



PilotHub Studiedata en AI

# MBO x CEDA





# Data Analyse Instroom

22 January 2026

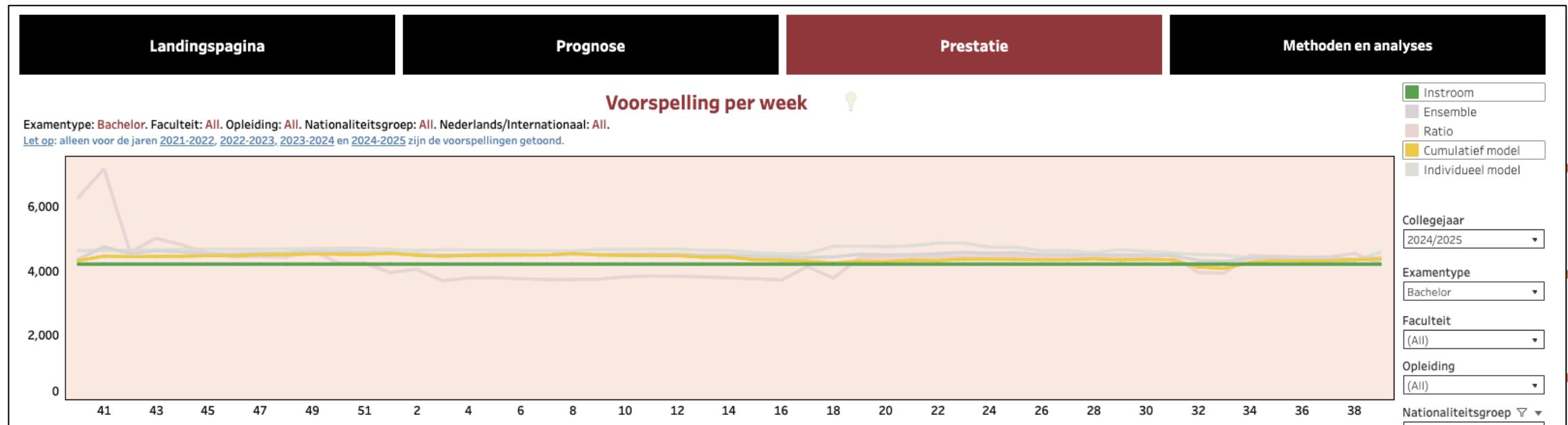


# AI-apps

- Binnen CEDA zijn samen met instellingen al minimum viable products toegankelijk gemaakt.
- In workshops zien we met name enthousiasme over AI-apps.
- Instellingen hebben dit vaak nog niet en hebben ook niet de expertise om dit naar ‘productie’ te brengen.
- Tenminste 2 AI apps uit CEDA zijn rijp voor opschaling:
  1. Instroomprognose
  2. Uitval jaar 1

# Instroomprognose HO

- Op basis van Studielink bestand dat iedere instelling heeft
- Aantal ‘dubbele’ aanmeldingen van studenten wordt meegenomen





