

# Inzicht in instroom

ROC Friese Poort heeft de ambitie om meer datagedreven te gaan werken om wendbaarder te zijn en beter aan te sluiten bij de omgeving. Ervaring en onderbuikgevoel zijn goede raadgevers, maar er is behoefte om aannames met onderzoek te toetsen of onderbouwen. Beschikbare data omzetten naar zinvolle informatie om de kwaliteit van het onderwijs te verbeteren, is waar het om gaat.

ROC Friese Poort richt zich binnen het project op de instroom van nieuwe studenten. De stip op de horizon is om een kwaliteitsslag te maken op de manier waarop nieuwe studenten binnenkomen en begeleid worden. Bovendien stellen toeleverende scholen vaak de vraag: hoe doen onze studenten het? Natuurlijk kunnen docenten en begeleiders daar met hun praktijkervaring antwoord op geven. Maar hoe mooi is het om op basis van data te vertellen hoe het verloop van groepen studenten aan het mbo is, zodat overdracht en begeleiding daarop aangesloten kunnen worden?

De onderzoeksvraag luidt als volgt: hoe ziet de instroompopulatie van ROC Friese Poort er uit en wat maakt dat studenten in het eerste studiejaar switchen of uitvallen? ROC Friese Poort is nieuwsgierig of het van tevoren bezoeken van een open dag hier impact op heeft. Elisabeth Leenstra, medewerker marketing & communicatie bij ROC Friese Poort vertelt: "De vraag is heel bewust breed opgepakt met één kartrekker en diverse disciplines in de wagen. De combinatie van mensen uit onderwijs, marketing, informatiemanagement en studentenadministratie levert enthousiasme op en spin-off naar de rest van de organisatie."

Een flinke (gepseudonimiseerde) dataset met allerlei kenmerken over studenten, het onderwijs en de regio is de basis van het onderzoek. Met behulp van machine learning zijn twee modellen gemaakt. Eén model om voorspellingen te doen over switchen en uitvallen in het eerste jaar en één model om studenten die in het model gelijk gedrag vertonen, in te delen in clusters. De clusterindeling is ontwikkeld op basis van de variabelen die in het eerste model naar voren kwamen als effectieve voorspellers van uitval of switch. Daarnaast is een uitgebreide analyse van de instroompopulatie gemaakt.

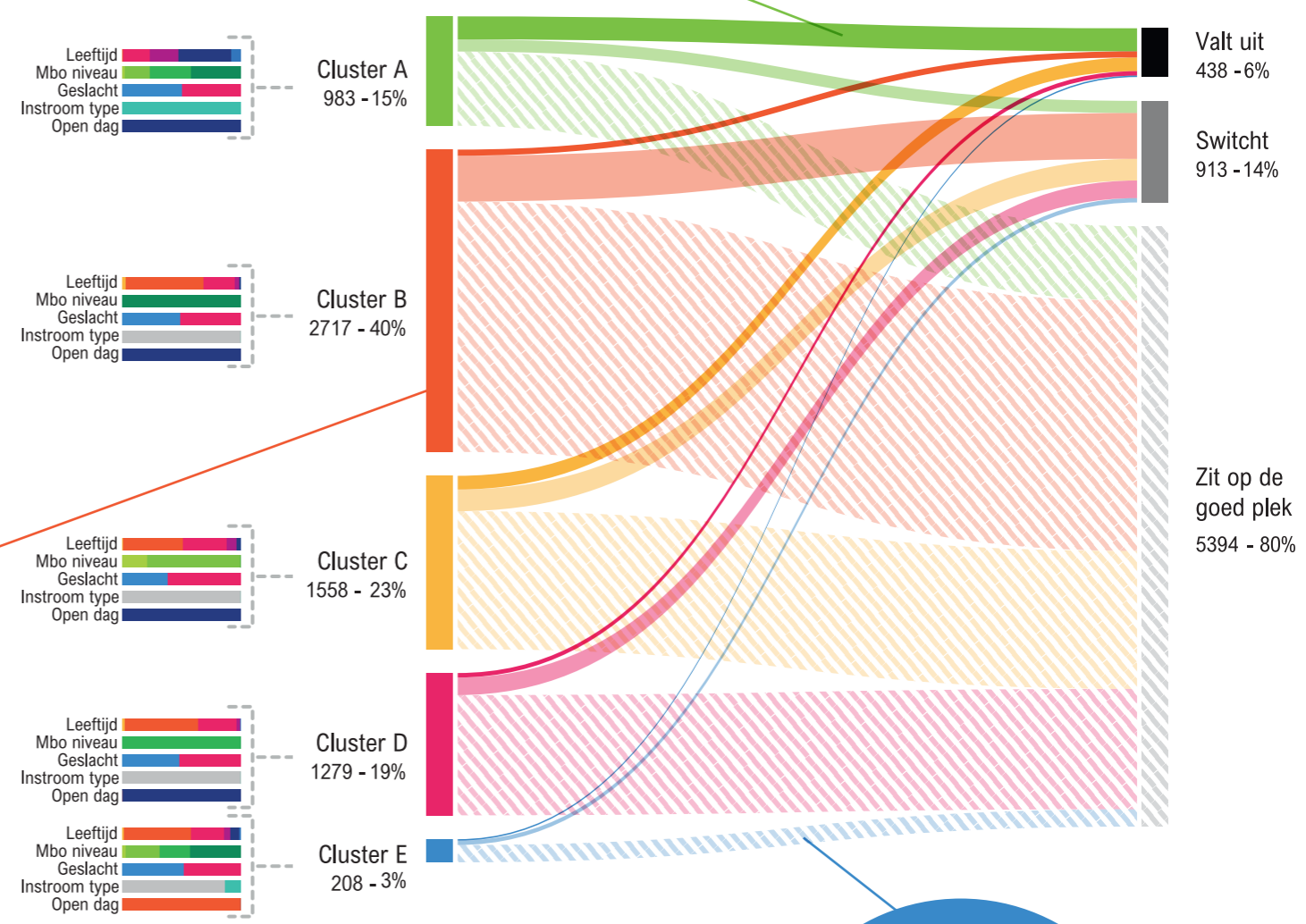
Van alle factoren die zijn meegenomen in de data analyse, bleek het bezoeken van een open dag geen doorslaggevende factor in het model. Dit is grotendeels te wijten aan de kwaliteit van de data die daarover wordt verzameld. Goed nadenken over welke data op welke manier moet worden vastgelegd om van waarde te zijn in de toekomst, is één van de lessen die het onderzoek heeft opgeleverd.

