

Data en algoritmes



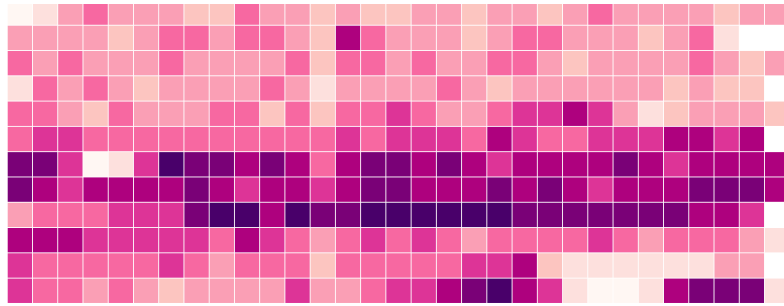
Presentatie voor Avond voor de Stad op 16 mei 2024 over data en algoritmes voor de gemeenteraad van Nieuwegein.

Door: Eelco Rommes (M&I/Partners.)

Vragen? Opmerkingen? Mail: eelco.rommes@mxi.nl

(versie 4 - laatst gewijzigd: 21 mei 2024.)

Data



Waar kijken we naar?

Wie nauwkeurig telt, zal zien dat er 12 rijen van 29 tot 31 gekleurde blokjes staan. Je hoeft geen Sherlock Holmes te heten om te begrijpen dat het om de dagen van het jaar gaat. Maar die kleuren kunnen van alles betekenen: *zonder context zijn data betekenisloos*.

Hier staan de kleuren voor *het aantal jarigen* per datum. Hoe donkerder, hoe meer jarigen. Daarom is het 29^e blokje van de tweede rij heel licht gekleurd: er zijn maar weinig mensen geboren op een schrikkel dag.

Nu je dit weet, kan je uit patronen in de kleurverdeling opmaken om welk land het gaat. Bijvoorbeeld: op kerstmis en nieuwjaarsdag en op elke dertiende van de maand is er een dipje. Maar ook op 4 juli, wanneer de VS hun onafhankelijkheid vieren.

Maar waarom? Mensen kiezen hun verjaardag toch niet zelf?

Tot op zekere hoogte wel: in de VS worden veel bevallingen ingeleid. Artsen plannen graag om de feestdagen heen. En bijgelovige, aankomende ouders proberen de dertiende te vermijden. (En naïeve, aanstaande ouders vinden het blijkbaar romantisch om op Valentijnsdag te bevallen.)

Wat is 'goed'?

Relevant	Actueel
Compleet	Frequent
Precies	Snel
Consistent	Betekenisvol
Geldig	Plaatselijk
Juist	Rechtmatig
Significant	Ethisch
	...



Wie wel eens in een auto met een defecte brandstofmeter heeft gereden, snapt hoe belangrijk het is om te sturen op goede data. Maar wat is kwaliteit? En wanneer is goed ook goed genoeg? Dat verschilt enorm per toepassing.

Stel je voor: we kunnen van alle woningen in de gemeente met 75% zekerheid voorspellen of er in het dak asbest is gebruikt. Is dat goed genoeg om met inwoners in gesprek te gaan? In een kwart van de huizen zal geen asbest aanwezig zijn.

Een ander voorbeeld: de zorgspiegel van Vektis levert cijfers over het zorggebruik in een gemeente. Maar die gegevens zijn actueel tot twee jaar geleden.

Dat is nuttige informatie voor wie zich bezighoudt met beleid voor langdurige zorg. Wie nu moet weten hoe het er de afgelopen maand voorstond, kan er veel minder mee.

Het veranderen van datakwaliteit is vaak ingewikkeld en kostbaar, zeker als de data door anderen worden verzameld en beheerd. Maar ook interne werkprocessen aanpassen om betere of andere data te verzamelen, is lang niet altijd gemakkelijk.

Weet wat je meet



Sturen op data is een kunst op zich. Harde cijfers zijn vaak niet zo hard als ze lijken. Vaak meten we andere dingen dan we dachten, of we veranderen de werkelijkheid alleen al omdat we gaan meten.

Een groot ziekenhuis wilde de poliklinische zorg verbeteren en besloot daarom de patiënttevredenheid te meten. Na een half jaar gebeurde er iets opvallends: de afdeling die het hoogst scoorde op tevredenheid, had daar het minst in geïnvesteerd. Geen lean procesverbetering, geen klantreizen, geen 'de patiënt centraal' veranderprogramma...