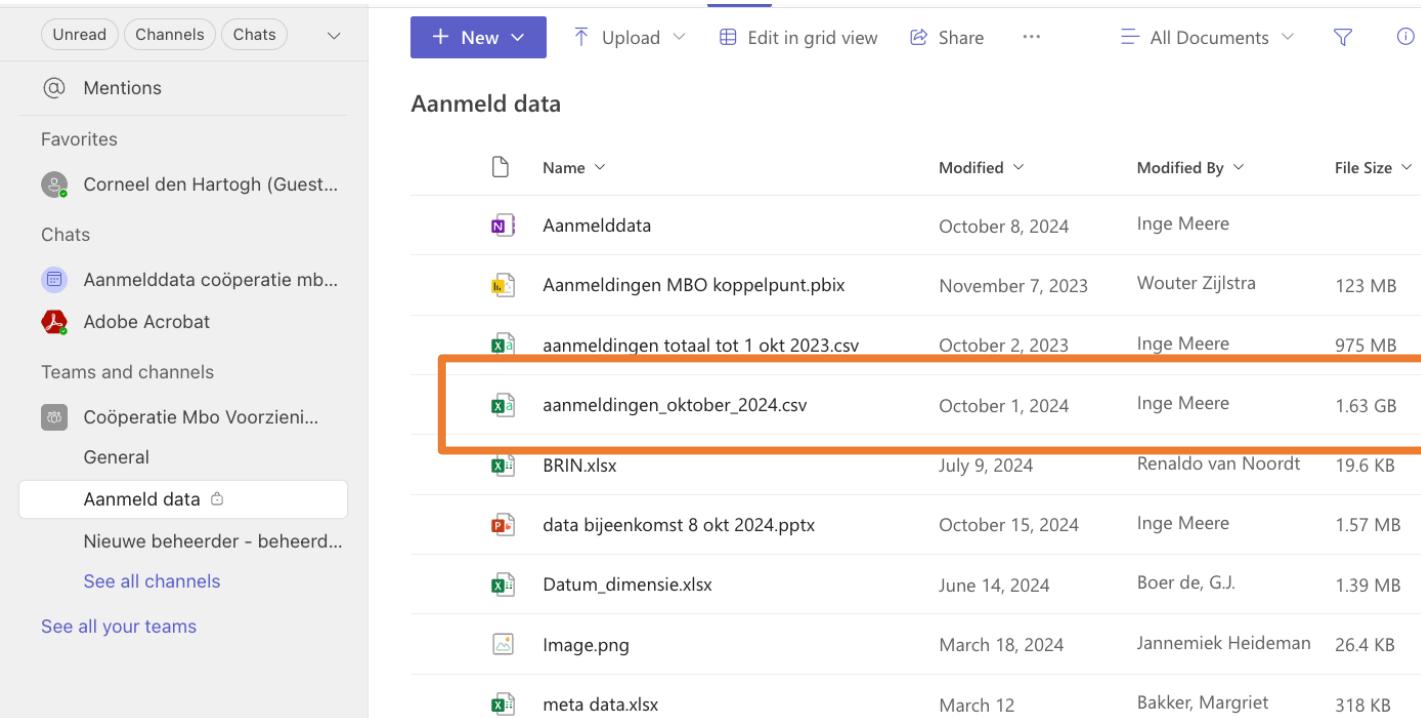
# Koppelpunt MBO

22 January 2026



# Wat is het koppelpunt?

- Veel instellingen laten aanmeldingen verlopen via **CAMBO**
- Iedere instelling is verplicht data te leveren aan het koppelpunt
- Beheer:
  - [i.meere@mbovoorzieningen.nl](mailto:i.meere@mbovoorzieningen.nl)



Name	Modified	Modified By	File Size
Aanmelddata	October 8, 2024	Inge Meere	
Aanmeldingen MBO koppelpunt.pbix	November 7, 2023	Wouter Zijlstra	123 MB
aanmeldingen totaal tot 1 okt 2023.csv	October 2, 2023	Inge Meere	975 MB
aanmeldingen_oktober_2024.csv	October 1, 2024	Inge Meere	1.63 GB
BRIN.xlsx	July 9, 2024	Renaldo van Noordt	19.6 KB
data bijeenkomst 8 okt 2024.pptx	October 15, 2024	Inge Meere	1.57 MB
Datum_dimensie.xlsx	June 14, 2024	Boer de, G.J.	1.39 MB
Image.png	March 18, 2024	Jannemiek Heideman	26.4 KB
meta data.xlsx	March 12	Bakker, Margriet	318 KB

# Wat voor data in HO (Studielink)?

- **Geaggregeerde data per opleiding, instelling, achtergrondkenmerken**
  - Per week geleverd
  - Iedere Hogeschool / Universiteit krijgt data van alle hogescholen / universiteiten
  - Dubbele aanmeldingen zichtbaar
- **Beschikbaar via portal**
- **Zeer stabiele bestanden**
- Levering geaggregeerd vanwege concurrentie en privacy
- [CEDA Voorspellingen repository](#)

# Wat voor data?

- **1 Regel per statusverandering van aanmelding**

- Student (pseudoniem), opleiding, instelling, status, tijdstip, startmoment, bron
- Koppelbaar aan SIS van instelling (keten id)
- Wat achtergrondkenmerken student (postcode4, leeftijdscategorie)
- Iedere ROC krijgt data van alle ROC's

