



BETER VOORSPELLEN MET DE JUISTE GEGEVENS

INNOVATIEBOOST WHITEPAPER



MAART 2021

vereniging van
woningcorporaties



INNOVATIEBOOST 

COLOFON

© maart 2021 Aedes vereniging van woningcorporaties Den Haag

Whitepaper in opdracht van Aedes vereniging van woningcorporaties

Met medewerking van:

Theo Pronk (3B Wonen)

Tekst:

Martijn Vet

Redactie en vormgeving:

Aedes vereniging van woningcorporaties

Beter voorspellen met de juiste gegevens: datascience bij woningcorporaties

Innovatieboost Whitepaper

Inhoud

Inleiding	5
1 Waaron waningcorporaties moeten innoveren met data	6
2 Hoe kan datascience waningcorporaties helpen?	7
3 Datascience in de praktijk	9
4 Meer doen met data	16
<i>Bijlage</i> Woordenlijst	17

Inleiding

Veel bedrijven maken gebruik van datascience om meer business- en klantwaarde te bereiken. Ook woningcorporaties kunnen veel meer met datascience doen om de veranderingen in de markt bij te houden en hun klanten beter van dienst te zijn. Dankzij datascience kunnen we zeer nauwkeurige voorspellingen doen, bijvoorbeeld wanneer we mutaties kunnen verwachten, wanneer klachten van huurders binnenkomen of wanneer welke cv-ketel storingen zal geven.

In 2019 is Aedes begonnen met de werkgroep Big Data. Hieraan nemen verschillende woningcorporaties deel. Theo Pronk van 3B Wonen bracht namens de groep tijdens de Innovatieboost het idee naar voren om meer kennis over datascience te ontwikkelen.

De groep kreeg een golden ticket om vanuit een opleiding in datascience en procesbegeleiding twee praktijkvoorbeelden uit te werken die door alle andere corporaties kunnen worden toegepast. Daarnaast kreeg de groep de opdracht te bekijken hoe de opgedane kennis kan worden gedeeld met de rest van de sector.

Binnen de werkgroep hebben verschillende deelnemers workshops over datascience gevolgd en inmiddels is een eerste voorbeeld uitgewerkt en vrij beschikbaar: een kant-en-klaar template waarmee ook andere corporaties de mutatiegraad kunnen voorspellen. Verschillende andere kant-en-klare toepassingen zullen de komende tijd volgen.

In dit whitepaper lees je waarom datascience onmisbaar is voor woningcorporaties. Een eenvoudig voorbeeld uit de dagelijkse praktijk van 3B Wonen geeft inzicht in de concrete toepassing van datascience.

1 **Waarom woningcorporaties moeten innoveren met data**

Alle woningcorporaties hebben dezelfde taak: kwalitatief goede woningen bouwen, beheren en voor een betaalbare prijs verhuren. Ook al verandert de functie van woningcorporaties niet, de wereld verandert wel. Zo verwacht het CBS dat de bevolking de komende tientallen jaren flink zal groeien. Het aantal goede woningen zal steeds schaarser worden, wat betekent dat de huizenprijzen nog hoger zullen worden. Dat zet de betaalbaarheid van woningen onder druk. Wat ook verandert, is dat huurders van hun woningcorporatie dezelfde dienstverlening verwachten die ze van (web)winkels gewend zijn.

In het licht van deze veranderingen moeten woningcorporaties zich blijven ontwikkelen en innoveren. Datascience kan een belangrijke bijdrage leveren aan de verbetering van de dienstverlening. Door gegevens over woningen en huurders die beschikbaar zijn op een slimme manier te combineren, wordt het gemakkelijker om beslissingen voor een groot deel geautomatiseerd te nemen, klanten efficiënter te bedienen, voorspellingen te doen, processen te verbeteren en gericht medewerkers te werven.

'Datascience zorgt voor meerwaarde bij organisaties én klanten' – Theo Pronk (3B Wonen)

'Datascience is op dit moment hét innovatieve vakgebied dat bij veel bedrijven en hun klanten meerwaarde biedt. Bij woningcorporaties zien we dat nog te weinig,' vertelt Theo Pronk van 3B Wonen. 'Daarom dongen we mee naar de Innovatieboost voor de ontwikkeling van datascience.'



