

Onderzoek naar invloed van datakwaliteit op performance

# BUSINESS PERFORMANCE

**Binnen de bedrijfswetenschappen krijgt het onderwerp 'datakwaliteit' steeds meer aandacht. Zowel op het gebied van Business Intelligence en bedrijfsprocessen als marketing wordt het belang van datakwaliteit groter. Dit heeft voor een belangrijk deel te maken met de toegenomen automatisering van bedrijfs- en marketingprocessen.**

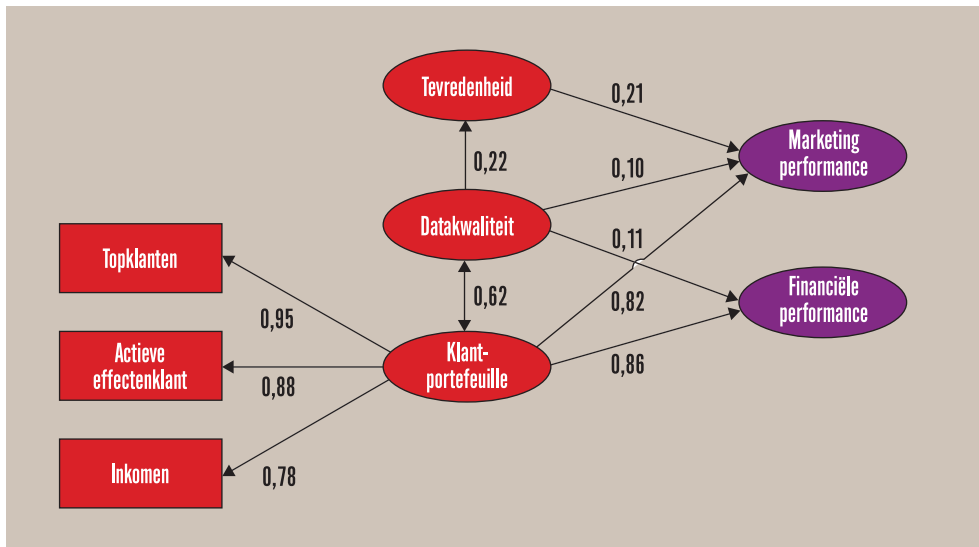
Door Gonneke Zwennes-Van der Plank

**D**it was een reden om de relatie tussen de kwaliteit van informatie en de performance van een organisatie nader te onderzoeken. Uit eerdere onderzoeken is gebleken dat gebrek aan datakwaliteit een behoorlijke kostenpost kan opleveren, waaraan zowel directe kosten als indirecte kosten kunnen worden gekoppeld<sup>1</sup>. Er is duidelijk een direct verband tussen datakwaliteit en kosten. Dit kan ook logisch worden beredeneerd; immers als een organisatie communiceert op basis van foutieve gegevens, dan leidt dit tot een toenemende ontevredenheid bij de klant, foutieve facturatie (lees onbetaalde facturen) en een overvol callcenter waar klagende klanten de medewerkers dagen bezighouden. Toch blijft de vraag overeind in welke mate dit nu invloed uitoefent op de performance van een organisatie, zowel financieel als wat betreft marketing performance. Tevens is de vraag wat oorzaak en gevolg is. Realiseren betere klanten nu meer datakwaliteit of zorgt datakwaliteit voor betere klanten? 'Hoe' loopt de relatie? Om deze vragen te kunnen beantwoorden is in samenwerking met een grote Nederlandse bank een onderzoek verricht naar de rol van datakwaliteit

bij de performance van bankkantoren. De onderzoeksvraag hierbij is geweest: "In welke mate wordt de performance van een onderneming beïnvloed door de kwaliteit van de klantcontactgegevens?"

