

Opdracht 1.

Column Profiling

Completeness / Fill Rates, Data Type, Size/Lengths
Unique Values, Frequency, Patterns, MIN/MAX

Als tester heeft u de opdracht gekregen om testen te definiëren en u heeft ervoor gekozen om "pairwise testing" toe te passen. Daarom bent u op zoek naar alle mogelijke waarden voor het begrip "geslacht".

De aangeleverde specificaties geven aan dat "geslacht" kan worden gevuld met de volgende waarden:

Omschrijving	Waarde
Male	M
Female	F
Unknown	U

U heeft ook de beschikking over een set (gepseudonimiseerde) productiegegevens, het betreft de database AdventureWorks2012_bron, tabel HumanResources.Employee en kolom Gender in het bijzonder.

Taak 1a.

Realiseer met de beschikbare omgeving via SQL een selectie op de gegevens waarmee u kunt vaststellen welke waarden de kolom Gender daadwerkelijk bevat. Eventueel met de spreiding (aantallen) per waarde.

Uitwerking

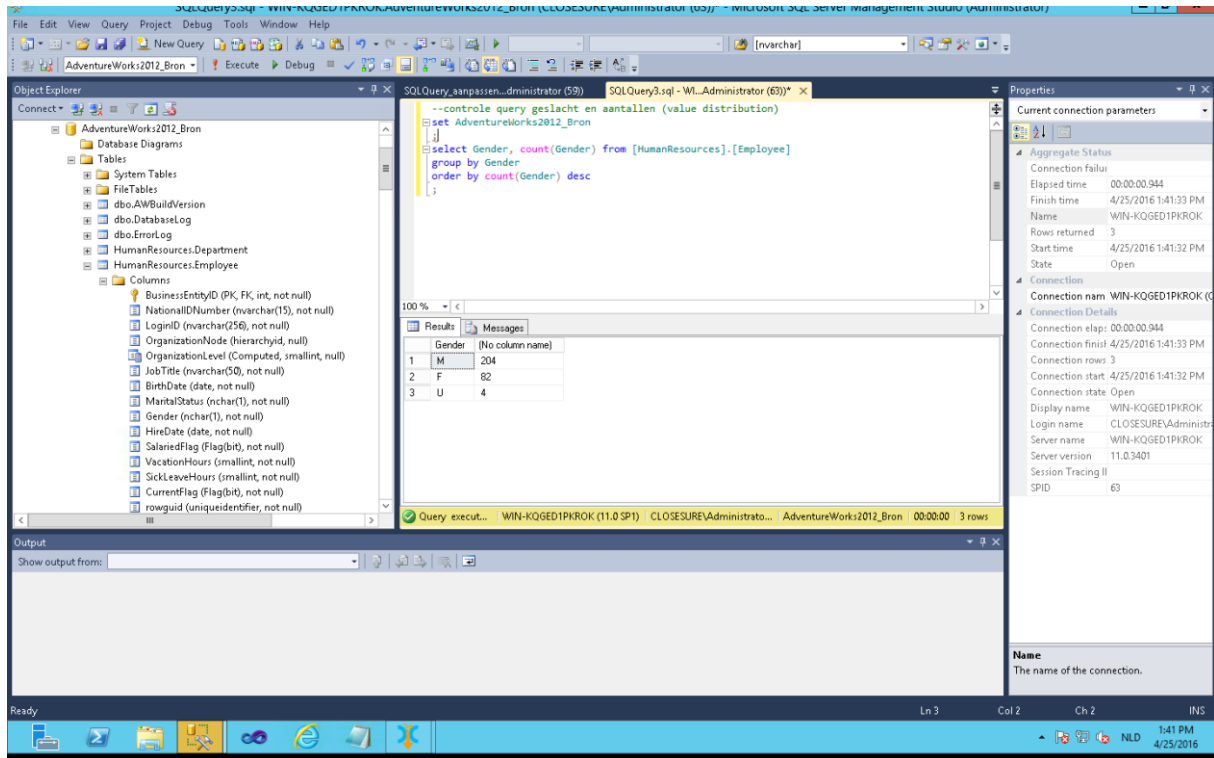
Voorbeeld SQL:

```
-- onderstaande GROUP BY query is een van de mogelijkheden om de verdeling te tonen

--controle query geslacht en aantallen (value distribution)
-- stel eerst de juiste database in
set AdventureWorks2012_Bron;

-- Doe vervolgens de selectie
select Gender, count(Gender) from [HumanResources].[Employee]
group by Gender
order by count(Gender) desc;
```