2 Hoe kan datascience woningcorporaties helpen?

Het doel van datascience is waarde uit data halen. Die data bevinden zich op allerlei plekken in de organisatie. Het kan gaan om gegevens uit databases of ERP-systemen, maar ook berichten op social media of chatgesprekken met klanten kunnen bruikbare data opleveren.

Het proces van datascience bestaat uit een aantal stappen die zich telkens herhalen.

Wil je goede voorspellingen doen op basis van data, dan moet je eerst bepalen welke data je gaat gebruiken. Zijn de data eenmaal verzameld, dan moeten ze worden gefilterd en naar een model worden getransformeerd. Dat doen programmeurs met talen als Python en R. De modellen die hieruit ontstaan, kunnen worden gebruikt om voorspellingen te doen.

Vervolgens kan de data-analist aan de slag met statistiek. Door causale verbanden te leggen, kan de data-analist accurate voorspellingen doen.

Naast de traditionele statistiek is ook machine learning een steeds bruikbaarere manier om voorspellingen te doen. Dankzij deze vorm van kunstmatige intelligentie kunnen algoritmes worden gebruikt die zelflerend zijn en zichzelf steeds verbeteren.

Aan het einde van het proces worden de data visueel in kaart gebracht. Daardoor zijn bepaalde verbanden beter zichtbaar.

Datascience levert vaak een bevestiging op van de vooraf gestelde verwachtingen, maar het is ook mogelijk dat een veronderstelling wordt verworpen of dat er verbanden worden gevonden waar vooraf geen rekening mee werd gehouden.

Voorbeelden van datascience bij woningcorporaties

Datascience kan bij woningcorporaties op verschillende terreinen toegevoegde waarde hebben. Een paar voorbeelden:

1. Onderhoud en assetmanagement

- a. Is er een verband tussen reparatieverzoeken, klachten en betaalgedrag van huurders?
- b. Kunnen we voorspellen wanneer er onderhoud nodig is?
- c. Is te voorspellen wat het verwachte rendement wordt op investeringen in duurzaamheid?

2. Communicatie met klanten

- a. Optimaliseren van de bezetting van de klantenservice bij de woningcorporatie op basis van historische gegevens met betrekking tot klachten.
- b. Mogelijkheden voor het afvangen van eenvoudige vragen door middel van het inzetten van een chatbot.
- c. Analyseren van klantcommunicatie door middel van tekstanalyse om klachten automatisch te rubriceren en clusteren. En hier aanvullende analyses op doen om zo de klachtafhandeling te verbeteren en eventueel nieuwe klachten voor te zijn door social media in de gaten te houden.

3. Optimalisatie van processen

- a. Het steeds verder automatiseren van processen bij de aanvraag en afhandeling van huurcontracten kan worden geanalyseerd met behulp van process mining.
- b. Hiermee krijgt de organisatie inzicht in de efficiëntie van deze processen en kan ze zo op de kritieke punten verbeteringen doorvoeren.

- c. Bestaande processen kunnen dus inzichtelijk gemaakt worden op bijvoorbeeld de doorlooptijd en het aantal keren dat een proces first-time-right wordt afgehandeld. Deze KPI's kunnen in een realtime dashboard beschikbaar worden gesteld zodat er een duidelijk en up-to-date beeld bestaat van de kwaliteit van de geleverde diensten.

Big data

Als het over data gaat, kom je vaak de term *big data* tegen. Het is goed om te beseffen dat data niet 'big' hoeven te zijn. Het is zelfs goed om klein te beginnen met een eenvoudige, pragmatische vraag. Daarbij is het wel verstandig om niet zomaar 'iets' met data te gaan doen, maar planmatig te werk te gaan.

3 Datascience in de praktijk

3B Wonen heeft verschillende vraagstukken beantwoord met behulp van datascience. Hierbij is steeds gebruikgemaakt van de zogeheten survivalanalyse.

Survival-analyse is een statistische methode die wordt gebruikt voor data die informatie geven over de tijd tot het optreden van een bepaalde gebeurtenis, bijvoorbeeld het aantal jaren, maanden of weken vanaf de start. Van oorsprong komt de technologie uit de medische wereld, waarbij de gebeurtenis het overlijden van een patiënt is (vandaar de naam), maar survival-analyse is ook toe te passen op andere terreinen, bijvoorbeeld in de criminologie, in het bedrijfsleven of bij overheidsinstellingen.

