

# Het proces van Data Analyse

Data-Analyse is een proces waarbij er ruwe data omgezet wordt naar nuttige informatie. Deze informatie kan ingezet worden binnen de besluitvorming rondom vraagstukken. Binnen het data analyse proces kunnen we verschillende stadia onderscheiden.

## De vereiste gegevens voor Analyse

De noodzakelijke input voor de analyse van gegevens. De data wordt gespecificeerd op basis van de eisen die gesteld worden door het management of de klanten van het eindproduct.

## Verzamelen van te analyseren data

Vanuit verschillende bronnen moeten er gegevens verzameld worden om aan de eisen van het management of de klanten te voldoen. Beheerders van de systemen moeten de data aanleveren bij analisten. Data kan verzameld worden vanuit databases, interviews, downloads en andere online bronnen. Data komt ook steeds vaker uit sensoren die bijvoorbeeld het verkeer op een snelweg meten.

## Verwerken van data

De verkregen gegevens moeten zo georganiseerd worden dat zij geschikt zijn voor analyse. Dit kan inhouden dat de gegevens binnen een vaste kolom structuur opgeslagen moeten worden voordat er verdere analyse uitgevoerd kan worden.

## Data cleaning

Zodra de data verwerkt en op de juiste manier georganiseerd is moet de data geschoond worden. Gegevens kunnen onvolledig zijn of de georganiseerde data bevat duplicaten. Data cleaning is nodig om fouten die ontstaan zijn bij de invoer van de data te corrigeren.

## Exploratieve data-analyse

Wanneer de gegevens opgeschoond zijn kan er gestart worden met de analyse van de data. Analisten zullen een eerste exploratieve data analyse uitvoeren. Deze eerste analyse heeft als doel te onderzoeken of er voldoende data en correcte data beschikbaar is om het doel van de analyse te behalen. Het gaat hier vooral om het begrijpen van de data. De eerste analyse kan leiden tot een tweede data cleaning of verzoeken voor extra data.

## Modellering en algoritmen

Na de exploratieve data-analyse zullen er wiskundige formules of modellen toegepast worden op de data. Door middel van deze algoritmen kan er correlatie of causaliteit identificeert worden tussen de variabelen.

## Data producten

Afhankelijk van de frequentie van de analyse kan er een applicatie ontwikkeld worden die de input automatisch omzet naar bruikbare informatie. Deze applicatie kan gebaseerd zijn op een model of algoritme. Een goed voorbeeld is een applicatie die de klant adviseert om op basis van zijn inkoop geschiedenis andere aankopen te doen.

## Communicatie

Zodra de gegevens geanalyseerd zijn kunnen er rapporten opgemaakt worden. De rapporten kunnen voor elke gebruiker anders van inhoud zijn.

Bron: <http://www.marqit.nl/data-analyse>

