

Template R notebook met SQL code

++++++
Instructies: + + Onderstaande is nodig om je notebook werkende te krijgen. + + Bij het opstarten van je notebook dien je de chunk van regel 28 t/m 40 en regel 42 t/m 46 te runnen. + + Daarna spring je naar regel 76 waar je je eerste SQL-chunk kan maken. + + +
++++++

Dit is een voorbeeld van hoe je aan de slag zou kunnen gaan met R chunks in een Notebook. Hierin ga je: + opmaak toepassen + functies uit Databcamp toepassen op een “eigen” dataset + oefenen met “knitten” Er is met zogenaamde Markdown allerlei opmaak toegepast op de tekst rondom de chunks. Dit is anders dan hoe het in Word gaat (daar is het meer What You See Is What You Get), het is wel vergelijkbaar met de opmaak van pagina's in Wikipedia.

Importeren libraries en dataset

Hieronder wordt een library geladen waar functies in zitten die we verderop in het Notebook gaan gebruiken: tidyverse. Daarnaast wordt hier de dataset geïmporteerd zodat we daar vragen aan kunnen stellen.

Setup

```
UPDATE dataset SET besteldatum = DATE(besteldatum, 'unixepoch'), leverdatum = DATE(leverdatum, 'unixepoch'), betaaldatum = DATE(betaaldatum, 'unixepoch');
```

Als je bovenstaande code uitvoert door op het groene pijltje rechtsbovenin de chunk te drukken, gebeurt er nog vrij weinig. Als het goed is verschijnt er alleen even een groen balkje naast de code regels. Tenminste, als je het Excel bestand op de juiste plek hebt geüpload. Het laden van de libraries hoeven we maar 1 keer te doen en het is wel zo overzichtelijk om dat allemaal aan het begin van je document te doen voor je de inhoud in duikt.

De dataset is nu opgeslagen in een database waarvoor we een connectie hebben gemaakt die 'con' heet. Elke keer als we dus gegevens uit die database willen halen zullen we dus eerst die connectie moeten gaan noemen.

eerste SQL-chunk

In de onderstaande chunk is de meest basale vorm van een query opgesteld. Je kan deze laten uitvoeren door op de groene knop te drukken of door de query-code te selecteren en op *Ctrl+Shift+Enter* te drukken. Merk ook de eerste regel op waarin 'connection=con' staat. Deze is nodig om contact met de database te maken.

```
select *  
from dataset
```

Displaying records 1 - 10

regioCode	regiomgr	mgrnm	regelOmzet	bestelnr	verkoper	verknm	besteldatum	leverdatum	betaaldatum	klantnr	naam	postcodehuisnr
GVGH	806	Harrie	146.25	1001	805	Karel	2012-01-06	2012-01-09	2012-01-18	15	Boer, de	5463ZK1003
GVGH	806	Karel	52.50	1004	806	Karel	2012-01-14	2012-01-20	2012-01-24	19	Zuur	5463KK103
GVGH	806	Karel	473.00	1014	806	Karel	2012-01-27	2012-01-31	2012-02-11	19	Zuur	5463KK103
GVGH	806	Karel	80.55	1014	806	Karel	2012-01-27	2012-01-31	2012-02-11	19	Zuur	5463KK103
GVGH	806	Karel	373.75	1014	806	Karel	2012-01-27	2012-01-31	2012-02-11	19	Zuur	5463KK103
GVGH	806	Harrie	97.50	1016	805	Karel	2012-02-04	2012-02-06	2012-02-18	19	Zuur	5463KK103
GVGH	806	Bertha	207.00	1017	804	Karel	2012-02-04	2012-02-05	2012-02-15	14	Kraymans	5472ZK103
GVGH	806	Bertha	125.30	1017	804	Karel	2012-02-04	2012-02-05	2012-02-15	14	Kraymans	5472ZK103
GVGH	806	Harrie	162.50	1019	805	Karel	2012-02-05	2012-02-08	2012-02-14	15	Boer, de	5463ZK1003
GVGH	806	Herman	172.00	1023	801	Karel	2012-02-10	2012-02-11	2012-02-19	19	Zuur	5463KK103

++++++
+ Opdracht + + Bouw een aantal SQL-queries die interessante informatie opvragen uit de data. Je kan denken aan vragen zoals + + - Hoeveel verschillende verkopers zijn er? + + - Hoeveel verkopers per regio zijn er? + + + + Maak zo meer SQL-queries op de gegeven dataset (of een

eigen dataset) waarin gebruik gemaakt is van onder andere + + WHERE, ORDER BY, GROUP BY, HAVING en subqueries; + + +
+++++

```
SELECT
  omschrijving as 'Most popular item',
  COUNT(omschrijving) AS "Amount of distinct purchases"

FROM
  dataset

GROUP BY
  omschrijving

ORDER BY
  "Amount of distinct purchases" DESC

LIMIT 1;
```

1 records

Most popular item	Amount of distinct purchases
Magere Kwark	56

Nieuwe query's kan je aanroepen door in de balk van dit subscherm op het pijltje naast het groene C-knopje te drukken en te kiezen voor SQL. Vergeet niet de connectie 'con' heet en daar aan moet linken, net zoals de bovenstaande query's.

```
SELECT
  woonplaats as 'Most popular city',
  COUNT(woonplaats) AS "Amount of times mentioned"

FROM
  dataset

GROUP BY
  woonplaats

ORDER BY
  "Amount of times mentioned" DESC

LIMIT 1;
```

1 records

Most popular city	Amount of times mentioned
Eindhoven	159

De resultaten van het notebook kan je opslaan als een HTML-pagina. Om dat te doen klik je in de bovenstaande balk op de pijl naast *Preview* en selecteer je *Knit to HTML*. Een andere optie is om *Ctrl+Shift+K* in te drukken om een HTML bestand te krijgen.