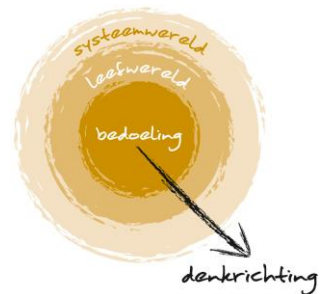
In plaats daarvan had de afdelingsmanager in de wachtkamer een heel goede espressomachine neergezet. (De theedrinkers kwamen er weer eens bekaaid af...)

Van gemeenten wordt steeds vaker verwacht dat ze rapporteren over harde, meetbare resultaten, bijvoorbeeld om subsidies te verantwoorden. Dat klinkt logisch en transparant. Maar dat is nog niet altijd eenvoudig.

Hoe meet je het gevoel van veiligheid in een wijk? Of de leefbaarheid van een buurt? Hoe voorkom je dat een espressomachine het beeld vertroebelt?

Wat is er nodig?

- Een doel
- Een aanpak
- Toegang tot data
 - Goede data
 - Nu en straks
- Interpretatie
 - Betekenis geven aan data



Om met data te werken, moet je bepaalde zaken op orde hebben.

Het begint met een duidelijk *doel*. Gemeentes hebben enorme hoeveelheden gegevens. Dan kan het verleidelijk zijn om eens te kijken wat je daar zoal mee kan. En het is best lastig is om een doel te formuleren als je niet weet wat er mogelijk is. (Voor de boomers onder ons: weet u nog hoe we mobiele telefoons maar onzin vonden? Altijd bereikbaar zijn, waar kon dat nou goed voor zijn?)

Aanpak: een uitweg uit deze impasse is om gecontroleerd te experimenteren, daarvan te leren en bij te sturen: wat kan er? Wat niet? Wat is waardevol? Wat is daarvoor nodig? Welke risico's zijn er? Wat doen we daarmee?

De juiste aanpak is (deels) maatwerk: wie de bezetting van een call center wil optimaliseren, stelt een andere vraag dan iemand die werkt aan meerjarig beleid voor de jeugdzorg.

Toegang: soms zijn afspraken nodig met externe partijen. Soms moet je data uit eigen systemen putten, maar wel in het juiste formaat, op de juiste tijd, in de juiste frequentie, met de juiste kwaliteit. Misschien moet een werkproces daarvoor veranderen, om data goed en consequent vast te leggen.

Interpretatie: het **verhaal** achter data

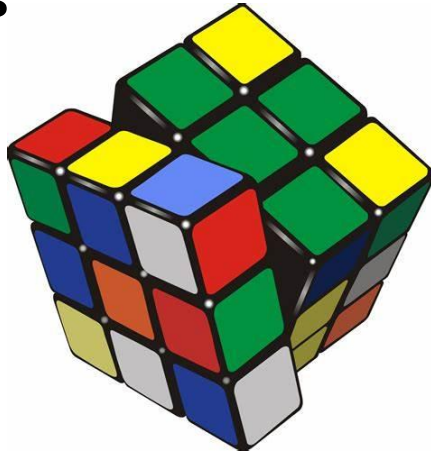


En als de juiste data er dan is, van goede kwaliteit, voorzien van context, is het nog altijd belangrijk om oog te houden voor de werkelijkheid die achter de cijfers schuilgaat.

Een ambtenaar in het sociaal domein was in overleg met een woningcorporatie over een lastige inwoner, een wanbetaler. Deze mevrouw had jarenlang trouw haar huur betaald, maar de laatste maanden schoot ze consequent tekort. En omdat ze ook niet reageerde op de brieven die haar waren gestuurd, vond de woningcorporatie het genoeg: tijd om een incassobureau in te schakelen. De ambtenaar bood aan eerst eens bij deze mevrouw langs te gaan.

Zij bleek een vriendelijke, iets oudere dame te zijn. Ze betaalde haar huur altijd automatisch, maar die was met enkele euro's verhoogd. Het moderne, digitale bankieren vond ze lastig: het lukte haar niet de aanpassing door te voeren. De bank om de hoek was allang geleden gesloten – die 'stuurde op digitale kanalen'. Van de eerste brief van de woningcorporatie werd ze wat zenuwachtig en volgende brieven legde ze netjes in een la. Totdat deze ambtenaar bij haar langsging, was ze niet als mens benaderd: voor de bank en de woningcorporatie was ze 'een zaak', met rekenregels en workflows en automatisch gegenereerde brieven.

