

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.

Tip:

Login met **remote desktop** op de omgeving (zie Handleiding voor details).
Gebruik de **Microsoft SQL Server Management Studio** om een query uit te voeren op de Database "Adventureworks2012_Bron".
Bv. een GROUP BY query.

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.

Tip:

Login met **Remote desktop** op de omgeving (zie Handleiding voor details).

Start **DataCleaner** op en selecteer "Adventureworks2012_Bron" als datasource.

Sleep tabel "Employee" in het veld, selecteer een functie uit DataCleaner en verbind de twee. Start de job en beoordeel het resultaat.