## **Bankkantoren als bedrijven**

Om een uitspraak te kunnen doen over de invloed die datakwaliteit uitoefent op de performance van een onderneming is het noodzakelijk om te kunnen beschikken over 'performance'-gegevens van veel bedrijven. Daarbij komt het de kwaliteit van het onderzoek ten goede als de onderlinge verschillen tussen deze bedrijven tot een minimum zijn beperkt. Immers, organisaties die opereren in een vergelijkbare omgeving zullen over het algemeen een vergelijkbare set aan beïnvloedbare factoren kennen met betrekking tot hun performance. Vandaar dat voor het onderzoek is gekozen om bankkantoren van een groot financieel concern in Nederland te gebruiken als onderzoekseenheid. Bankkantoren zijn over het algemeen zelfstandig opererende ondernemingen binnen het bankconcern, met een eigen verzorgingsgebied, een eigen profit & loss account en een min of meer zelfstandig commercieel beleid. De overeenkomsten



Afbeelding 1: Model Datakwaliteit en bedrijfsperformance.

als het gaat om het productportfolio, de prijspropositie en de verkoopstrategie, maakt bankkantoren uitermate geschikt om de invloed van factoren als datakwaliteit te kunnen onderzoeken. Voor het onderzoek hebben 160 bankkantoren naast hun klantenbestand gegevens aangeleverd op kantooniveau, variërend van het aantal en type klanten, de financiële performance en marketing performance tot het personeelsverloop en de concurrentie in het marktgebied. Op deze wijze zijn op kantooniveau de meest relevante factoren verzameld die invloed uitoefenen op de performance van de onderneming<sup>2</sup>. Het klantenbestand van de 160 bankkantoren is geanalyseerd op datakwaliteit. Zo is op kantooniveau bekend hoeveel procent van de klantgegevens in de database foutief is. Met dit gegeven en de andere data op kantooniveau proberen we de financiële performance en marketing performance te voorspellen. Financiële performance is gedefinieerd als de gemiddelde Net Banc Income per klant. Marketing performance is geoperationaliseerd met de RFM score.

### Statistische correlatie en verbanden

Dat er statistisch een relatie bestaat tussen datakwaliteit enerzijds en financiële en marketing performance anderzijds zal niemand erg verbazen. Er zijn al verscheidene onderzoeken uitgevoerd die deze relatie naar voren halen. Maar een statistische relatie zegt op zichzelf niets. Een bedrijf opereert in een werkelijkheid waarin veel factoren invloed uitoefenen op de performance. Denk aan concurrentie, kwaliteit van mensen, het weer etcetera. Om de invloed van deze en andere beïnvloedende variabelen alsmede hun onderlinge correlatie mee te wegen, is getracht alle relevante variabelen mee te nemen in het onderzoek en deze via *Structural Equation Modeling* (SEM) te modelleren.

Met behulp van SEM kunnen zeer complexe modellen getoetst worden. Er wordt verondersteld dat de onderlinge samenhang tussen de geobserveerde variabelen veroorzaakt

wordt door één of meer onderliggende variabelen, dat wil zeggen, variabelen die 'ten grondslag liggen aan' de geobserveerde variabelen. Deze onderliggende variabelen worden latente variabelen of factoren genoemd. Daarnaast kan men oorzakelijke relaties specificeren tussen de factoren onderling.

Structurele modellen geven meer inzicht in de oorzaak/gevolg-relaties tussen variabelen. Met deze structurele modellen kunnen ook allerlei indirecte effecten gemodelleerd worden. Deze indirecte effecten zijn vaak juist zeer interessant. Een voorbeeld van een indirect effect is de relatie tussen de lengte van een grootvader en de lengte van zijn kleinzoon.

## Tot voor kort was de belangrijkste argumentatie om datakwaliteit te verbeteren een defensieve

De lengte van de grootvader heeft een directe invloed op de lengte van de vader, evenals die van de vader op de kleinzoon. Daarmee heeft de lengte van de grootvader dus een indirecte invloed (namelijk via de vader) op die van de kleinzoon.

### Resultaten

Allereerst is gekeken naar de enkelvoudige samenhang tussen datakwaliteit en financiële en marketing performance. Deze samenhang blijkt erg sterk te zijn: de Pearson-correlatie is respectievelijk 0,64 en 0,65.

Andere variabelen die sterk samenhangen met zowel financiële performance als marketing performance, zijn het

## Met behulp van SEM kunnen zeer complexe modellen getoetst worden

percentage topklanten, het percentage actieve effectenklanten en het inkomen dat het kantoor gemiddeld per klant ontvangt. Dit zijn variabelen die betrekking hebben op de klantportefeuille en waarvan we op basis van logica al kunnen veronderstellen dat er een sterk verband is met performance. Een kantoor met veel topklanten zal immers meer omzet per klant genereren dan een kantoor met weinig topklanten.