Screenshot:



The screenshot displays the Microsoft SQL Server Enterprise Manager interface. The left pane shows the 'AdventureWorks2012_Bron' database structure. The central pane shows a SQL query being executed: `--controle query geslacht en aantallen (value distribution)`
`set AdventureWorks2012_Bron`
`select Gender, count(Gender) from [HumanResources].[Employee]`
`group by Gender`
`order by count(Gender) desc`

The 'Results' pane shows the output of the query:

Gender	(No column name)
M	204
F	82
U	4

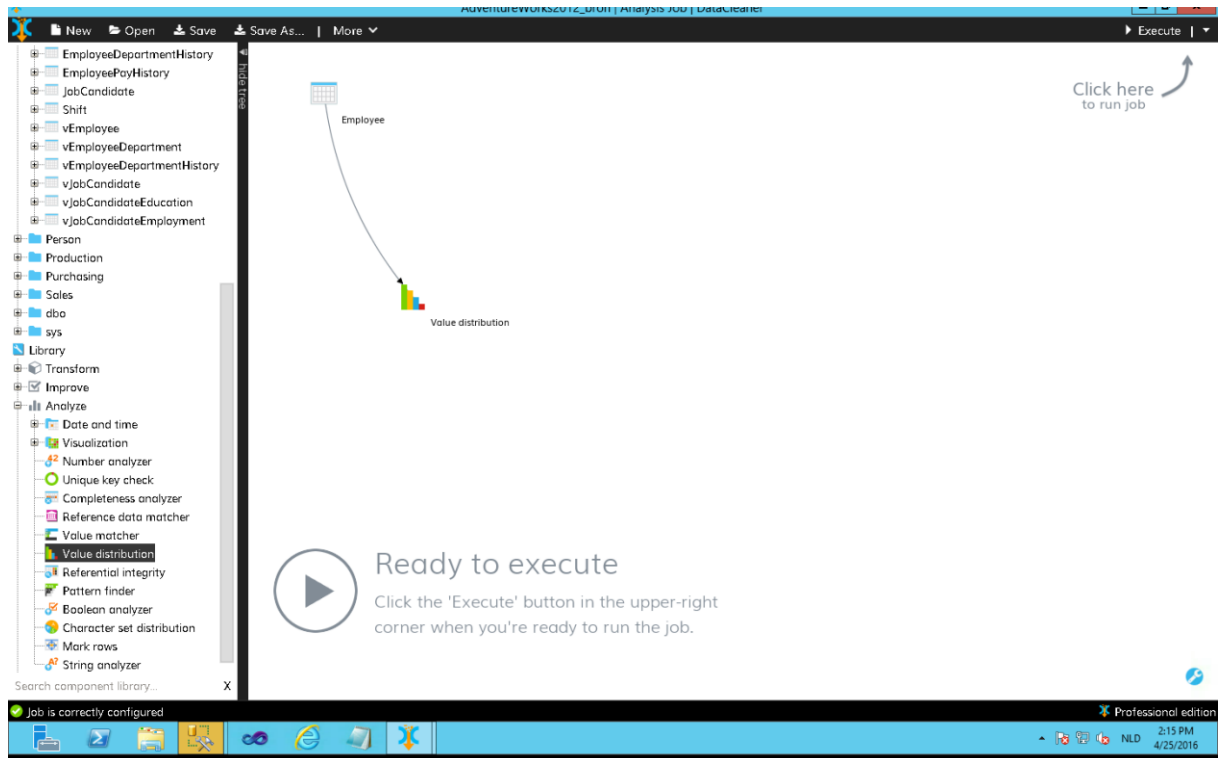
The 'Properties' pane on the right shows connection details for 'WIN-KQGED1PKROK'.