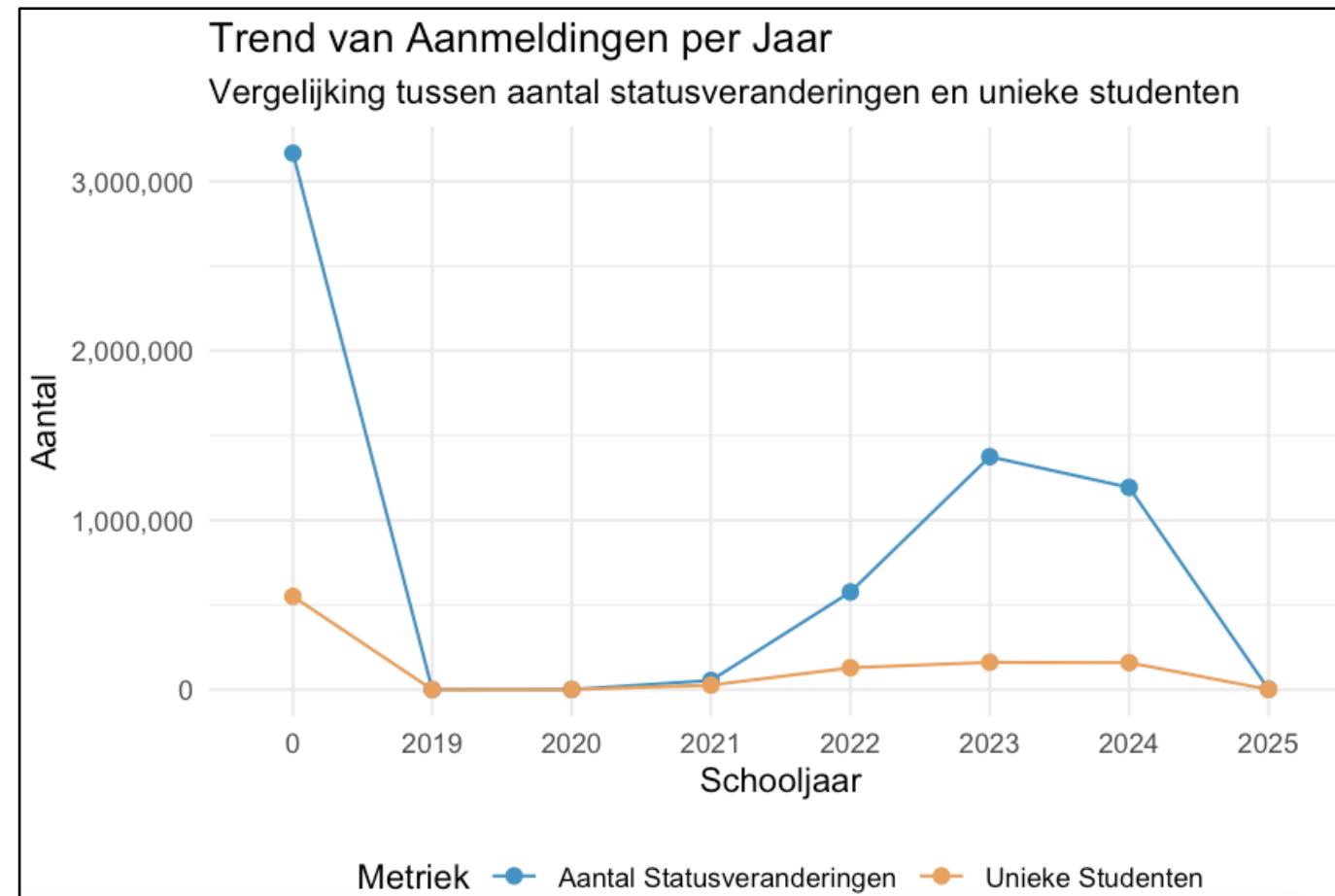
- Levering na startmoment vanwege concurrentie
- Achtergrondkenmerken beperkt vanwege privacy
- Meer data per student aanwezig bij Koppelpunt

# Data kwaliteit en validatie?

- Bestanden zijn nog nieuw en niet geanalyseerd of gevisualiseerd
- Voor alle aanmeldingen van 2023 en 2024 zou data ‘goed’ moeten zijn
  - Geen formele validatie → dit kan ook niet als koppelpunt
- Data kwaliteit is niet absoluut, maar ‘fit for purpose’
  - CEDA/NELA Sessie op Community Dag 8 oktober
  - 6 miljoen regels met codes, datums en lege waardes