Bij 3B Wonen wordt de survival-analyse gebruikt om in verschillende onderzoeken verschillende gebeurtenissen te voorspellen:

- Wanneer treedt een eerste klacht op na een mutatie? Deze analyse wordt in dit whitepaper verder uitgewerkt.
- Welke cv-ketels blijven het langst storingsvrij na installatie?
- Welke factoren maken de kans op mutaties het grootst?

De resultaten van de analyse kunnen worden weergegeven in een survival curve, een grafiek waarin de kans op de gebeurtenis wordt weergegeven op basis van bestaande gegevens.

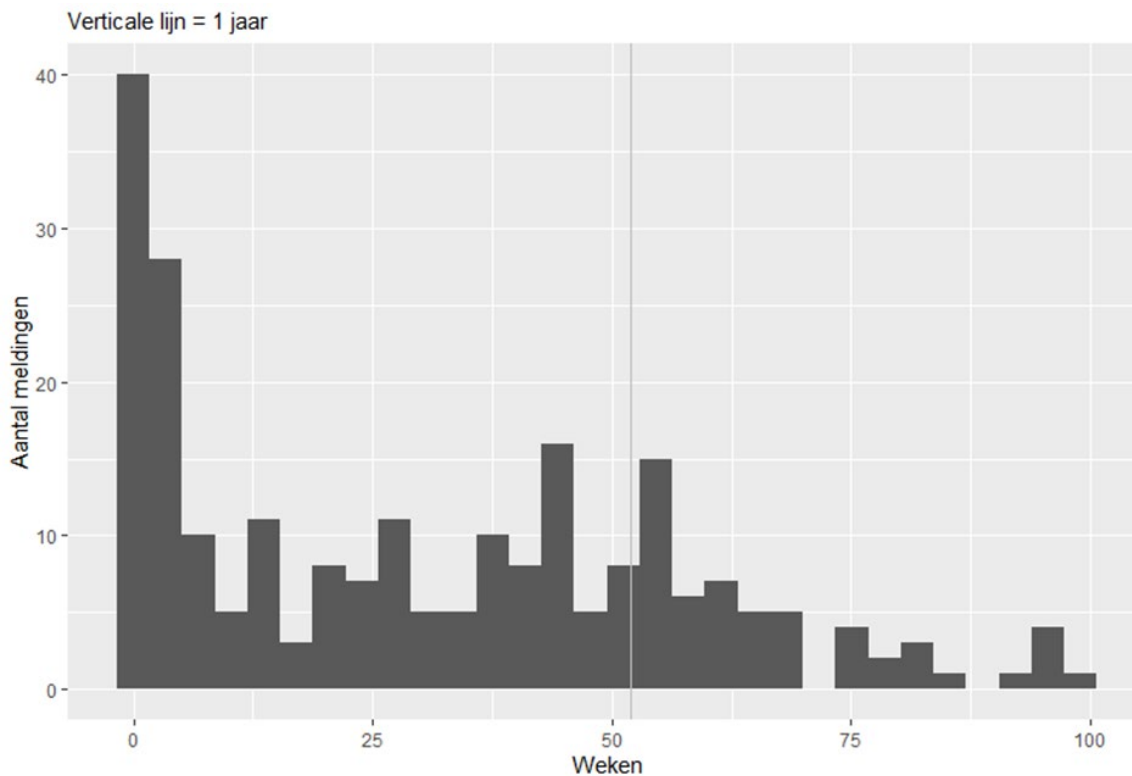
Voorspellen wanneer de eerste klacht optreedt

In de ideale (woningcorporatie)wereld wordt bij een mutatie de vrijgekomen woning opgeknapt en zal de volgende huurder geen klachten melden in of aan de woning. 3B Wonen heeft als doel gesteld te achterhalen of dit klopt door de vraag te onderzoeken: wanneer treedt een eerste klacht op na een mutatie? Op basis van de resultaten van deze analyse kan de woningcorporatie bekijken of klachten te voorkomen zijn of het budget voor onderhoud plannen.

De algoritmes komen voort uit de data van de onderhoudsadministratie van 3B Wonen vanuit de jaren 2019 en 2020. Het gaat om 382 mutaties met 234 eerste meldingen.