Algoritmes



Toen ik in 1994 begon aan een studie informatica, wist niemand wat een algoritme was. Ja, mijn medenerts en een enkele verdwaalde wiskundige. Tegenwoordig staat nu.nl er vol mee. Maar wat zijn algoritmes nou precies?

In de kern is een algoritme niet meer dan een stappenplan om een probleem op te lossen. Vrijwel ieder van ons heeft op enig moment in het leven een strikdiploma gehaald en sindsdien gebruiken we dagelijks een algoritme om onze schoenen vast te maken aan onze voeten.

Een ander voorbeeld: op internet zijn volop filmpjes te vinden van mensen die in een ontstellend tempo een Rubiks kubus oplossen. Dat kunnen ze niet omdat ze genieën zijn die razendsnel nadenken: ze hebben een algoritme uit hun hoofd geleerd en net zolang geoefend tot ze het zonder nadenken kunnen uitvoeren.

Een kookboek staat vol algoritmes. Kortom: algoritmes zijn er al heel lang en we gebruiken ze voortdurend in allerlei situaties.

Maar over zulke algoritmes maakt De Telegraaf zich natuurlijk niet druk. Het gaat nu om algoritmes die we door computers laten uitvoeren. Hoe dat dan?

Voorbeeld



Eerder constateerden we hoe belangrijk context is om data te kunnen begrijpen. Computers hebben geen besef van context of interpretatie. Ze voeren blindelings uit wat hen wordt opgedragen.

Een voorbeeld kan helpen om computeralgoritmes beter te begrijpen.

Laten we een klassiek probleem nemen: sorteren van een rij getallen (of, in dit geval: Uno-kaarten. In het echte leven ben ik vaker bezig met het ontsorteren, ofwel schudden, van zulke kaarten. Maar het is een leerzaam voorbeeld.)

Laten we, stap voor stap, een algoritme maken om deze kaarten op volgorde te zetten, van klein naar groot.

Selection sort

Vind de laagste kaart.
Wissel om met kaart 1.

(Voor de computernerds onder ons: ik laat *selection sort* zien in pseudocode, om het voor iedereen begrijpelijk te houden.)

Stap 1: vind de kaart met het laagste getal en wissel die om met de eerste kaart. Nu staat de kaart met het laagste getal vooraan.

We weten dus dat we aan die voorste kaart niks meer hoeven veranderen: die staat op de goede plek. We hoeven alleen nog de rest van de kaarten te sorteren.

Selection sort

Vind de laagste kaart.
Wissel om met kaart 1.

Vind de op één na laagste kaart.
Wissel om met kaart 2.

Stap 2: welke kaart op de volgende plek? De op één na laagste kaart. Die vinden we en verwisselen we met de kaart die nu op plek 2 staat.

Nu staan de eerste twee kaarten op de juiste plek.

Selection sort

Vind de laagste kaart.
Wissel om met kaart 1.

Vind de op één na laagste kaart.
Wissel om met kaart 2.

Vind de op twee na laagste kaart.
Wissel om met kaart 3.

Stap 3.

Oplettende kijkertjes beginnen een patroon te zien...

Selection sort

Vind de laagste kaart.
Wissel om met kaart 1.

Vind de op één na laagste kaart.
Wissel om met kaart 2.

