



# Dr. Marc Lange

Mathematician, Analyst  
Data Scientist, Generalist

- Hamburg, Germany
- <https://github.com/Vassago1911>
- [marc@lange-data.org](mailto:marc@lange-data.org)

## About me

problem solver  
*analytical, creative, abstract, thorough*

Automation, Prediction, Reporting, Analysis.. no kind of data I can't force into an AHA!, that will work for years even without me :)

fluent in  
*python, sql, latex, zsh, english, german, sets, graphs, categories*

Blender, Numpy, Pandas, ~~MatX~~, Scikit-Learn, (Tensorflow), BERT, Tesseract, Tableau, Matomo, Google Ads, ..

models in  
*data, geometry, dynamical systems, algebra, algebraic topology*

Natural Language Processing, (Realtime) Audio/Video Processing, automated understanding, Chatbots, Revenue Prediction, Budget Optimisation

experienced with  
*developing people, relationships, models, processes, datasets*

team lead experience, consultant experience, lots of experience in educating various people, extensive experience in giving presentations, sales experience

## Skills

- Math, Python, SQL, zsh
- Data Science
- git, cloud, team lead tasks :)

## Education

2005,2011,2015Abitur, Diplom, Doktor (Ph.D.) "magna cum laude"  
 Itzehoe, Hamburg, Hamburg  
 Major in Mathematics, Minors Physics and Computer Science

## Experience

2007-2015 Various Tutoring jobs University of Hamburg

2016-2017 Business Intelligence Analyst AppLike GmbH, Hamburg

2018-2019 Lead Business Intelligence / Data Science AppLike GmbH, Hamburg

- doubled the profit share of daily revenue by a transparent mathematical optimisation  
Revenue Prediction, Lead Optimisation, Margin Optimisation
- presented on a sales event for AppLike's Top10 partners
- joined career and business fairs to approach business and tech applicants as well as business leads
- interim Product Owner for a backend-team in March 2018
- communication center between tech and business: tracking, reporting, predictions, analytics
- implementation with external partners for integrations - e.g. Appsflyer, Singular, King, Playrix, etc.  
Tracking, Reporting, Predictions, Analytics,..
- instructed 10+ classes in Linear Algebra, Set Theory, Mathematics for Physicists
- advised a master thesis in AppLike to synthetically generate fraudulent behaviour (July 2019), informally advised about 5 more (2017-2020)
- trained 5+ (2017-2022) colleagues for durations of either two weeks up to a year in data science methods or business generally as applicable to their interests, skills, expertise, job and tasks
- defined half year goals for Team BI, joined lead offsites
- actually provided the safe space for "my" analyst to be able to hit these half year goals :D

2019- Data Science Consultant elbformat consulting, Hamburg

- interacted with customers at a big union, "Land Schleswig-Holstein", specifically their department of education, a marketing service, some at content management systems, and a few others,
- ...usually providing the single source of truth on how the data science can be done, hence in direct customer management while instructing them on the use of their custom-built analysis, prediction, etc., which are hard to explain products tailored and created for these settings.
- (Re-)Started and intensely pursued my own research, *please see back* :)

## Other information

I love gaming, digital, table top, cards, whatever :) I love math and love sharing math :D generally I love science, from astrophysics down to quanta, from graph algorithms to edges of known complexity theory, but also queer theory, intersectional feminism, disability theory, and many more.

As a mathematician and a general tech enthusiast I am not fixed to specific areas of application. Currently my business expertise mostly stems from a well-scaling business model in mobile marketing for the gaming sector, content management for huge websites (about 20.000 pages), media management (text, links between content, audio, video, etc.). So personally I'm on a research path towards a "General Theory of Data", think "world formula", but from someone who might know :D *More on back, please look* :)

TLDR: I just love figuring out all sorts of stuff :D

**My Research Project** I'm working on understanding a "general theory of data", i.e., I'm conceptualising what EVERY type of data a data scientist could ever come across can be mathematically described as. As example types I have already investigated: Plain text in natural languages like English or German, sequences of user events as in an app or on a website, code in an arbitrary programming language, audio like music, but also street noise, gaming sounds / music, artificially layered 5 source signals, video like from netflix shows or from games, video signals with ui elements like in games, memory dumps of an arbitrary running program on a pc, the sequence of how an arbitrary website is built up in a browser, keylogs, sequences of mouse movements and clicks, webcam and microphone feeds, python scripts that generate a blender image or video,...

In short: All over 2022 so far I've educated myself in processing all this type of data to a certain extent, obviously I have strengths and weaknesses, as well as favourites and dislikes, but, I can do all these things.

Now remember, I'm a PhD in mathematics. Which means, while learning all of this I've been observing the way I am learning these things, and distilling this into a few python scripts each of about 200 lines doing various things like:

"Keylog": Recording Key- and Mouse-Movements realtime, modelling these, hopefully soon automating these with a very general approach.

"MemDump": Recording e.g. one full copy of a fixed program's memory every second, which depending on the program can enable tricks like finding hex addresses of some values of high interest or read out clean text from memory to process like code or language.

"Screengrab": Recording an arbitrary display section realtime (more than 16 fps), doing ETL-style transformations while preserving realtime, soon fitting the text recognition onto this, to find all UI elements automatically.

"Narc": This paper is a quiet revolution <https://arxiv.org/pdf/2106.07688.pdf>, <https://github.com/quantinfo/ng-rc-paper-code>. They model a chaotic very noisy mathematical system with a very elegant reorganisation of its data. Essentially collapsing the need for neural nets onto regressions and a few cleverly chosen features. I'm modeling the "past -> future" transition from keylogs / memdumps / screengrabs data to providing automated robot output just the way this paper is doing it.