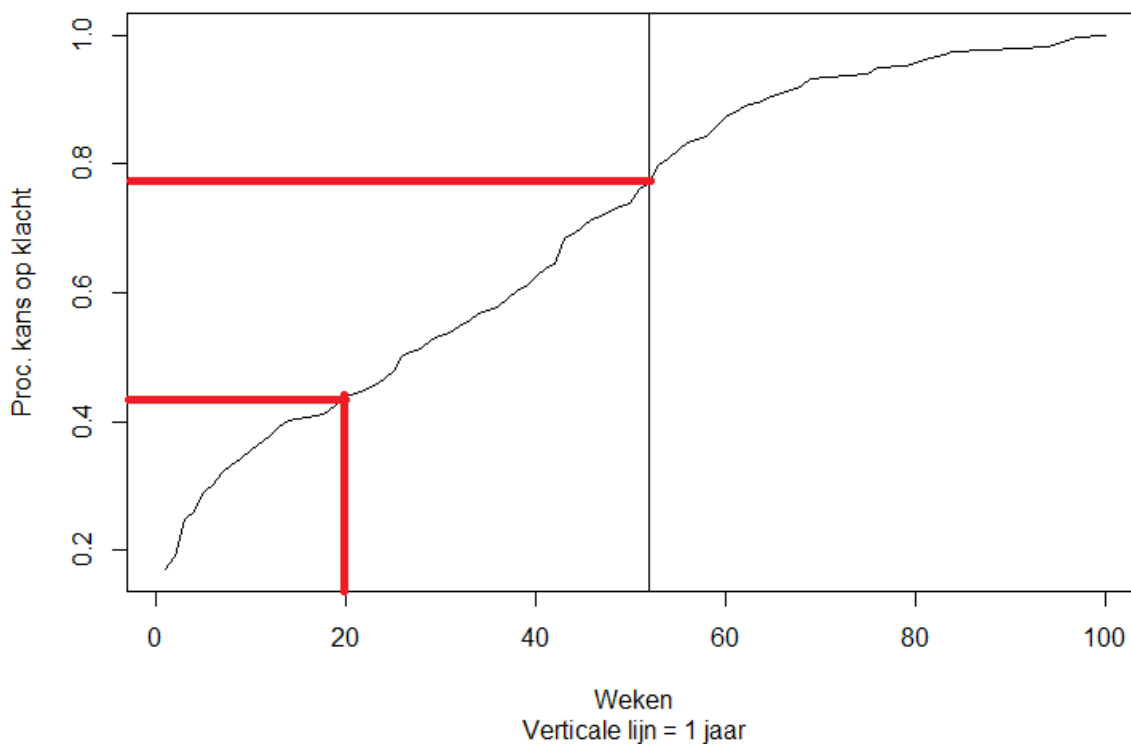
Van de 382 mutaties hebben 234 vhe's (verhuureenheden) een eerste melding binnen twee jaar (61 procent). Als we diverse narooiwerkzaamheden buiten beschouwing laten (opknapwerk dat nog wordt uitgevoerd direct nadat een nieuwe huurder de woning heeft betrokken), hebben 191 vhe's een eerste melding binnen twee jaar (50 procent).

In de volgende grafiek is de eerste klacht na mutatie van de jaren 2019 en 2020 te zien. De klachten (meldingen) zijn per week opgeteld.

