategorieën waarvan de omzet aan de kassa hoger was dan € 500. Controleer daarna de uitkomst.

Table 29: 1 records

COUNT(DISTINCT CategorieID)
3

Hoe vaak was de omzet aan de kassa van een productcategorie aan het einde van een dag minder dan € 10? Controleer de uitkomst

Table 30: 1 records

COUNT(CategorielD)
694

Bij hoe veel productcategorieën is het voorgekomen dat de omzet aan de kassa aan het einde van een dag minder dan € 10 was? Controleer de uitkomst

Table 31: 1 records

COUNT(DISTINCT CategorielD)
4

3.6.3 SUM

De andere vorm van het samenvatten van kolomwaarden is het optellen van waarde. Kijk of je met SUM de totale omzet van het filiaal in Den Bosch (7) over april kunt berekenen.

Table 32: 1 records

SUM(Bedrag)
78513.9

Wat was het budget van het filiaal over april 2025 als je weet dat deze periode week 14 tot en met 18 is?

Table 33: 1 records

SUM(Budget)
93100

3.6.4 MIN en MAX

[TODO: UITLEG EN OPDRACHTEN TOEVOEGEN]

3.7 ORDER BY clause

[TODO: UITLEG EN OPDRACHTEN TOEVOEGEN]

3.8 GROUP BY clause

[TODO: UITLEG EN OPDRACHTEN TOEVOEGEN]

3.9 HAVING clause

[TODO: UITLEG EN OPDRACHTEN TOEVOEGEN]

4 Meerdere tabellen

Tot nu toe hebben we bij het opvragen van gegevens we ons beperkt tot één tabel, maar het is ook mogelijk om gegevens op te vragen vanuit meerdere tabellen.

Voor het koppelen van gegevens van de ene tabel aan de andere gebruiken we gegevens die in beide tabellen aanwezig zijn. In de data die je hebt gekregen zijn dit de verbindingen:

tabel kolom	Verwijst naar
FiliaalID in tabel kassa	ID in tabel filiaal
CategorieID in tabel kassa	ID in tabel categorie
Datum in tabel kassa	Datum in tabel datumtabel
FiliaalID in tabel budget	ID in tabel filiaal

De linke tussen de tabellen budget en datumtabel is lastiger omdat er niet één kolom de verbinding maakt, maar twee. De kolommen Jaar en Week in budget verwijzen naar JrNr en WkNr in de tabel budget.

Voor het gebruiken van meer dan één tabel voor je vraag, voeg je extra informatie toe tussen de FROM en de WHERE clause. In eerste instantie maken we gebruik van de INNER JOIN clause.

Achter INNER JOIN voeg je de naam van de extra tabel toe. Daarna geef je achter ON de verbinding aan. Let op dat als je meer dan één tabel gebruikt je voor alle kolommen aan moet geven uit welke tabel de informatie moet komen.

Hieronder het voorbeeld voor het opvragen van de locatie, de manager en het budget voor week 2 van alle filialen in Frankrijk.

```
SELECT Filiaal.Locatie
      , Filiaal.Manager
      , budget.Budget
FROM Filiaal
INNER JOIN Budget
  ON Budget.FiliaalID = Filiaal.ID
WHERE Budget.Week = 2
      AND Filiaal.Land = 'FR'
```

Table 35: 4 records

Locatie	Manager	Budget
Les Quatre Temps	Sofia Almeida-Rousseau	18100
La Part-Dieu	Jean-Baptiste Chevalier	16200
Centre Commercial Cap 3000	Élise Moreau	16100
Polygone Riviera	Camille Dupont-Leclerc	16800

Vraag nu zelf de kassagegevens (jaarweek, datum, filiaal, kassa en bedrag) op van de maandag van week 24 van 2025. Gebruik daarvoor de datumtabel en de tabel kassa.

LET OP:

Om de koppeling te leggen moet je de datum uit te tabel kassa omzetten naar tekst met de functie `date(kassa.datum, 'unixepoch')`. De datum in de tabel datumtabel is al als tekst opgeslagen.

Table 36: Displaying records 1 - 10

JrWk	Datum	FiliaalID	KassaID	Bedrag
25/24	2025-06-09	31	1	699.02
25/24	2025-06-09	31	1	498.48
25/24	2025-06-09	31	1	357.20
25/24	2025-06-09	31	1	184.03
25/24	2025-06-09	31	1	400.45
25/24	2025-06-09	31	1	192.83
25/24	2025-06-09	31	1	193.68
25/24	2025-06-09	31	1	82.31
25/24	2025-06-09	32	1	1134.32
25/24	2025-06-09	32	1	682.88