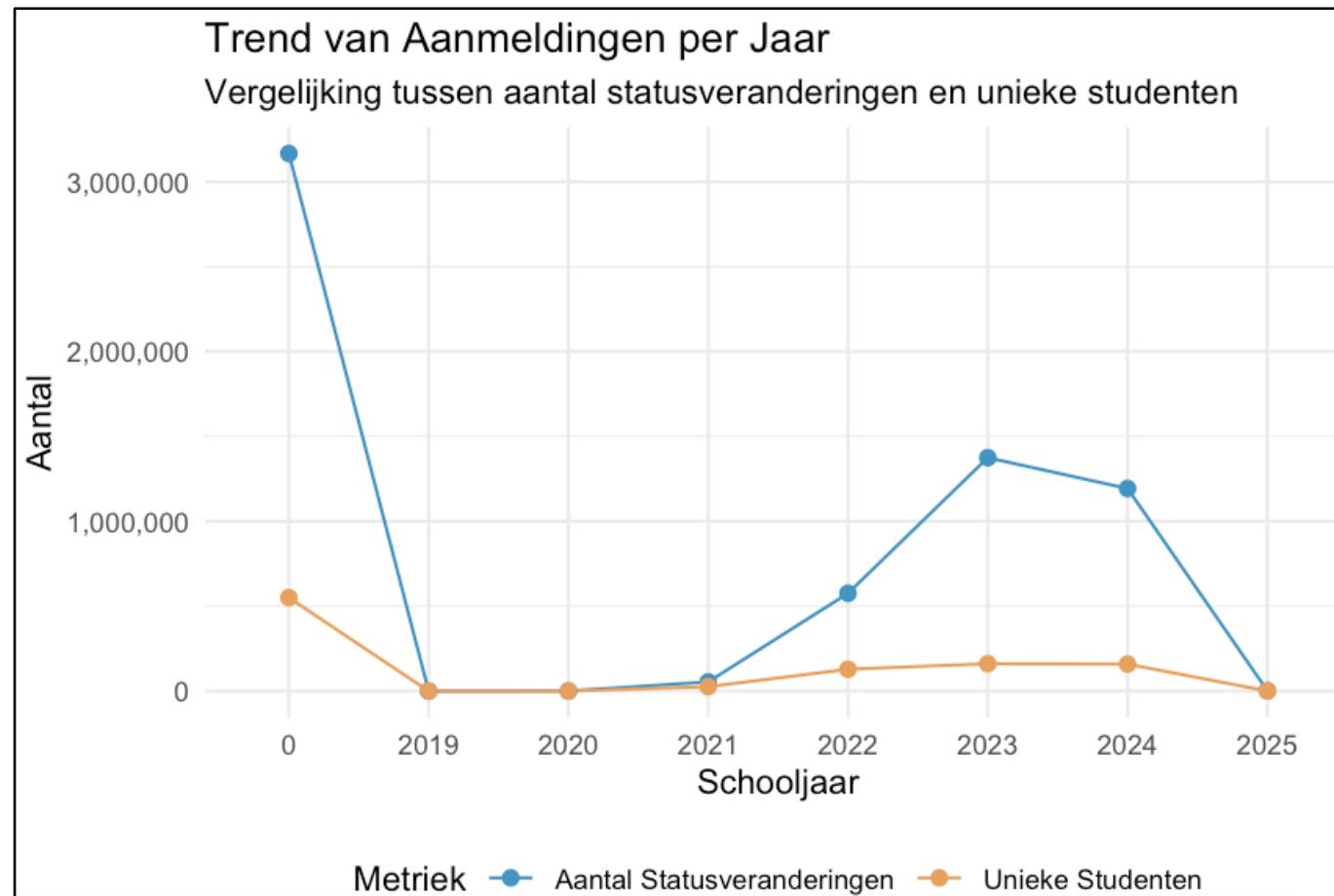
ketenid	bsnhash	leeftijdscategorie	postcodecijfers	instellingserkenningscode	onderwijsaanbiedercode	onderwijslocatiecode	opleidingcode	leertrajectmbo	startmoment	status	begindatum
25LH:90040198	Gx5HboWWq6MCSn2gP4smqs8Uo\$/{KnKgForSqC4Z0V...	TUSSEN_16_18	3703	25LH	100A598	101X218	25297	BOL	2022-08-01	OFFERED	2023-08-17 11:15:46
25LH:90045473	9DxJv9PC3XVG0Ww8cbvXSU8A7LsNGMMk841bb9w8b...	ONDER_16	3438	25LH	100A585	101X218	25104	BOL	2022-08-01	ENROLLED	2023-08-17 11:16:47
27DV:98561381	5RR3YGDvyqfYvnVBLj0588UFoKtrK3Tu0H6NHvZsJU=	TUSSEN_18_23	8271	27DV	NA	NA	25669	BBL	NA	RECEIVED	2023-08-17 11:17:18
25PZ:00291538	thV5g0SakbuowvxrhMGu490xgfADkib5Zj6xq1FaNTQ=	TUSSEN_16_18	3755	25PZ	NA	NA	25182	BOL	NA	OFFERED	2023-08-17 11:17:19
25LH:90082717	ayBzXbSvfIKMEz01pEyMlgXod678Tz49MVAEXP13U=	TUSSEN_16_18	3465	25LH	100A585	101X218	25104	BOL	2022-08-01	RECEIVED	2023-08-17 11:17:48
25LH:90068860	L9Y5KwRqzgw00HHdOXKrBnlaYn/KagMEMh+3f0Hu...	TUSSEN_18_23	3161	25LH	100A585	101X218	25104	BOL	2022-08-01	OFFERED	2023-08-17 11:18:49