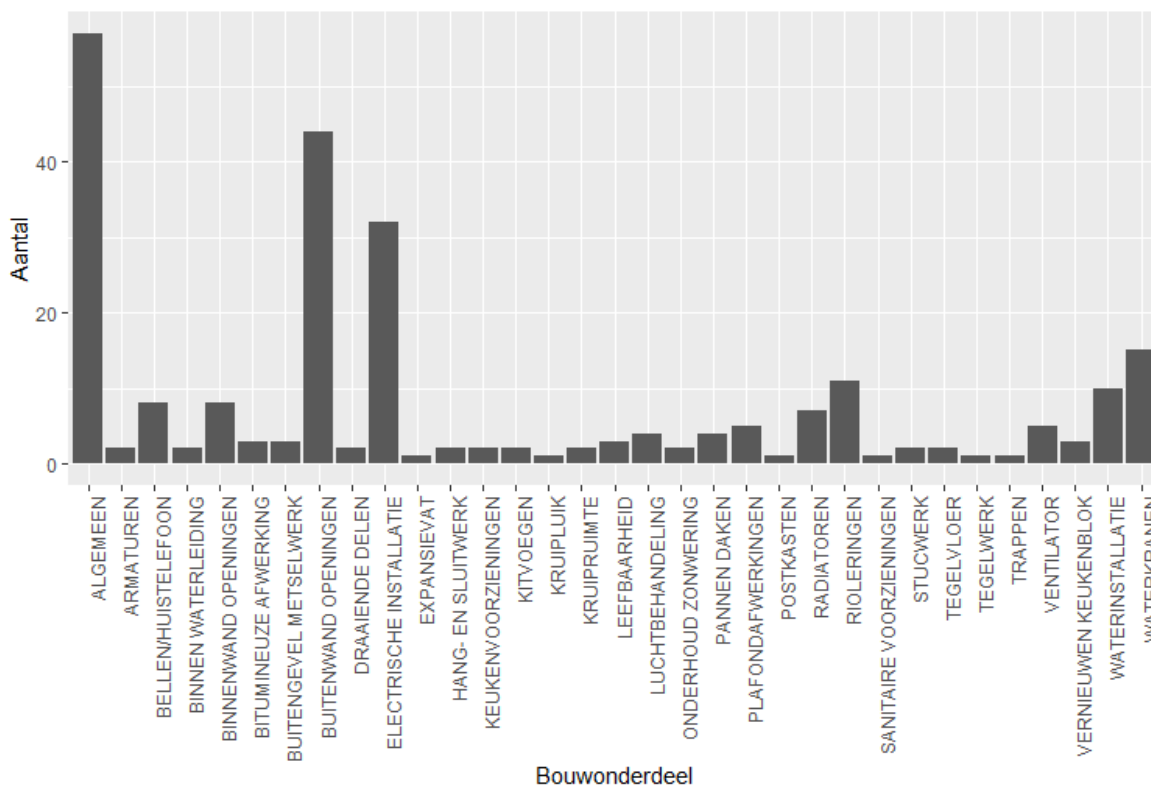
Grafiek 1 Eerste klacht na mutatie

De piek in week 0 is te verklaren door narooiwerkzaamheden nadat het nieuwe huurcontract is ingegaan. Na een jaar (week 52) is een verticale lijn getrokken.

Na een survival-analyse kunnen we een survival curve tekenen. Deze toont de kans op een eerste storing na een bepaalde periode. In de grafiek is te zien dat na 20 weken de kans op een klacht ongeveer 42 procent is, na een jaar is dat ongeveer 80 procent.

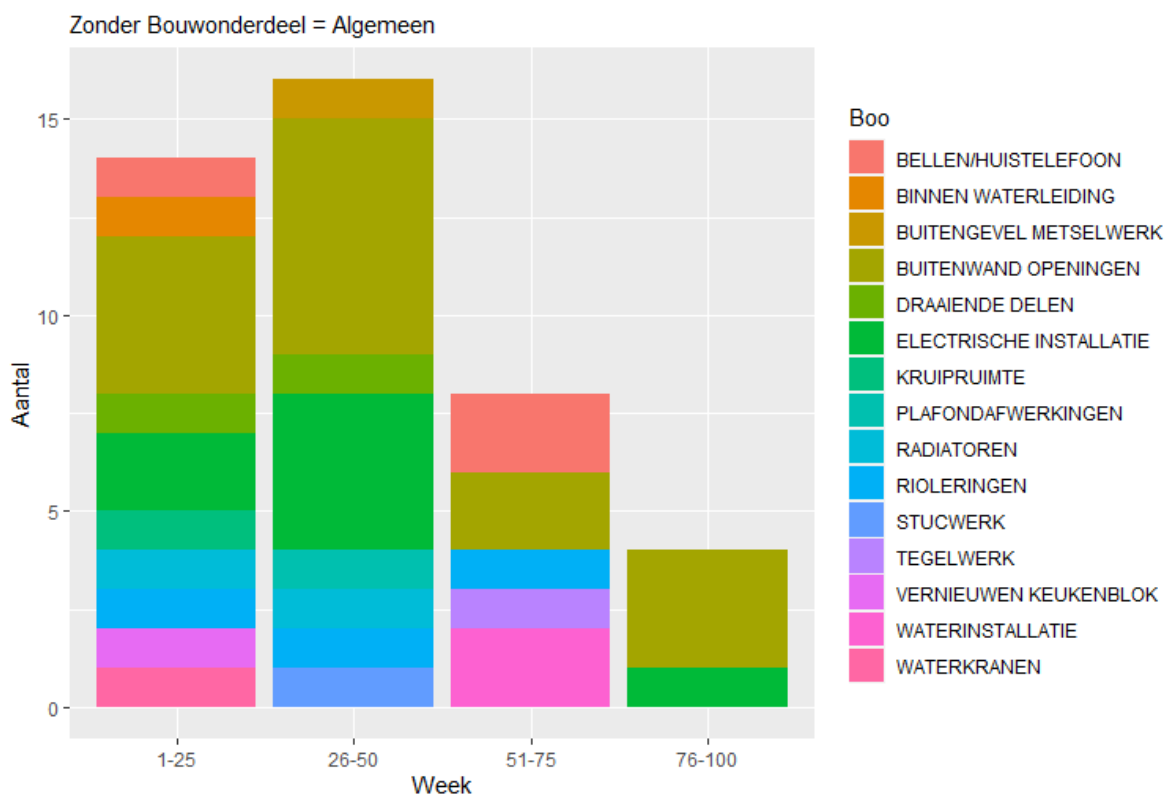
Grafiek 2 Procentuele kans op klacht na mutatie

