

Opdracht 3

Profiling across Tables or Files

Compare data, determine overlapping or identical sets of values, indicate foreign keys, indicate orphan records

U test een datawarehouse. De data van de brontabel Production.Product uit database AdventureWorks2012_Bron, hierna: Source) wordt bewerkt en geladen naar doeltabel Production.Product in de database AdventureWorks2012 (hierna: Target).

Een eerste laadactie heeft plaatsgevonden in de testomgeving en uw collega's van het ontwikkelteam vragen u om kort te beoordelen of dit goed lijkt te zijn gegaan. U beschikt nog niet over gedetailleerde testgevallen maar u wilt wel de aantallen en de spreiding vergelijken.

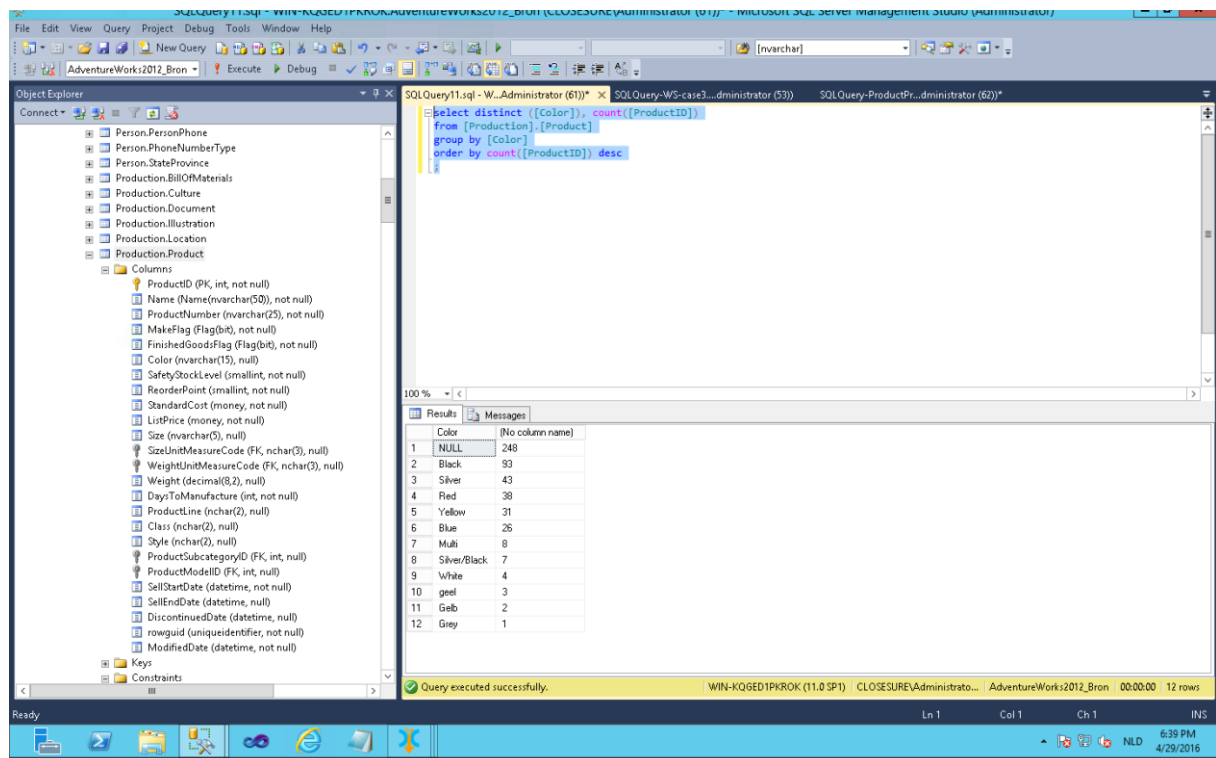
Taak 3a)

Realiseer een overzichtje middels SQL waarmee u een goede indruk kunt geven van de spreiding en de aantallen (compleetheid).

Uitwerking

Query voorbeeld:

```
select distinct ([Color]), count([ProductID])
from [Production].[Product]
group by [Color]
order by count([ProductID]) desc
;
```

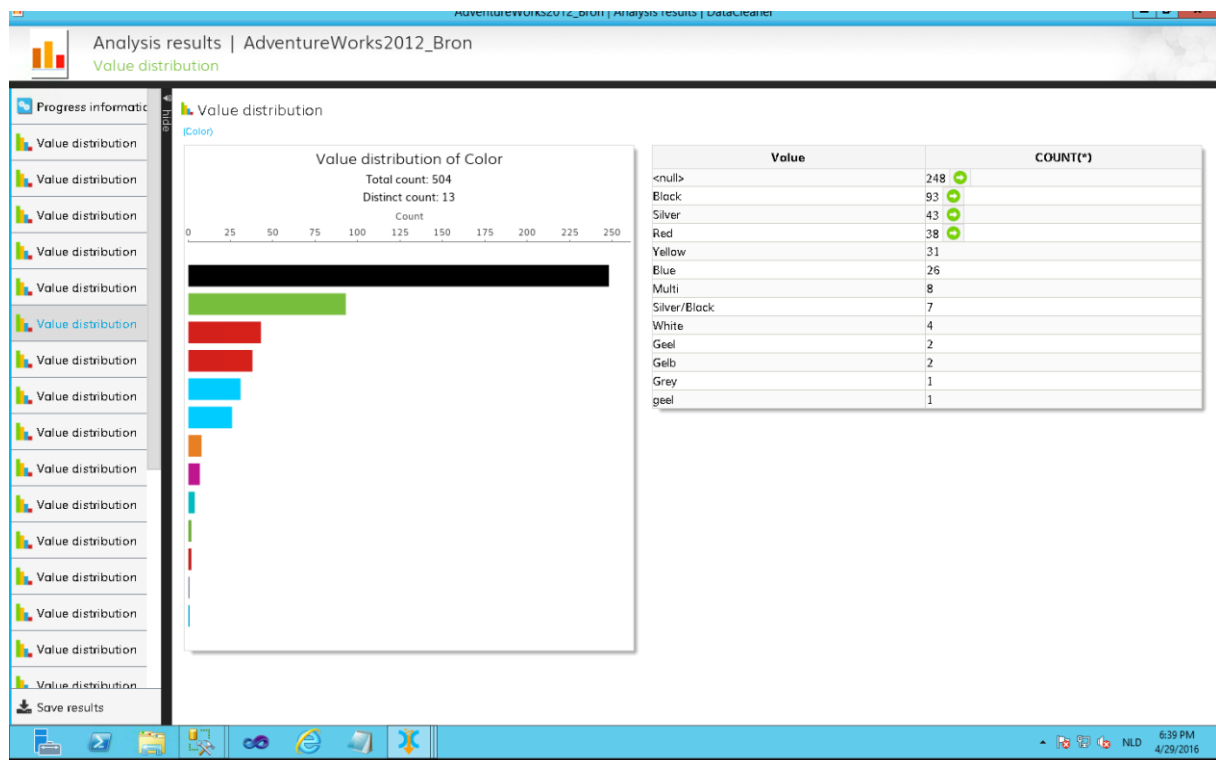


Taak 3b)

Realiseer een overzichtje middels DataCleaner waarmee u een goede indruk kunt geven van de spreiding en de aantallen (compleetheid).

Uitwerking

Selecteer de datasource Adventureworks2012_Bron en doe een analyse over alle aanwezige tabellen. Hieronder als voorbeeld een screenshot van de waarden in "Color".



Selecteer vervolgens de datasource "Adventureworks_2012" en voer dezelfde handelingen uit.