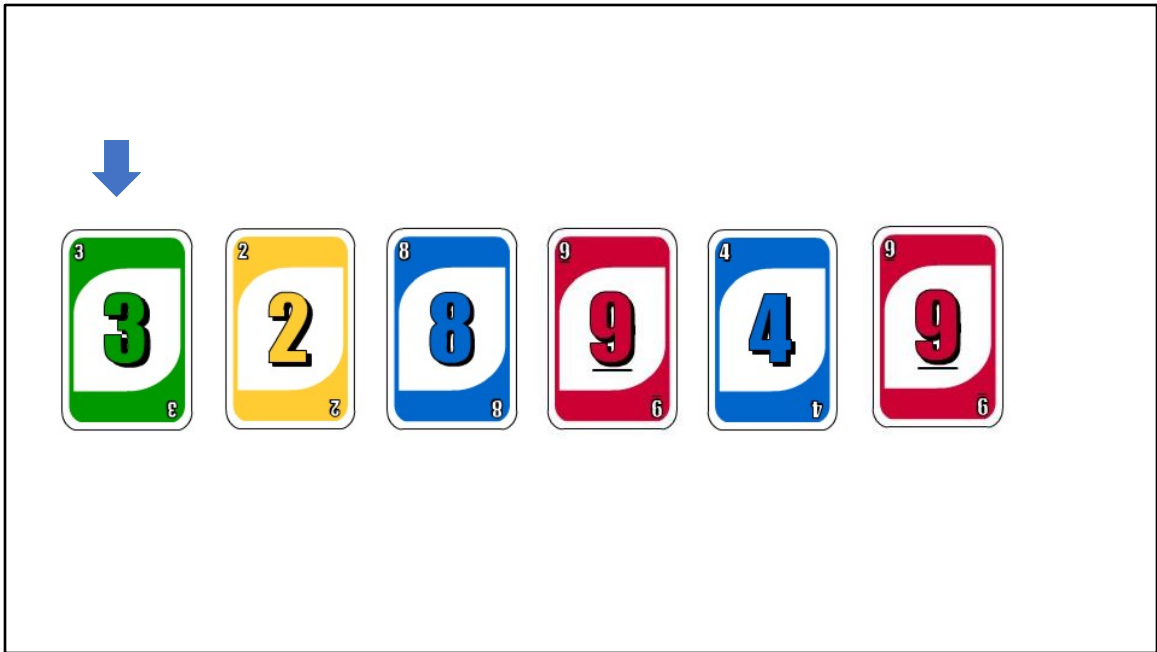
Vind de op twee na laagste kaart.
Wissel om met kaart 3.

[Herhaal tot de laatste kaart]

Op deze manier kunnen we elke rij kaarten sorteren, hoe lang of kort ook.
En eigenlijk: elke rij getallen, of elke rij van zaken die zich onderling laten vergelijken.

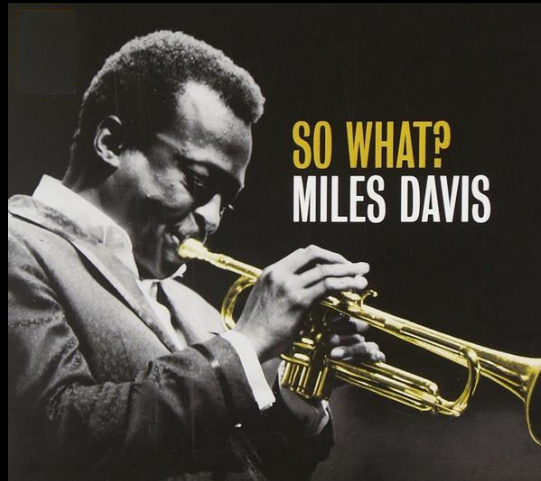
Voor zulk saai, repetitief werk zijn computers uitermate geschikt. Als we een miljoen kaarten moeten sorteren, zal een mens verveeld raken, fouten maken, de tel kwijt raken en (vooral) heel lang bezig zijn. Een computer kan dat in een oogwenk.

En zolang het algoritme klopt, klopt de uitkomst.



En zo ziet dat eruit. (Deze slide bevat een animatie van het algoritme in actie.)

De blauwe pijl wijst steeds naar de eerste kaart die nog niet gesorteerd is. De groene pijl wijst de laagste, ongesorteerde kaart aan.



(Of, zoals Hennie Vrienten zong: Is dit alles?)

