

Gijs Ybema

Phone: +31 6 52887539
Email: gijsybema94@gmail.com
Portfolio: gijsybema.nl

 [linkedin.com/in/gijs-ybema](https://www.linkedin.com/in/gijs-ybema)
 github.com/gijsybema

Profiel

Data professional met een achtergrond in statistiek en sociale wetenschappen, en ervaring als Data Scientist, Data Engineer en Kwantitatief Onderzoeker. Ik combineer sterke statistische en data-analytische vaardigheden met praktische software engineering — van data-pipelines tot AI/LLM-toepassingen zoals semantic search — en werk gericht aan mijn door groei richting AI engineering.

Vaardigheden

Programmeertalen en platformen

- **Programmeren:** Python, SQL
- **Python libraries:** Numpy, Pandas, Scikit-learn, statsmodels, keras, requests, BeautifulSoup, psycpg2
- **Statistical software:** SPSS, STATA, MPLUS, R
- **Databases & query engines:** Microsoft SQL Server, Apache Impala, PostgreSQL
- **Data Visualization:** Tableau, Power BI, Python (Matplotlib, Seaborn, Plotly, Streamlit)
- **Andere software/tools:** git (Github), Databricks, Vercel, Railway

Statistiek, Data Science, Machine Learning:

- AI & LLM's: embeddings & semantic search, prompt engineering, RAG, OpenAI API, LangChain, ChromaDB (vector database), Claude Code
- Supervised Learning: Classification and Regression algorithms (Regression, Logistic regression, SVM, Random Forest, Neural Networks)
- Unsupervised Learning: Clustering (k-means), Dimensionality Reduction (Principal Component Analysis)
- Statistics: Multivariate regression (linear, non-linear, logistic), ANOVA, Structural Equation Modelling, Longitudinal Analysis (panel analysis), Multilevel Modelling
- Overig: Natural Language Processing (Spacy, nltk), Time Series (ARMA, ARIMA, LSTM)

Werkervaring

Nuffic

Den Haag

Data Engineer / Data Coordinator

jul 2025 - heden

Als Data Engineer / Data Coordinator bij Nuffic ben ik verantwoordelijk voor het voorbereiden van onderzoeksdata voor het onderzoeksteam binnen Nuffic, en voor de coördinatie van alle data-gerelateerde processen binnen het onderzoeksteam.

Verantwoordelijkheden:

- Voorbereiden van onderzoeksdata voor onderzoekers binnen Nuffic met Alteryx en Python. Hierbij coördineer ik een andere data engineer.
- Bewaken van datakwaliteit, governance en privacy van onderzoeksdata, en planning van dataverwerking.
- Verbeteren van datakwaliteit door geautomatiseerde datachecks en een vierogenprincipe in te voeren.
- Verbeteren van het proces door bestaande Alteryx-workflows te herschrijven, en documenteren van proces via een wiki, inclusief databronnen, datakwaliteitseisen en workflow-standaarden.
- Coördinatie van alle publiek beschikbare Tableau dashboards van Nuffic: borgen dat eigenaren de periodieke update uitvoeren en dashboards voldoen aan kwaliteits- en vormgevingseisen. Hiervoor

een vast proces ontwikkeld: eigenaarschap per dashboard, uniforme procedures, test- en acceptatiestap voor livegang.

Kwantitatief Onderzoeker

mrt 2023 - jun 2025

Als Kwantitatief Onderzoeker bij Nuffic was ik verantwoordelijk voor het uitvoeren van kwantitatief onderzoek, het schrijven van onderzoeksrapporten en het leveren van beleidsadvies aan interne en externe stakeholders in de overheid en het hoger onderwijs.

Verantwoordelijkheden:

- Uitvoeren van kwantitatief en kwalitatief onderzoek en het schrijven van verschillende onderzoeksrapporten (zie Publicaties).
- Samenwerken met interne collega's en externe stakeholders bij het ministerie van Onderwijs, Cultuur en Wetenschap en onderwijsprofessionals in het hoger onderwijs, om onderzoeksvragen te definiëren, geschikte onderzoeksmethoden te bepalen en beleidsadvies te leveren.
- Projectleider van verschillende onderzoeken. Verantwoordelijk voor de planning, uitvoering en coördinatie van het projectteam om onderzoeksdoelen te behalen en resultaten tijdig en effectief te presenteren aan stakeholders.
- Data verzameling, data cleaning, statistische data-analyse en data visualisatie met Python, Tableau en Excel.
- Toepassen van Machine Learning voor het ontwikkelen van prognoses voor het aantal inkomende internationale studenten met regressie-algoritmen in Python (statsmodels, scikitlearn, keras).
- Ontwikkelen van dashboards in Tableau om onderzoeksresultaten visueel begrijpelijk te presenteren en online toegankelijk te maken voor stakeholders.

Decades Collective

Amsterdam

Cofounder

nov 2024 – sep 2025

Decades Collective is een creatieve community in Amsterdam die zich richt op persoonlijke ontwikkeling. Als cofounder was ik verantwoordelijk voor het ontwikkelen van strategische plannen voor de groei van de organisatie, en het beheren van operationele, juridische en financiële processen.