# Wat zien we?

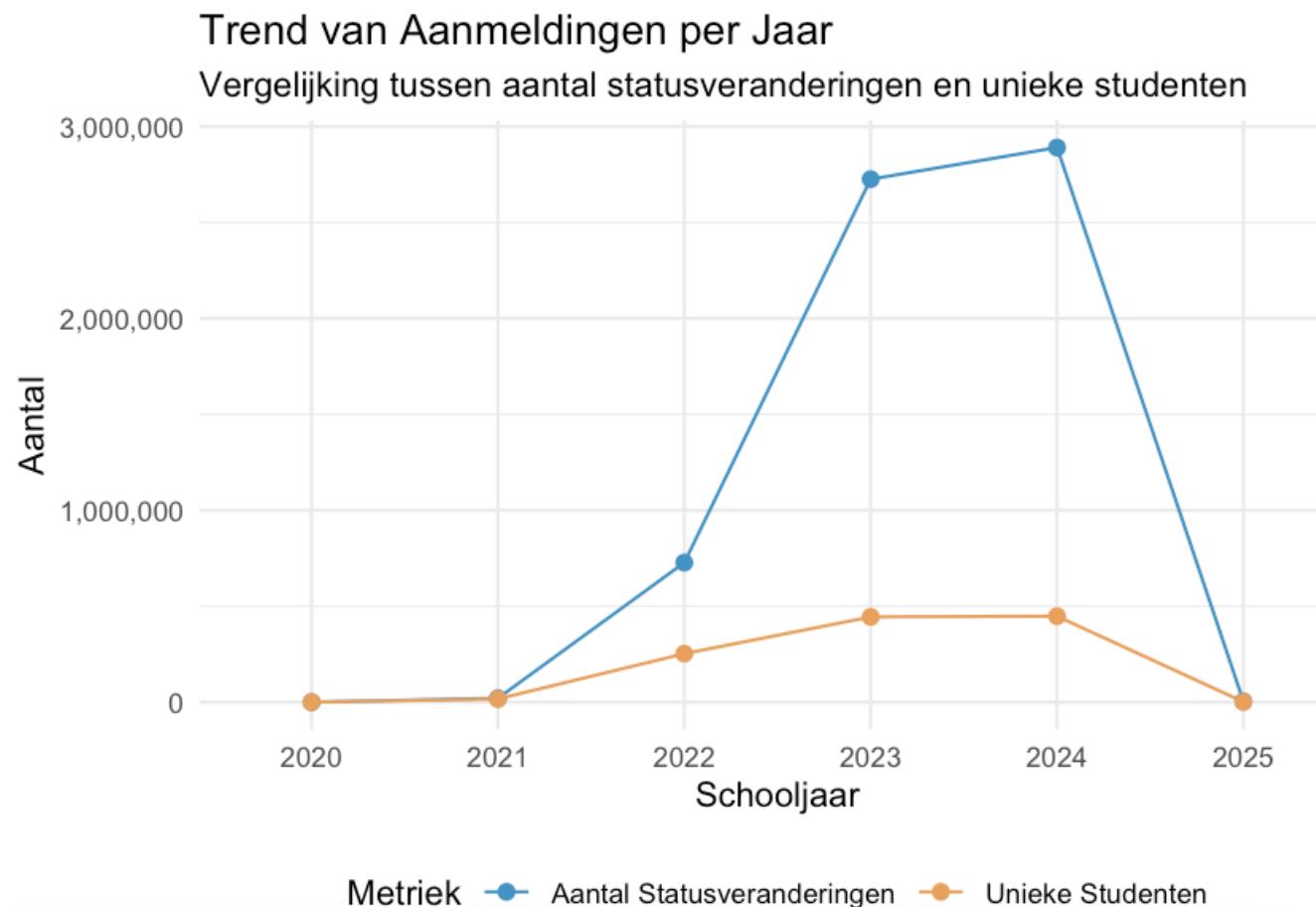


# Wat zien we?

- Jaar 0 scoort goed
- 2024 ongeveel evenveel studenten als 2023
  - minder statusveranderingen
- Suggestie:
  - Obv datum statusverandering en startmoment het schooljaar bepalen

