Joshua Prettyman, Ph.D.

Data Scientist, Mathematics Ph.D., Python Developer

j.prettyman@icloud.com
linkedin.com/in/joshuaprettyman
tel: 07470 885 347

Personal Statement

I developed Blink SEO's database and internal software from scratch and oversaw its evolution into the ML-powered SaaS product *Macaroni Software*, boasting a 20× increase in SEO productivity. I created the role after being hired as a one-man data/dev team with the brief: 'use data to improve processes'. The team has grown and I have been refining the product in an agile environment. I am looking for a new role where I can implement my data science skills and continue my professional development.

Data Science Skills and Tech Stack

- Regular use of Python libraries: ScikitLearn, Numpy, Pandas, NLTK, MatPlotLib, Plotly, etc..
- Full stack development with Python, Google Cloud Platform, BigQuery, PostgreSQL and JavaScript.
- Huggingface transformers and LLMs, ollama LLM, OpenAI Assistant API.
- CodeCademy pro *Data Scientist* course: ScikitLearn, TensorFlow, PyTorch, NLP and data visualisation.
- Experience C++, shell scripting, Matlab, Retool, data engineering, dashboards, scientific computing.

Professional Experience

Data Scientist — Blink SEO / Macaroni Software *Nov. 2021 to Dec. 2024*

- I delivered Macaroni Software, allowing SEO teams to deliver a year's work in a single month.
- I created this role by working with the team to replicate and then improve their (spreadsheet-based) processes with continual feedback and deployment.
- Macaroni does the (typically manual) data-engineering for SEO by aggregating data from various sources, streamed daily (or on request) to a *BigQuery* database via the python backend (*GCP Compute Engine* — Debian) which utilises various REST APIs and preprocessing steps.
- Client onboarding, data imports, data analysis tasks and ML-assisted recommendations can be triggered in the app and run asynchronously in the backend using a (*PostgreSQL*-based) job-queue system which I wrote for this purpose.
- Quantitative data are used for various visualisations in the app, powered by *Plotly*.
- Text data (on-page content and search terms) are processed and combined with quantitative data (clicks, impressions, etc.) to surface actionable recommendations for new content and site taxonomy. This uses NLP and Clustering algorithms (*NLTK*, *huggingface*, *ScikitLearn*) but also a deep understanding of SEO, necessitating regular exchanges with the SEO delivery team and frequent iterations.
- I also integrated an LLM (*huggingface*) to generate ready-to-go suggestions for content gaps, allowing customers to see immediate impact.
- Besides conceiving of and working on Macaroni, I also did frequent, one-off data tasks for the SEO delivery and management teams including data engineering, content generation, and providing analyses and visualisations for clients and investors.

Data Science Researcher — National Physical Laboratory (NPL) *Sep. 2015 to Feb. 2021*

- I completed my doctoral research in industrial collaboration with Dr. Valerie Livina at NPL.
- I presented data and results for stake-holder meetings at NLP, as well as in published articles and at international conferences.
- My research developed methods to detect tipping points in dynamical systems, working with large time series datasets and building stochastic models to test hypotheses.
- Code written in *MatLab* and *Python*. Visualisations produced using *MatPlotLib* and *XMGrace*.
- Signals in hourly sea-level pressure were used to produce early warning signals for approaching tropical storms. This required the cleaning, interpolating and correct interpretation of large meteorological datasets downloaded (raw) from public sources.

Professional Experience (continued)

Associate Lecturer — Sheffield Hallam University *Sep. 2017 to Jul. 2019*

- I taught mathematics, statistics and computing courses from Foundation Degree to Masters.
- Responsibilities included lesson preparation, course planning, marking, leading tutorials.

Informatics developer (internship) — UK Met Office Informatics Lab *Jun. 2017 to Aug. 2017*

- A three-month project assessing the viability of using live data from amateur meteorologists to augment official observation data.
- This involved accessing internal and external data via APIs using Java, then performing data cleaning and analyses in Python.
- I worked as part of a team and participated in daily stand-ups.

Academic background

Ph.D. Mathematics — **University of Reading** *Feb. 2021*

Thesis: *Tipping Points and Early Warning Signals with Applications to Geophysical Data*.

- My thesis develops spectral analysis methods to detect correlations in multi-dimensional time series, which are used as Early Warning Signals for tipping events in stochastic dynamical systems.
- Publication of three papers in respected journals:
  - *A novel scaling indicator of early warning signals* (EPL, 2018)
  - *Generalized early warning signals in multivariate and gridded data* (Chaos, 2019)
  - *Power spectrum scaling as a measure of critical slowing down* (ERL, 2022)
- Code written in *MatLab* and *Python*. Visualisations produced using *MatPlotLib* and *XMGrace*.
- In many cases I implemented mathematically derived methods —both my own and those of others— numerically by discetising and linearising.
- Participation in several international conferences and workshops.
- Three-month data-focussed internship at the UK Met Office (see *Professional Exp.* section).

M.Res. Mathematics — **Imperial College London** (*Distinction*) *Sep. 2015*

Dissertation: *Mesh Generation Using Solutions of an Optimal Transport Problem*.

- My dissertation uses solutions to a linearisation of the Monge-Ampère equation to generate an adaptive, optimally distributed mesh for numerical models.
- Working in C++; monitoring CPU usage and convergence time over various scenarios.
- Participation in weekly soft skills workshops and monthly internal showcases.
- Attendance at weekly Mathematics of Planet Earth seminars hosted by the Center for Doctoral Training.
- Completed taught courses in Python, R, Dynamical Systems, Statistic differential equations, Bayesian Analysis and Probability.

MA Mathematics — **University of Edinburgh** (*First Class Hons.*) *Jul. 2013*

Dissertation: *Growth and its Applications in Graph Theory*.

- My dissertation explores patterns in paths through connected digraphs.
- For the dissertation project I wrote a program with a simple UI to investigate the properties of these digraphs: this then became a teaching aid, earning a letter of thanks from the principal.
- I completed the requisite taught courses for the Pure Mathematics degree, besides optional modules in computing and mathematics education.