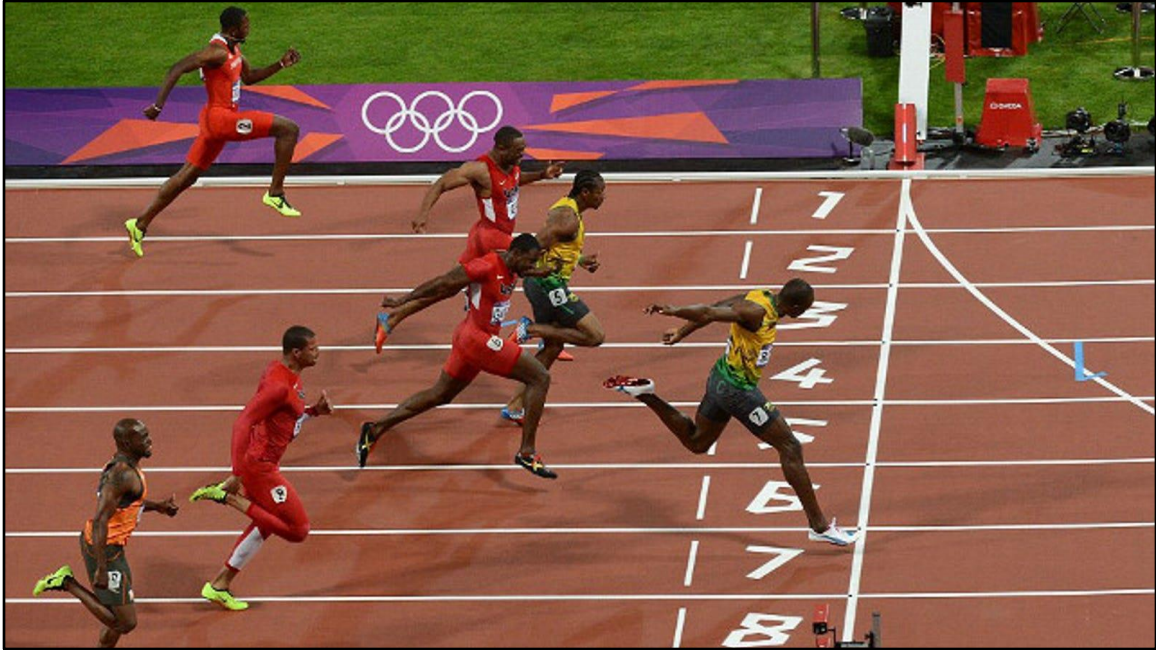
Als algoritmes gewoon recepten zijn om bepaalde problemen op te lossen, waarom zijn ze dan zo controversieel?

Profilering



Bedenk wat er gebeurt als we algoritmen inzetten om iets anders te sorteren dan kaarten (of e-mails, of word-documenten,...)

Wat als we een algoritme gebruiken om mensen te sorteren? En wat als we op basis van de uitkomst, besluiten nemen over die mensen?



Dat hoeft niet spannend te zijn.

We doen dat al heel lang. En vaak is er niks mis mee.

Niet alleen in de sport behandelen we mensen anders op basis van objectief vast te stellen eigenschappen. Kinderen sturen we vanaf hun vijfde naar school, ouderen hebben vanaf hun 67^e recht op AOW, enzovoort enzovoort.

En omdat het om zoveel regels en zoveel mensen gaat – repetitief en saai werk – zetten we daarvoor computers in.

Maar computers begrijpen de context niet: ze voeren domweg heel precies uit wat hen is opgedragen. Zonder gezond verstand en zonder moreel oordeel.

En daar kleven risico's aan.

“Risicomodel”



De Belastingdienst maakte bij het controleren van de aanvragen voor kinderopvangtoeslag gebruik van iets dat ze eufemistisch ‘een risicomodel’ noemden.

Dat model was onderdeel van een algoritme om te berekenen hoe groot de kans was dat iemand fraudeerde. Daarvoor gebruikte het algoritme allerlei eigenschappen van mensen, bijvoorbeeld nationaliteit: niet-Nederlanders kregen van het model sneller het stempel ‘vermoedelijke fraudeur.’

Lang was het buiten de Belastingdienst niet bekend dat dit stempel überhaupt bestond. Daarna was het vrijwel onmogelijk om te weten te komen wie dat stempel had gekregen. En er weer vanaf komen was ook ondoenlijk.

Intussen werd het stempel ook voor allerlei andere doeleinden ingezet.

(nu.nl, 25 oktober 2021)

“Scorekaart”



Gemeentehuis Nieuwegein. Foto: RTV Utrecht

Nieuwegein stopt met omstreden algoritme tegen fraude-opsporing

De WIL gebruikte een algoritme om inwoners die bijstand aanvroegen te sorteren in twee groepen. De ene groep doorliep een snelle procedure, de andere een uitgebreide procedure met meer controles en checks.

Sommige media interpreteerden het als: mensen worden verdacht van fraude als ze nog niets verkeerd gedaan hebben. Dat was niet de bedoeling, maar WIL had de schijn tegen.