De resultaten zijn gevisualiseerd in een zogenaamde Sankey-diagram. Een Sankey-diagram toont "stromen". Elke stroom bestaat uit studenten die hetzelfde gedrag vertonen. In dit geval studenten die uitvallen, switchen van opleiding of juist op hun plek zitten in het eerste studiejaar. De grootte van de stroom symboliseert het aantal studenten. Daarnaast geeft het de lezer direct inzicht in de verhouding tussen de clusters: de drie stromen van een cluster bij elkaar is het totaal aantal studenten in dat cluster. Cluster E is in het geval van ROC Friese Poort heel klein. Het is zelfs kleiner dan het aantal uitvallers in Cluster A. Het helpt om de aandacht en focus te leggen op de groepen waar de impact het grootste is.

De diagram toont studentenstromen en hun werkelijke gedrag in het verleden (cohorten 2016-2017 en 2017-2018). Het machine learning model dat aan ROC Friese Poort is opgeleverd is een voorspellend model. Dit model levert inzichten op die iets zeggen over toekomstige studentenstromen.

Cluster B is de grootste groep en doet het ten opzichte van de rest goed. Het cluster bevat weinig uitvallers en een gemiddeld aantal switchers. Het bestaat uit studenten die doorstromen uit het VO naar een niveau 4 opleiding. Deze studenten zijn gemiddeld 16,4 jaar en hebben een korte reisafstand naar de instelling.

Cluster A doet het minder goed. Het cluster bevat de meeste switchers en uitvallers. Dit cluster bestaat uit zij-instromers. De gemiddelde leeftijd is 19,5 jaar en de gemiddelde reisafstand naar de instelling is 6 km verder dan die van studenten in cluster B.



In cluster E zitten de studenten waarvan bekend is dat ze een open dag hebben bezocht. Het betreft vooral jonge instromers vanuit het VO op een niveau 3 of 4 opleiding. De kans om uit te vallen of te switchen is voor deze groep niet opvallend anders. We weten niet of dit daadwerkelijk zo is, of komt door de beperkte hoeveelheid data die hierover beschikbaar is.

Leeftijd	Mbo niveau	Geslacht	Open dag
16-	1	Vrouw	wel bezocht
16	2	Man	niet bezocht
17	3		
18	4		
19-23			
23+			

Instream Type
vanuit VO
switch