

## DATA-ANALIST (ZATERDAG TRAINING)

### HET ANALYSEREN EN VERWERKEN VAN DATA

## OMSCHRIJVING

### Introductie

De **gecertificeerde opleiding tot Data-Analist** wordt ingericht op twee verschillende momenten:

- Woensdagavond-groep
  - Startdatum: 9/9/2020
  - Lesmomenten telkens op woensdagavond van 18u30 tot 22u, aangevuld met 4 lesmomenten op zaterdagvoor- en namiddag (periode januari - februari 2021)
  - Einddatum: eind juni 2021
- Zaterdag-groep:
  - Startdatum: 3/10/2020
  - Lesmomenten telkens op zaterdagvoor- en namiddag van 9u tot 12u30 en van 13u15 tot 16u45
  - Einddatum: eind juni 2021

### Omschrijving

Bedrijven beschikken over heel veel intern gestructureerde data (bvb: financiële gegevens, producten, diensten, klantgegevens, prospecten, interne kennis,...) die in databases aanwezig zijn. Om voeling met de markt, met hun consument of met hun klanten te behouden, is het voor bedrijven noodzakelijk om ook strategisch externe informatie te capteren en vervolgens binnen hun bedrijfsdatabase-systemen als gestructureerde data te registreren.

De kunst om al deze interne en externe databronnen om te zetten naar informatie die strategisch bruikbaar is, vormt een hoeksteen van de huidige en toekomstige bedrijfsvoering. De CIO (Chief Information Officer) op managementniveau, aanvullend geflankeerd door een data-analist, zijn de twee sleutelfiguren om het toekomstig succes van nieuwe producten en diensten op een objectieve onderbouwde en systematische wijze te helpen realiseren. Binnen deze ' **first-time-right-filosofie**' is de snelheid en tegelijk de nauwkeurigheid waarmee nieuwe (of aangepaste) producten of diensten in de markt worden gezet een duurzame succesfactor gebleken.

De (big) data-analist zet data om naar informatie in nauwe samenwerking met CIO, waarbij vervolgens deze informatie opnieuw geanalyseerd en geïnterpreteerd wordt door het managementteam. De CIO is verantwoordelijk voor de juiste duiding en context van cijfers en alle gepresenteerde informatie afkomstig uit bedrijfsdatabases.

De data-analist moet vaardig zijn in het begrijpen van grote stromen gegevens (bvb: financiële cijfers, groeicijfers, producteigenschappen, onderzoeksresultaten, tendenzen...) en deze gegevens overzichtelijk, helder en snel aan belanghebbenden kunnen presenteren.

De data-analist analyseert en verwerkt gegevens, maar is niet verantwoordelijk voor deze informatie. De eindverantwoordelijkheid van accuraatheid van de data, de methodes en de juiste duiding van de begrippen rond de data ligt bij CIO. De data-analist speelt hierbij echter een zeer belangrijke ondersteunende functie. In heel wat KMO's wordt de jobrol van CIO en data-analist namelijk door één en dezelfde persoon gerealiseerd.

### Voor wie is deze opleiding bestemd?

CIO's, IT-managers, business-analisten, ontwikkelaars, marketeers en database administrators die meer willen leren over data-analyse

### Voorkennis

Er zijn geen strikte toelatingsvoorwaarden, maar:

- er wordt van uitgegaan dat de cursisten een behoorlijke affiniteit hebben met wiskunde, logica en statistiek, wat onontbeerlijk is voor het volgen van deze opleiding.
- een goede kennis van Excel is een must. Omgaan met SAS en Power BI wordt in de cursus aangeleerd.
- een functionele kennis van het engels wordt vooropgesteld. Dit gezien alle documentatie en leerplatformen in het engels zijn opgesteld.

## Bijkomende info

- De lessen worden ingepland op zaterdag, waarbij zowel lessen in de voormiddag (9u00-12u30) als in de namiddag (13u15-16u45) worden georganiseerd.
- Je volgt een traject met ordebepaalde modules, het curriculum legt de volgorde van de modules op. Dit betekent concreet dat je moet geslaagd zijn voor de examens van de modules van een lagere orde vooraleer kan deelgenomen worden aan de examens van modules met een hogere orde. Aanwezigheid in de lessen is dus van groot belang.
- Maaltijden over de middag zijn niet voorzien.
- De eerste module is vrij theoretisch. Daarna gaan we praktisch aan de slag met SQL, SAS en Power BI. Programmeren in R en/of Python is niet opgenomen in de leerdoelen.  
Voor nieuwjaar wordt een eindwerkthema gekozen.
- De eindproefverdediging vindt plaats eind juni 2020.

## Methodologie

- De lessen worden klassikaal georganiseerd in campus Kortrijk, Doorniksesteenweg 220.
- Deelnemers dienen zelf een laptop mee te brengen naar de opleiding.
  - Minimum vereiste hardwareconfiguratie van de laptop: 4 GB RAM - 1GB vrije HDD ruimte - Windows 7 of hoger - x64-bit processor aanbevolen.

## PROGRAMMA

### Module 1: data structuur en data modeling (36 uren)

- Het beroep van data-analist: jobrollen ,verantwoordelijkheden en toepassingsmogelijkheden
- Herkennen van behoeftes en belangen van verschillende bedrijfsafdelingen, GDPR (General Data Protection Regulation)
- (Enhanced) Entity Relationship Diagram (ERD & EERD)
- Data modeling
- Ontwerpen en aanmaken van tabellen
- Werken met datatypes
- Relaties, indexen en Unieke ID
- Ontwerpen: normalisatieniveaus
- Aanmaken van tabellen
- Datanormalisatie
- Datakwaliteit
- Data Warehouse (DWH), Big Data en Business Intelligence
- Toepassen van ETL (Extract, Transfer , Load) en ECTL (Extract, Clean, Transfer, Load)
- Beroepservaringen en korte praktijkoefeningen

### Module 2: data-management (60 uren)

- Inleiding en definities: wat is een databank, data, soorten datatypes?
- SQL
  - Select Query
  - Joins
  - Group by
  - Subqueries
  - Standaard functies
- SAS
  - SAS software installeren en basisstructuur van de verschillende toepassingen creëren van rapporten dmv filters, groeperingen en sorteren.
  - SAS formats aanmaken en gebruiken
  - Data sets in- en uitlezen, aanpassen
  - Excel en CSV-bestanden importeren
  - SAS data manipulatie
  - Data join & summary reports

### Module 3: analyse en visualisaties (36 uren)

- Visual Analytics

- Data Analyse
- Power BI

De cursisten dienen een uitgewerkte analyse en rapportering te maken met een zelfgekozen toepassing op een gegeven dataset en met duidelijke richtlijnen (d.i. een duidelijke probleemstelling en/of vraag die beantwoord dient te worden op basis van een grondige analyse van de dataset).

Gezien de aard van de oefening zal de dataset vermoedelijk van een publieke bron afkomstig zijn, zodat de cursist er zonder problemen aan kan.

*Er kan ook geopteerd worden om met eigen bedrijfsgegevens te werken. In dat geval dient er expliciete goedkeuring verkregen worden van de bedrijfsleider en wordt een DNA opgemaakt.*

Voorbeelden van probleemstellingen zijn:

- Evolutie van de faillissementen doorheen het jaar (gebaseerd op KBO-gegevens)
- Demografische analyse van bepaalde landen (gebaseerd op VN-gegevens)
- Analyse eigen loopgedrag (FitBit data)
- Analyse voetbal data (invloed van weer, ...)