Het algoritme zelf was (ongeveer) een tabel met indicatoren die samen een score opleverden. Sommige van die indicatoren waren dubieus, bijvoorbeeld: ‘woont op een woonwagenkamp’.

(Illustratie: Molenkruier, 17 juni 2022)

Kaders



Om zulke problemen te voorkomen, is de inzet van data en algoritmen gebonden aan allerlei wet- en regelgeving.

Soms gericht op een specifiek doel: de AVG voor privacy, de BIO voor informatiebeveiliging en de AI-act over de inzet van lerende algoritmes.

Er zijn ook meer algemene kaders: de grondwet, de universele rechten van de mens,...

Manifest Digitaal Nieuwegein

Inclusief

Zeggenschap

Slim en zorgvuldig

Menselijke maat

Legitiem en gecontroleerd

Open en transparant

Veilig



Uw raad heeft begin 2022 een ethisch kader vastgesteld: het manifest Digitaal Nieuwegein. Bij (spannende) wijzigingen in de informatiesystemen, wegen die waarden mee.

Dat wil niet zeggen dat daarmee de kous af is.

Ethiek is geen probleem dat zich met een algoritme laat oplossen. Daar is altijd maatwerk voor nodig, discussie, het afwegen van waarden en belangen vanuit verschillende perspectieven.

Maar het manifest helpt om duidelijk te maken waar de gemeente waarde aan hecht, wat voor soort overheid ze wil zijn. Het helpt het bestuur én de ambtelijke organisatie om daar ook naar te handelen.

Data en algoritmes kunnen de stad
beter maken.

Zo **kun** je bijvoorbeeld...

Gemeenten staan onder druk. De opgaven worden groter en complexer, regelgeving wordt ingewikkelder, middelen groeien niet even hard mee.

De inzet van data en algoritmen kan helpen, ook binnen de gegeven kaders, om de stad beter te maken.

De vraag is dan: wat wil je wel en wat niet?

Hier volgen enkele voorbeelden van hoe een gemeente data *zou kunnen* inzetten.

...drukte voorspellen

Aan de balie en de telefoon.
Om inwoners sneller te helpen,
en medewerkers slimmer in te zetten.



...looproutes aanpassen

Op basis van **overlastdata** vanuit verschillende systemen,
om **boa's** effectiever in te zetten,
en zo de stad **veiliger** te maken.



Denk aan data over afval, overlast van jongeren, ...

(Een risico is het zelfversterkend effect: waar vaak Boa's lopen, doen Boa's vaker meldingen, zodat het systeem 'nieuwe' hotspots vindt, waar vaker Boa's gaan lopen, etc. Maar je kunt het systeem inrichten en gebruiken om zulke effecten te voorkomen.)

...woonfraude bestrijden

Een opdracht van de raad: een systematische steekproef van risico-adressen.

Hoe kies je die adressen?



Een mogelijk algoritme is: een willekeurige steekproef. Dat betekent wel een (extreem) lage pakkans en vaker overlast voor onschuldige inwoners die onterecht worden gecontroleerd.

Op basis van data kun je gericht kiezen. Welke data? Hoe weeg je die?

De VNG heeft een lijst opgesteld met risico-indicatoren. Sommige daarvan zijn best... spannend:

“Persoon staat als enige ingeschreven in een relatief grote woning.”

“Panden waarop zowel een winkel als een woonbestemming zit.”

“Meerdere volwassenen met verschillende achternamen op een adres. Kijk naar ‘rare’ combinaties van nationaliteiten.”

...gericht voorlichten over **ondermijning**

Om preventief in gesprek te gaan
met **kwetsbare** ondernemers
zodat ze **weerbaar** zijn.



De gemeente heeft te weinig capaciteit om met elke ondernemer te gaan praten.
Wie kies je? Wanneer is een ondernemer 'kwetsbaar'?

...inwoners gericht hulp bieden

Soms laten inwoners geld of hulp liggen omdat ze niet weten dat ze er **recht** op hebben of de weg naar de gemeente niet kennen.



Keer op keer blijkt dat inwoners geld of hulp laten liggen. Om hen pro actief te benaderen, moet je eerst weten wie ze zijn.

Met inzet van data kan dat effectiever.

Wil je dat?

...maar **willen** we dat?

(En hoe?)