Aon

Rotterdam

Data Scientist

mrt 2019 – feb 2023

Als Data Scientist bij Aon was ik verantwoordelijk voor het toepassen van geavanceerde data-analysemethoden en Machine Learning om waardevolle inzichten te verkrijgen en strategische beslissingen te ondersteunen.

Verantwoordelijkheden:

- Adviseren van stakeholders in management, marketing, customer experience en business development met inzichten op basis van data analyse die helpen bij het nemen van datagedreven beslissingen.
- Verantwoordelijk voor data verzameling, data cleaning, data analyse, data visualisatie met SQL, Power BI, Python en Excel.
- Toepassen van Machine Learning voor voorspelling van customer churn met classificatie-algoritmen in Python (scikitlearn), en het leveren van datagestuurde aanbevelingen aan het marketingteam voor gerichte retentiecampaagnes.
- Toepassen van Machine Learning en Natural Language Processing (NLP) voor topic classification en sentiment analysis van klantfeedback, met als doel klantgerichte processen te verbeteren en de klantervaring te optimaliseren.
- Samenwerking met business stakeholders om te bepalen welke meetwaarden belangrijk zijn voor rapportages, en deze vertalen naar relevante datasets en dashboards.

- Automatiseren van het creëren en bijwerken van datasets die gebruikt worden in Power BI dashboards, door middel van Python- en SQL-scripts, wat de efficiëntie en tijdigheid van het rapportageproces heeft verbeterd.

Working Talent

Landsmeer

Big Data Trainee

okt 2018 – okt 2019

- Training (met certificering) in verschillende Data Science tools, zoals R, Python, SQL, Power BI, Machine Learning
- Gedetacheerd bij Aon als Data Scientist

Universiteit Utrecht

Utrecht

Onderzoeksassistent/Onderwijsassistent/Programme Advisory Committee

sep 2015 – jun 2017

- Bijdrage aan verschillende onderzoeken binnen afdeling Sociologie. Voorbereiding en data collectie van grootschalig veld-experiment. Data preparatie en data analyse met SPSS en STATA
- Instructies en begeleiding computer labs over SPSS aan masterstudenten Sociologie
- Vertegenwoordiging masterstudenten in besluitvormingsproces universiteit. Adviseren universitair docenten over verbetering van onderwijs.

Projecten

TechTracker

Dagelijkse price-tracking pipeline (Python, PostgreSQL) met een Next.js-frontend, AI-gegenereerde productbeschrijvingen en semantic search op basis van embeddings.

Movie Dashboard

Interactief Streamlit-dashboard dat eigen kijkgeschiedenis analyseert en visualiseert (pandas, Plotly, TMDb API).

Opleiding

City University of New York, The Graduate Center

New York City, USA

PhD: Sociology

sep 2017 - jun 2018

Eerste jaar succesvol afgerond (GPA: 3.9/4.0). Gespecialiseerd in kwantitatieve data-analyse, geografische data-analyse en sociologische theorie.

Universiteit Utrecht

Utrecht

Master of Science (MSc): Migration, Ethnic Relations and Multiculturalism

sep 2015 - jun 2017

Cum Laude. Gemiddelde: 8.2/10

Master Thesis: *The First Phase after Arrival. A Longitudinal Approach to the Relationship between the Socio-Cultural and Structural Integration of Recently Arrived Immigrants* (Cijfer: 8.0)

Universiteit Utrecht

Utrecht

Bachelor of Science (BSc): Sociologie

sep 2012 - jun 2015

Overige trainingen en cursussen

Datacamp

- *Certificaten:* [Data Engineer Associate](#), [AI Fundamentals](#), [Python Data Associate](#) (links naar Datacamp)

- *Tracks (selectie):* Developing AI applications, Machine Learning Fundamentals in Python, Natural Language Processing in Python, Python Programming

- *Projecten (selectie):* Building RAG Chatbots for Technical Documentation, Topic Analysis of Clothing Reviews with Embeddings, Planning a Trip to Paris with the OpenAI API

Coursera - Machine Learning (provided by Stanford University)

Infotopics - Tableau Fundamentals

Microsoft

- Querying Data with Transact SQL (Exam 761)
- Analyzing and Visualizing Data with Power BI

Vrijwilligerswerk**WWOOF (World Wide Opportunities for Organic Farming)**

Huyro, Peru

Vrijwilliger op een boerderij in Huyro, Peru. Geholpen met dagelijkse werkzaamheden.

Sep 2022

Hobby's en interesses

Yoga, salsa, literatuur, arthouse/wereldcinema

Publicaties (selectie)Nuffic (2025) - [Vergelijking van prognosemodellen voor diplomamobiliteit](#)Nuffic (2024) - [Incoming degree mobility in Dutch higher education 2023-24](#)Nuffic (2024) - [De invloed van macro-factoren op diplomamobiliteit](#)Nuffic (2023) - [Staying after graduation](#)