Een belangrijke vraag is of datakwaliteit de financiële en marketing performance *autonoom* beïnvloedt, onafhankelijk van de klantportefeuille van een kantoor. Met andere woorden als we de invloed van de klantportefeuille op nul zetten, is er dan nog steeds een positieve correlatie tussen datakwaliteit en performance? Het structurele model in afbeelding 1 geeft antwoord op deze vraag. De getallen bij de pijlen geven de sterkte van het effect weer. Dit zijn de zogenaamde 'gestandaardiseerde coëfficiënten' die onderling vergelijkbaar zijn. Alle coëfficiënten zijn statistisch significant<sup>3</sup>.

Uit het model blijkt dat de klantportefeuille de meest verklarende factor is voor de financiële performance en marketing performance: hoe beter de klantportefeuille, hoe hoger de financiële en marketing performance. Indicatoren van de klantportefeuille zijn het percentage topklanten, het percentage actieve effectenklanten en het inkomen dat het kantoor gemiddeld per klant ontvangt. Datakwaliteit blijkt, onafhankelijk van de klantportefeuille, ook een positief effect te hebben op de financiële en marketing performance. Dit betekent dat van twee kantoren met een gelijke klantportefeuille, het kantoor met een betere datakwaliteit significant beter presteert. Marketing performance wordt naast de klantportefeuille en datakwaliteit beïnvloed door klanttevredenheid. Hoe tevredener de klanten, hoe sterker de loyaliteit. Tevreden klanten zijn met andere woorden loyaler aan het kantoor dan ontevreden klanten.

Datakwaliteit blijkt marketing performance op twee manieren te beïnvloeden: direct en indirect. Indirect heeft datakwaliteit effect op marketing performance doordat datakwaliteit van invloed is op klanttevredenheid. Slechte data (verkeerd geschreven namen, foute adressen etcetera) zorgen voor irritatie bij de klant.

De klantportefeuille blijkt ten slotte sterk samen te hangen met datakwaliteit. De kwaliteit van de data is over het algemeen beter bij kantoren met een betere klantportefeuille.

Dit kan wellicht twee kanten op gaan. Enerzijds zal een kantoor waarschijnlijk meer investeren in goede data met betrekking tot hun topklanten. Anderzijds zal een klant die veel producten afneemt en veel geld 'op de bank heeft staan' ervoor zorgen dat zijn of haar gegevens goed geregistreerd staan. Het model verklaart de financiële en marketing performance goed. Bijna 90 procent van de variantie in financiële en marketing performance wordt verklaard door de variabelen in het model.

### Wordt performance beïnvloed door datakwaliteit?

Het model toont aan dat performance direct wordt beïnvloed door datakwaliteit. Daarmee wordt de onderzoeksvraag positief beantwoord. De hamvraag is alleen of we deze invloed kunnen kwantificeren. Met andere woorden, hoeveel euro's verdient deze bank met 1 procentpunt meer datakwaliteit? Uit berekening blijkt dat dit naar schatting per klant enkele euro's meer omzet oplevert. Voor de betreffende bank levert dit in totaal miljoenen euro's extra omzet op. Door een betere datakwaliteit is een bankkantoor in staat de klantwaarde substantieel te verhogen door de verloren geraakte binding te reactiveren.

Tot voor kort was de belangrijkste argumentatie om datakwaliteit te verbeteren een defensieve, te weten: kostenbesparing. Met dit onderzoek wordt aangetoond dat een offensieve argumentatie – meer omzet of een hogere performance – eveneens als argument aangevoerd moet worden om de kwaliteit van klantgegevens te verbeteren.

### Noten

1. Het onderzoek 'Nationaal Onderzoek Datakwaliteit' is uitgevoerd door Cendris in 2004 en het onderzoek 'De kosten van slechte datakwaliteit' is uitgevoerd door het Instituut voor Commerciële Strategie en Beleidsvorming in 2005.
2. Een gedetailleerde omschrijving van het onderzoeksveld gaat in het kader van dit artikel te ver. Om de invloed van afzonderlijke factoren te kunnen destilleren, is naast geografische kenmerken vooral gekeken naar medewerkersmotivatie, concurrentiedruk, klantportefeuille, grootte van het bankkantoor en klanttevredenheid.
3. De financiële performance was erg scheef verdeeld en is daarom getransformeerd ten behoeve van de validiteit van de analyse.

**Gonneke Zwennes-Van der Plank**

Drs. G.G. Zwennes-Van der Plank is als Senior Data-analist werkzaam bij Cendris.