# Greg Lever | Ph.D. (Cantab.)

82 Martin Way, Morden, London SM4 4AG

+44 (0) 7814 676 385 • greglever@cantab.net • greglever.github.io
levergreg • greglever

I am a Data Science and Software professional with extensive Python and SQL experience. I gained my PhD in Computational Enzymology at the Cavendish Laboratory, University of Cambridge, where my thesis was nominated as outstanding and later published by Springer. I then worked as an independent Postdoctoral Research Associate at MIT. I have 18 months of commercial experience in startups building Machine Learning and AI products. I am passionate about continual and lifelong learning.

## Current Position

**Data Scientist (Data Science Developer), Arkera** **Westminster, London**

*A fintech startup at the intersection of Data Science and Wealth Management (established March 2015)* *Dec 2015 – present*

- I was Arkera's first Data hire and was instrumental in building the Data team, adding an additional three members.
- I am involved in the design, prototyping, implementing and maintaining of proprietary Machine Learning (Scikit-Learn/NumPy) and Natural Language Processing algorithms, along with ETL processes in production and our data models.
- Python (Flask/SQLAlchemy/Werkzeug) and SQL (complex queries for MySQL/PostgreSQL) for our backend and web scraping, along with the design of our database schema and dimensional data modelling.
- I am a strong advocate of Test-Driven Development and implemented the company's unit and functional testing framework.
- I played a key role in moving the Data team to an agile development process (stand-ups, sprint planning, code reviews, retrospectives) initially through Trello and then to JIRA.
- RESTful APIs, Amazon SQS event messaging, Docker and continuous integration (Jenkins) and exposure to iOS for the app.
- I perform integrations with third-party API providers and create and maintain in-house tools for our Content Team (javascript).

## Computation and Modelling Skills

**Languages**: *Strong in* Python (Scikit-Learn/Pandas/NumPy/SciPy/Matplotlib/Natural Language Toolkit/Flask/SQLAlchemy/pytest/unittest/pdb), SQL (PostgreSQL/MySQL/BigQuery), Linux (BASH scripting along with awk, sed and grep), R, Git (I encourage rebasing and gladly help others with merge conflicts), NoSQL (MongoDB) and LaTeX typesetting. *Confident with* Javascript, CSS, HTML, Mathematica, Java and the Google App Engine. *Experience of* Neo4j, Cypher Query, Perl, Fortran, C++ and Matlab.

**Operating Systems**: Strong capabilities in UNIX/Linux and High Performance Computing clusters, Mac OS X, and Windows-based environments.

**Certifications**: Medicinal Chemistry (EdX), R Programming (Coursera) and Data Scientist's Toolbox (Coursera).

## Experience

**Senior Data Scientist, Stylect** **WeWork, Liverpool St**

*An e-commerce startup with an app and website designed to enable women to find their perfect shoes* *April 2015 – Dec 2015*

- I headed the Data Science efforts at Stylect, utilising my skills in Python and SQL.
- I led the design, engineering and deployment of production-worthy Machine Learning architectures for Recommender Systems to improve the user experience and ultimately increase sales.
- Involved extensive Python development (web scraping, pytest) with Pandas (including NumPy, SciPy and Matplotlib) and Scikit-Learn.
- Required SQL to use services from the Google Cloud Platform including Big Query and the App Engine.
- Utilised Natural Language Processing (Natural Language Toolkit) to perform Feature Engineering for Machine Learning (Scikit-Learn) approaches for classifying retail products.
- Devised, implemented and interpreted A/B tests to assess the accuracy of the predictive models.

**Data Science Fellow, The ASI** **London**

*The ASI Fellowship enables academic scientists to become data scientists and engineers* *February 2015 – April 2015*

- Building a recommender system for an e-commerce startup (Stylect) to improve their user experience and increase sales.
- Developing strategies for growth hacking and user retention for a community-building startup (Quiet Riots).
- Expert-led training and practical workshops in Hadoop, Spark and MapReduce as well as Statistics and Business-specific concepts.

**Postdoctoral Associate, MIT and University of Cambridge** **Cambridge, MA (USA)**

*MIT Department of Chemical Engineering and Cavendish Laboratory* *March 2014 – February 2015*

- Further developed and galvanised existing skills in Python gained from my PhD.
- Built upon my knowledge of R from my MSci, learning new concepts and tools for analysing results of simulations.
- Mentoring of graduate students, taught undergraduates to use linux and High Performance Computing for the very first time.
- Independent exploratory analysis using Python/Pandas (including NumPy, SciPy and Matplotlib) of an existing body of data.
- Collaborated closely with teams based in the UK (Cambridge, Bristol and Southampton) and the USA (Yale).
- The work is currently being prepared to submit for publication in the *American Chemical Society* journal *Catalysis*.

## Education

**PhD, University of Cambridge** — **Cambridge**

*Computational Enzymology, Cavendish Laboratory* — *September 2010 – March 2014*

- Long-term independent research and analysis from self-driven projects that delivered scientific results and expanded my field.
- Applied mathematical theories to complex real-world problems and generated new methodologies to improve existing approaches.
- Extensive Python (Pandas, NumPy, SciPy) development for data analysis of computational chemistry simulations.
- Immersed in a High Performance Computing Environment utilising linux, BASH, perl, awk, sed and grep.
- Winner of thesis prize "Recognising Outstanding Research", *Springer* subsequently bought my intellectual property from me.
- Published first-name author papers in the *Journal of Chemical Physics* and the *Journal of Physics: Condensed Matter*.