Taak 1b.

Realiseer met de beschikbare omgeving via DataCleaner een selectie op de gegevens waarmee u kunt vaststellen welke waarden de kolom Gender daadwerkelijk bevat. Eventueel met de spreiding (aantallen) per waarde.

Uitwerking:

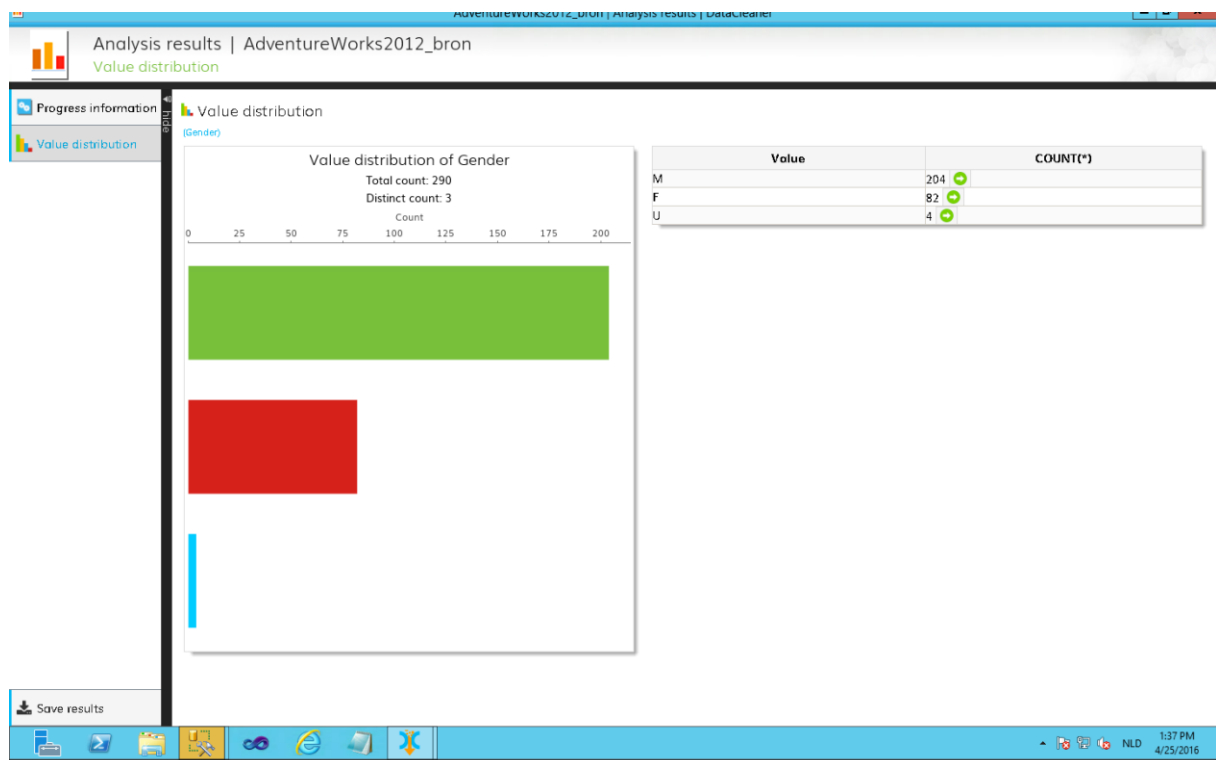
Selecteer in DataCleaner de juiste database (datasource) "AdventureWorks2012_Bron".
Sleep de tabel Employee in het vlak rechts.
Selecteer de functie Value Distribution in het vlak en verbind de tabel met deze functie.



Click here to run job

Ready to execute

Click the 'Execute' button in the upper-right corner when you're ready to run the job.



Klik eventueel door op de onverwachte waarden (groene pijlen) om vast te stellen welke rijen/voorkomens het betreft.