Onderzoeken wanneer een klacht zich voordoet, is een mooie eerste stap. We willen ook graag weten waar de meeste klachten uit voortkomen. In de onderstaande grafiek zijn de bouwonderdelen zichtbaar die de eerste klacht veroorzaakt.

Grafiek 3 Bouwonderdeel eerste klacht na mutatie

Als we kijken naar het tijdvak in weken waarin de klacht ontstond, kunnen we de volgende visualisatie maken (de narooiwerkzaamheden in de post Algemeen laten we hier buiten beschouwing).

Grafiek 4 Bouwonderdeel per tijdvak



Bij het bekijken van deze visualisatie valt het volgende op:

- In de eerste 25 weken zijn het vooral de buitenwandopeningen (kozijnen, ramen, deuren) die klachten opleveren.
- Van week 26 t/m 50 zijn het vooral de buitenwandopeningen en elektrische installatie.
- In het tweede jaar zijn er minder klachten. De meest voorkomende klachten die zich aan het begin van de huurperiode voordeden, zijn dan opgelost.

Om een dieper inzicht te krijgen in de aard van de klachten kunnen we nog bekijken welke woorden het meest voorkomen in de meldingen.

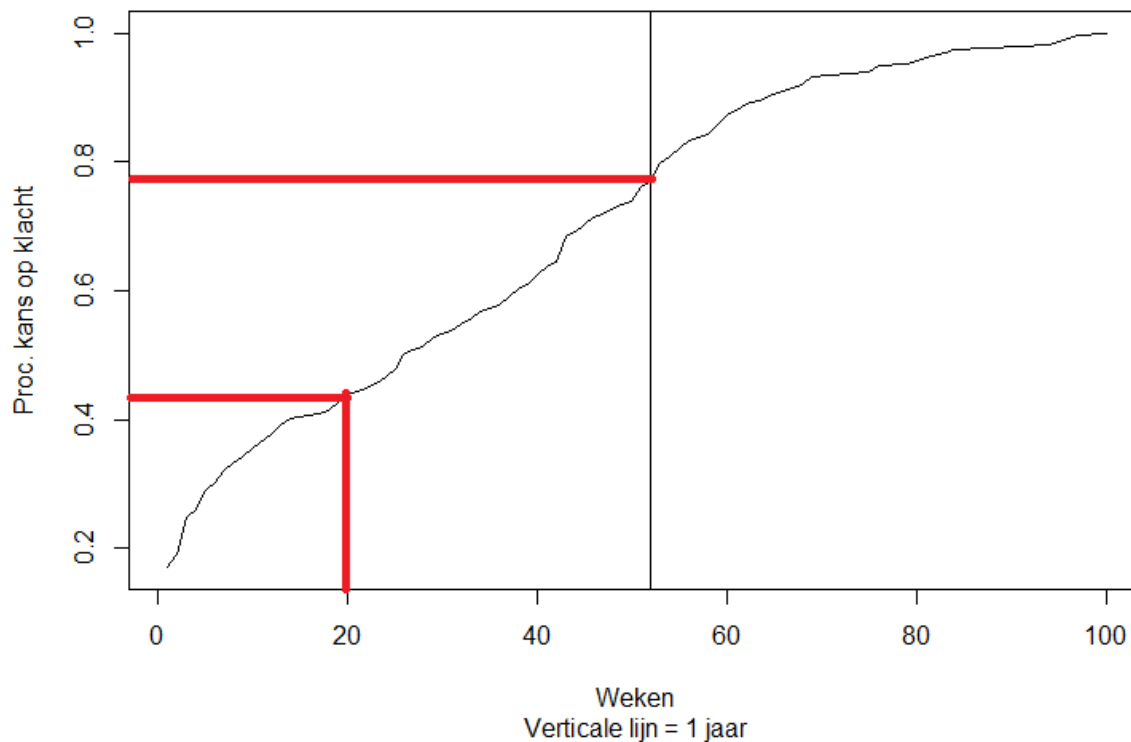
Woord	Aantal
Lekkage	30
Slot	15
Deur	13
Sluitend	13
Kraan	12
Los	12
Vastzetten	12
Raam	9
Sluitwerk	9
Keuken	8
Stroomstoring	8
Tocht	8
Voordeur	8
Intercom	6
Klemt	6
Toilet	6
Ventilatierooster	6
Balkondeur	5
Manchet	5
Stortbak	5
Waterleiding	5
Badkamer	4
Cv	4
Deuren	4
Keukenkraan	4
Lichtschakelaar	4
Reservoir	4
Rookmelder	4