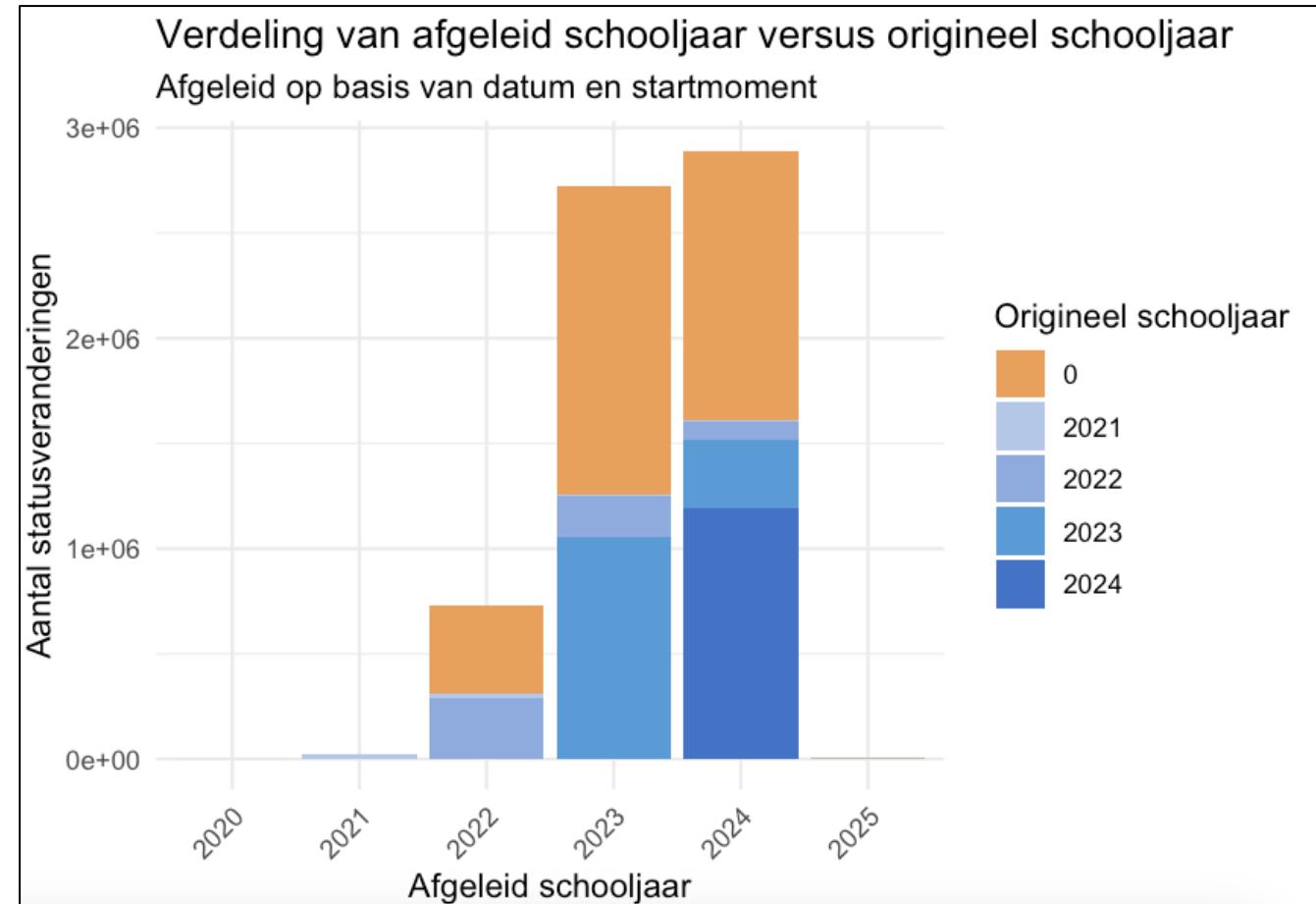


# Wat zien we NU?



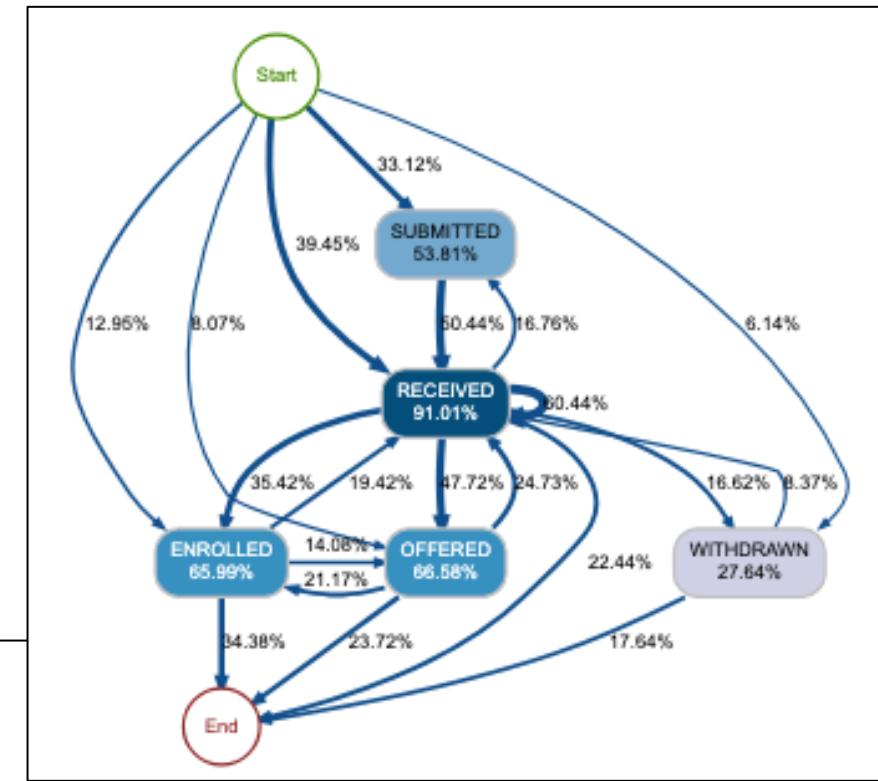
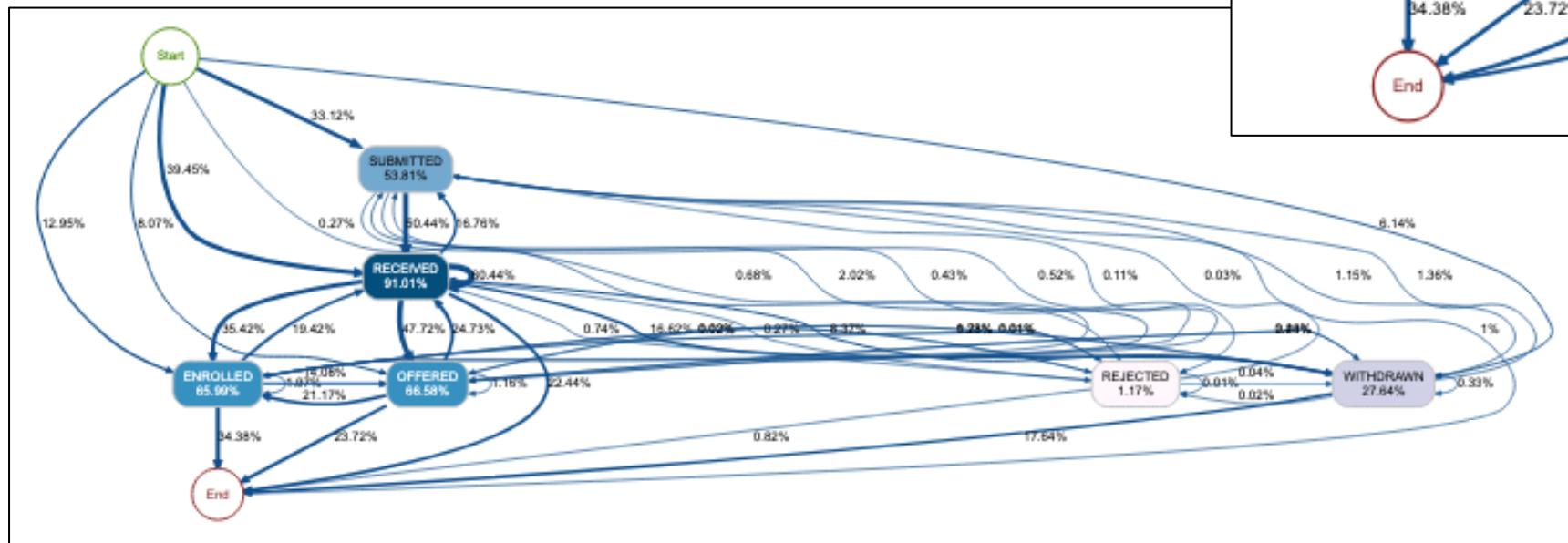
# Wat is er gebeurd?

- 2022 nog steeds zeer afwijkend
- 2024 nu (iets) meer statusveranderingen dan 2023
- Afgezien van jaar 0:
  - Aanmeldingen van voorgaande jaren worden ‘gerecycled’ in daaropvolgende jaren