**MSci (First Class Honours), University College London** — **London**

*Theoretical Physics* — *September 2006 – July 2010*

- Through summer projects in 2009 and '10, extended self-taught Linux and Python programming knowledge through real research.
- Contributed to Bayesian Inference software using R and Java and developed extensions in C++.
- Resulted in a publication in *Monthly Notices of the Royal Astronomical Society*, summer work published in *Physical Review B*.
- Courses Highlight: Java and C++, Mathematica, Advanced Quantum Theory, Stochastic Processes, Group Project.

## Extracurricular Activities

**Supervisor for MIT Undergraduate Research Opportunities Program (UROP)** — **Cambridge, MA (USA)**

*MIT Department of Chemical Engineering* — *July 2014 to Oct. 2014*

Provided modelling & simulation training for a candidate who had never previously run any simulations or used the linux terminal.

**Co-supervised Cambridge MPhil Student in Scientific Computing** — **Cambridge**

*Cavendish Laboratory, University of Cambridge* — *October 2012 to Oct. 2013*

Provided electronic structure expertise and biological expertise, to complement the MPhil candidate's additional co-supervisor who provided physics and analytical mathematical modelling expertise. Helped the candidate to prepare for PhD applications.

**Physics At Work - Outreach Event for the Cavendish Laboratory** — **Cambridge**

*Cavendish Laboratory, University of Cambridge* — *September 2011/'12/'13*

Organised exhibitors for and participated in a three day outreach exhibition in Cambridge aimed at inspiring 14-17 year olds to become the next generation of scientists, engineers and technology specialists by showcasing the many and varied ways computational physics is used in the real world.

**Procter & Gamble R&D European PhD Seminar 2013** — **Brussels**

*Procter & Gamble (P&G) Brussels Innovation Centre* — *April 2013*

- Applied technical knowledge to solve real problems found in the typical day of a Scientist or Engineer in R&D at P&G.
- Gained insight into daily business challenges, including how to make difficult and tough decisions, set priorities and communicate.
- Over the week, led a team of PhD students, practising the skills required to motivate people, via exercises and case studies.

**i-Teams: Commercialising Creativity** — **Cambridge**

*Finding real-world applications for a synthetic reversible molecular lego system* — *October 2011 – December 2011*

- Led a team of 6 PhD students and 1 MBA student asked to investigate an invention that came out of a Cambridge laboratory.
- Furthered our ideas for real-world applications by contacting external industry scientists and harnessing their expertise.
- Understood the practical processes needed to turn a new technology into a commercially-viable product.
- Learned how to work with sensitive intellectual property and to bring it to the attention of an existing market.

**Business Challenge hosted by Cambridge University Technology and Enterprise Club** — **Cambridge**

*Building a Business Case for more* in silico *modelling in the Pharmaceutical Industry* — *March 2011*

- Led a team of three fellow PhD students working in close collaboration with the Director of Clinical Pharmacology at Medimmune.
- Had to think about real-world research within a business context
- Gained a firm appreciation for what is expected from data-driven scientists within commercial organisations

**Researcher, London Centre for Nanotechnology** — **London**

*Successfully applied for funding from the Engineering and Physical Sciences Research Council* — *June 2010 – September 2010*

**Researcher, University College London** — **London**

*Successful in funding competition for all department undergraduates* — *June 2009 – September 2009*

References: Prof. Mike Payne (mcp1@cam.ac.uk), Dr. Danny Cole (daniel.cole@yale.edu)

# Greg Lever | Ph.D. (Cantab.)

82 Martin Way, Morden, London SM4 4AG

+44 (0) 7814 676 385 • greglever@cantab.net • greglever.github.io
levergreg • greglever


To Whom it may concern,

I would very much like to be considered for the role of Backend Developer (Python) at Wellcome.

I am currently working as a Data Science Developer at a FinTech start-up named Arkera where I was their first Data Science hire. I was instrumental in adding three additional members to the Data Team. At Arkera I am responsible for the backend and web scraping as well as designing, prototyping and implementing proprietary Machine Learning and Natural Language Processing algorithms. We predominantly use Python and PostgreSQL in Docker on AWS to serve iOS and web applications. I am a strong advocate of Test Driven Development and in my current role I was responsible for implementing Agile development processes within the team. I was previously the first Data hire at Stylect where I built their first recommender system for their app and website, utilising my expertise in Machine Learning. In December 2015 I spoke at Big Data Week 2015, co-hosting the workshop: "How to Become an Effective Data Scientist".

I was previously a Postdoctoral Research Associate at MIT. I completed my Ph.D. at the University of Cambridge where my thesis won the Springer Thesis Prize: "Recognising Outstanding Ph.D. Research" and is published in the Springer Thesis Program. I transitioned into industry through a Postdoctoral Fellow on the Data Science & Business Analytics Fellowship offered by The ASI. Since leaving academia I have been working at a couple of start-ups, officially as a Data Scientist but with a very hands-on development role productionising software. During this time I have come to realise that I really enjoy building tangible applications with real-world use. As such I am looking to build my career within a development focused role.

At Stylect I was responsible for creating and maintaining a unified and coherent set of APIs that served 100,000 visits each month on the e-commerce website and also served an app which had one million downloads. At Arkera I have built APIs with OAuth2 authentication whilst developing a secure linked data platform. I introduced code review at Arkera, ensuring the junior developers had an open forum within which to improve their skills. I implemented Arkera's first set of coding standards and documentation and integrated Test-Driven Development within the team, ensuring all code deployed into production has good test coverage.

I am excited by the prospect of joining the evolving and ambitious Web team at Wellcome and being a part of the friendly and creative working environment that will build a free digital space where people discover and engage with information and other people about the cultural contexts of health and the human condition.

I look forward to hearing from you.

Yours sincerely,

**Greg Lever**