Wat kunnen we uit deze analyse concluderen?

Van de 382 mutaties hebben 234 vhe's een eerste melding binnen twee jaar (61 procent).

Stel dat we 20 mutaties in één week hebben, dan kunnen we verwachten dat:

- na 20 weken ongeveer 5 woningen ($61\% * 42\% * 20$ woningen) een eerste klacht hebben
- na een jaar 10 woningen ($61\% * 78\% * 20$ woningen) een eerste klacht hebben gehad
- de klacht waarschijnlijk betrekking zal hebben op buitenwandopeningen of de elektrische installatie.

Grafiek 5 Procentuele kans op klacht na mutatie

In dit voorbeeld is een relatief kleine dataset gebruikt. Ook zijn alleen gegevens van 2019 en 2020 beschikbaar. Ondanks dat geeft de analyse inzicht in het tijdstip en de aard van de klachten die we kunnen verwachten.

In een vervolgonderzoek kunnen we bekijken wat het oplossen van de verschillende klachten kost en hoeveel duurder het ad hoc oplossen van klachten is vergeleken met oplossen tijdens de renovatie.

In een vervolgstap zouden we kunnen bekijken of we echt de kansen kunnen voorspellen op het moment dat we de renovatie uitvoeren. Dan kunnen we een toepassing maken die de top-3-aandachtspunten voor deze vhe aandraagt tijdens de renovatie zelf, bijvoorbeeld lekkages, kozijnen en deuren.

Andere voorbeelden van datascience

Met de techniek survival-analyse heeft 3B Wonen nog verschillende andere voorspellende onderzoeken gedaan.

Wat voorspelt het aantal mutaties?

De survival curve is ook te gebruiken om te laten zien hoe groot de kans is dat een huurder na een bepaalde periode verhuist. Een aantal interessante conclusies:

- Hoe langer een bewoner in de woning woont, hoe kleiner de kans op verhuizen. Na tien jaar is de kans op verhuizen 40 procent.
- Hoe groter de woning, hoe kleiner de kans op verhuizen.
- Hoe jonger de bewoner(s), hoe groter de kans op verhuizen.
- Wanneer een lift aanwezig is in hoogbouw, is de neiging om te verhuizen kleiner.

Welke cv-ketel blijft het langst storingsvrij?

In een ander onderzoek is bekeken welke van de verschillende geïnstalleerde cv-ketels na installatie het langst storingsvrij blijft.

In deze case is gekeken naar de twee merken met de meeste ketels. Bij deze ketels is gekeken naar de datum van de eerste storing na installatie.

Uit de data blijkt dat merk A een piek in eerste storingen heeft rond het derde jaar en rond het twaalfde jaar. Merk B heeft een piek rond het derde jaar en rond het achtste jaar. Uit de survival curve blijkt dat de kans op de eerste storing in het tiende jaar ongeveer 75 procent is.