# Process Mining

- Visualisaties van flow van statusveranderingen
- Conformance checking

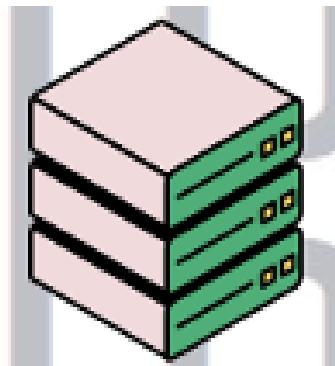


# Fit for purpose → gezamenlijk?

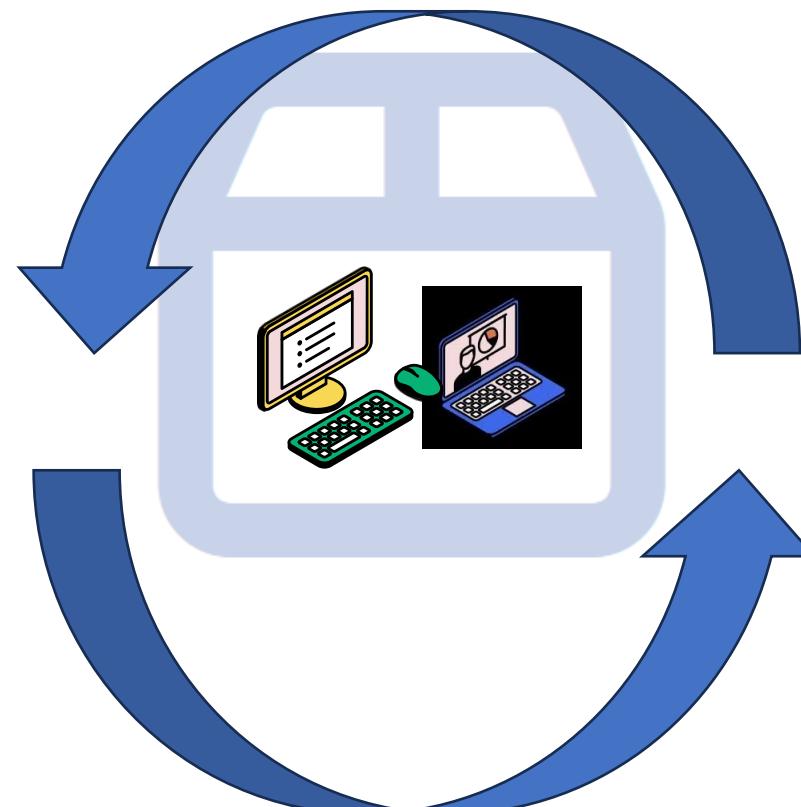
- Kleine andere issues
  - Slechts ~1% afwijking, vaak wordt snel gevuld
  - Postcodes → Kaarten?
- Vergelijking met interne data bij instelling met hoge ‘conformance’
  - Waarde van data zit deels in data kwaliteit andere instellingen
- Voorspellen?

# Data science is iteratief proces

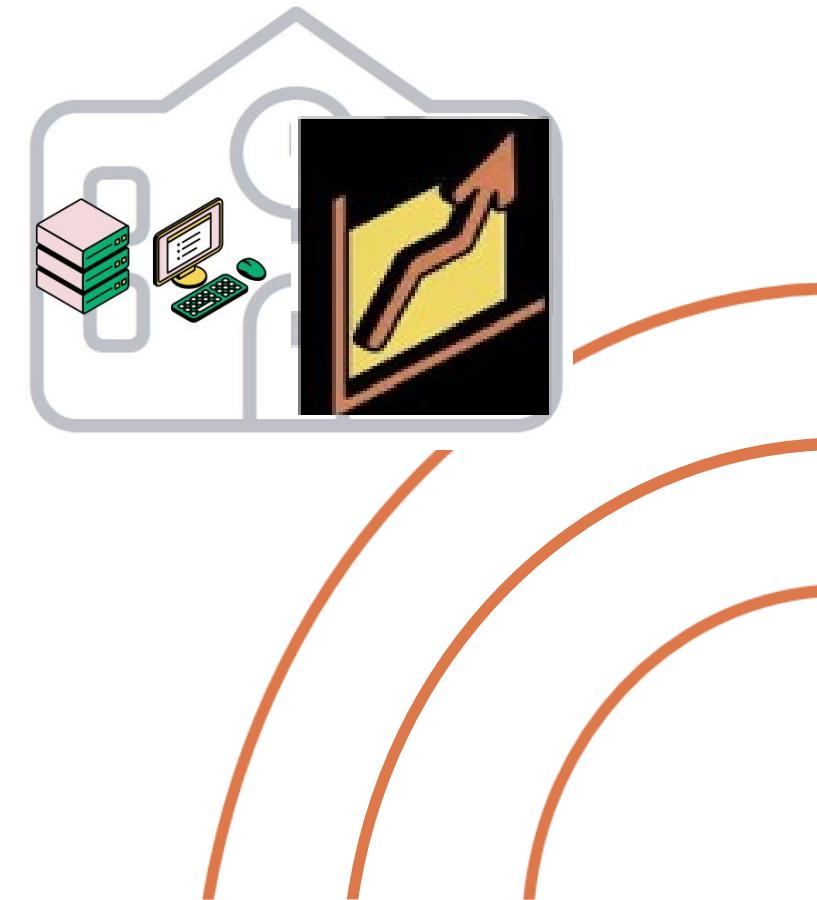
CAMBO, DUO, ..?



CEDA, Analyse Platform



Onderwijsinstellingen



# Contact & Delen



Corneel den Hartogh  
Projectleider CEDA  
[corneel.denhartogh@surf.nl](mailto:corneel.denhartogh@surf.nl)