Analyse van de twee merken ketels toont aan dat tot circa het zevende jaar de kans op de eerste storing bij merk B kleiner is dan bij merk A.

Hierna is gekeken naar verschillende typen ketels van beide merken. Er is een groot verschil in de periode van de eerste storingen.

Bij de beoordeling van de ketels die zo lang mogelijk storingsvrij blijven na installatie, scoort merk B beter dan merk A. De ketel die het langste storingsvrij blijft, is ook een type van merk B.

4 Meer doen met data

Zoals uit de voorbeelden in het vorige hoofdstuk blijkt, kunnen woningcorporaties met datascience heel nauwkeurig voorspellingen doen. Daardoor zijn processen beter in te richten, is geld te besparen en kunnen klanten beter worden bediend.

Door zelf meer te investeren in de kennis van datascience worden corporaties minder afhankelijk van externe bureaus. Bovendien wordt de branche aantrekkelijker voor jongeren die een studie in datascience hebben afgerond en daarmee nieuwe energie in de sector brengen.

Inmiddels heeft 3B Wonen een kant-en-klaar template ontwikkeld rond het voorspellen van mutatiegraden. Door de resultaten op een open platform aan te bieden, kunnen andere woningcorporaties heel snel aan de slag gaan met datascience en machine learning om snel tot inzichten en beslissingen te komen.

Aedes heeft een eigen plek (repository) op het open-sourceplatform Github. Hierin heeft 3B Wonen de analyses en datascience-scripts in de programmeertaal R beschikbaar gesteld voor het voorspellen van mutatiegraden.

Aanbevelingen van 3B Wonen

Wat te doen en wat vooral niet? Op basis van de ruime ervaring die 3B Wonen met datascience heeft opgedaan, zijn de volgende aanbevelingen te formuleren:

- Zorg dat de data op orde zijn. Data-analyse heeft alleen zin als de data betrouwbaar zijn. Hoe meer data je hebt, hoe betrouwbaarder de inzichten die je uit die data kunt halen. Zorg ook dat voor iedereen duidelijk is wat je precies gaat onderzoeken en definieer je data duidelijk.
- Creëer draagvlak. Onderzoek waar de organisatie behoefte aan heeft. Het koersplan, waarin de doelen van de corporatie duidelijk geformuleerd zijn, is daarbij leidend. Neem interne klanten mee in de zoektocht naar antwoorden op vragen.
- Zorg dat de uitkomsten uit te leggen zijn aan je collega's. Maak de conclusies niet te ingewikkeld met lastig te doorgronden modellen en grafieken. Maak visualisaties die begrijpelijk zijn.
- Zorg voor een klankbord, eventueel extern. Het is belangrijk om te sparren over oplossingen. Vaak ben je blind voor je eigen fouten.
- Begin klein en pragmatisch. Een klein succes geeft energie om verder te gaan.

Tip: Community Digitalisering en Informatievoorziening

Voor corporatiemedewerkers die zich met digitalisering en informatievoorziening bezighouden heeft Aedes een community. In deze [community](#) delen we inspirerende voorbeelden en innovaties met elkaar en ontwikkelen we kennis.

Bijlage

Woordenlijst

Datascience	Wetenschap met het doel om kennis en inzichten uit data te halen. Die data kunnen uit allerlei bronnen afkomstig zijn en kunnen gestructureerd of ongestructureerd zijn.
Algoritme	Een reeks instructies die een computerprogramma uitvoert om een taak uit te voeren, vergelijkbaar met een recept.
Kunstmatige intelligentie (of artificial intelligence, AI)	Het vermogen van een systeem of machine om gegevens juist te interpreteren, van die gegevens te leren en om die lessen te gebruiken om specifieke doelen te bereiken.
Machine learning	Het vermogen van een computerprogramma om zelf te leren, zonder dat het expliciet wordt geprogrammeerd. Hij kan zichzelf trainen en soms patronen en inzichten ontdekken die mensen niet hadden gevonden. Machine learning is een onderdeel van kunstmatige intelligentie.
Process mining	Een methode om de automatisering van processen te analyseren. Daardoor ontstaat inzicht in de efficiëntie van processen.
Survival-analyse	Een statistische methode die wordt gebruikt voor data die informatie geven over de tijd tot het optreden van een bepaalde gebeurtenis, bijvoorbeeld het aantal jaren, maanden of weken vanaf de start.

