Elliot Marsden

Address	Sarphatistraat 76/3 Amsterdam, 1018 GS, Netherlands	**Phone**	+31 644481822
		Email	elliot.marsden@gmail.com
Date of Birth	1989-02-14	**Website**	elliotmarsden.com
		Github	@eddiejessup

Employment

May 2016 – Present **Data Scientist**, **Booking.com**

- Answering business questions, helping people make decisions and developing ways to quantitatively measure their success.
- Building predictive models using machine learning, with Scikit-learn, Spark ML and H2O.
- Querying petabyte-scale data using Hadoop-ecosystem tools such as Hive-SQL, Spark and Oozie.
- Implementing A/B tests at scale. Doing meta-analysis of experimentation practice to improve results, and designing advanced statistical tests to increase ability to detect small effects.
- Other tools used include Matplotlib and Bokeh for data visualization, Grafana and Tableau for building dashboards, Jupyter, Git and MySQL.

Sep 2015 – Apr 2016 **Software Developer**, **Ecometrica**

- Developed the back-end and front-end of the Django-powered single-page web application.
- Helped design architecture, and migrate application to Amazon Web Services, becoming familiar with most of its components.
- Independently drove project to create a tool for interactive analysis of geo-spatial data, through Pandas, Numpy and Scipy.
- Other tools used include Flask, Mercurial, Celery, PostgreSQL and Redis.

Education

2011 – 2015 **PhD in Soft Condensed Matter Physics, University of Edinburgh**

- Thesis on *The collective dynamics of self-propelled particles in confining environments*, that is, on the movement of large numbers of bacteria in enclosed spaces such as water droplets.
- Statistical analysis and numerical simulation, both done mainly in Python using Numpy, Scipy and Cython.
- Close interdisciplinary collaboration with theoretical physicists and experimental biologists, on work of interest to audiences ranging from statistical physics to molecular biology.
- Publication of work in *Soft Matter*, and on the front cover of *Physical Review Letters*.
- Spoke at many academic conferences, including at *Stochastic Biology: from Cells to Populations*; as an invited speaker at the Niels Bohr Institute and at the NETADIS Summer School.

2007 – 2011 **MPhys in Theoretical Physics, Durham University**

- Graduated First Class with Honours.
- Thesis on *Computational Analysis of Shear Banding in Shear Startup and LAOS Regimes*, that is, on the complex behaviour of polymeric fluids as they flow.
- Contributed to publication in *Physical Review Letters*.

Experience

CUPED

- Worked to assess and implement the statistical technique CUPED, described in *Improving the Sensitivity of Online Controlled Experiments by Utilizing Pre-Experiment Data*, by Alex Deng et al.
- Aimed at increasing the ability of A/B tests to detect small effects.

TEX

- Ongoing work to re-implement the interpreter of the document preparation language TEX in Python.
- Required implementing complex lexing and parsing logic, containing a mixture of context-free and context-sensitive grammar.

Sneezly

- Service to allow users to easily log events.
- Back-end implemented using Django Channels, allowing support for websocket connections.
- Support for multiple front-end interfaces, including a Slack chat-bot.

Seavreeze

Tool to generate a CV in multiple output formats, including PDF and HTML, from a common input.

Transcrypt

Plugin for the text editor Sublime Text to do in-place text encryption.

Open source contributions

Contributed to various projects, including Pandas, knitpy and Jupyter.

Teaching

- Taught version control, Python and Linux skills to PhD students at two-day Software Carpentry Boot Camps in Edinburgh and Bath.
- Worked as a teaching assistant during my PhD, leading tutorials and helping to run undergraduate student workshops.

Other skills

- Used Tensor Flow to train a convolutional neural network to identify playing cards from images.
- Familiar with theoretical foundations of statistical learning, and with modern algorithms such as gradient boosting, random forests and various flavours of neural networks.
- Reasonable practical experience with Javascript, C++, Java and Go through personal projects, as well as more cursory experience with Scala, Rust and D.
- Used the NoSQL database MongoDB as part of a distributed system to record sports betting odds.
- Knowledge of computer science fundamentals, such as complexity theory, algorithm analysis and data structures.
- Aware of modern development practices and tools, such as unit testing frameworks, debuggers and performance profilers.

Summary of skills

Data Science

Hadoop

Hive, Spark and Oozie.

Machine learning

Scikit-learn, Spark ML, H2O and Tensor Flow.

Data analysis

Pandas, StatsModels and Jupyter.

Data visualization

Matplotlib, Bokeh, d3.js, Grafana and Tableau.

Theory

- Statistical techniques such as parametric significance testing, bootstrapping and permutation testing.
- Algorithms for machine learning such as regression, k-nearest neighbours, gradient boosting, random forests, neural networks and unsupervized clustering.
- General techniques of machine learning such as regularization, feature selection, dimensionality reduction and cross validation.

Software Development

Operating Systems

Comfortable working in Linux and MacOS.

Version control

Git and Mercurial.

Programming languages

Very capable using Python. Competent using Javascript, C++, Java and Go. Superficial experience with Scala, Rust and D.

Back-end web development

Django and Flask web frameworks.

Front-end web development

Standard web technologies, HTML, CSS, Javascript and related tools.

Databases

MySQL, PostgreSQL, MongoDB and Redis.

Amazon Web Services

Database storage (RDS), caching (Elasticache), storage (S3), hosting (EC2) and task scheduling (Lambda).

Scientific and numerical computing

Numpy, Scipy and Cython.

Theory

Complexity theory, algorithm analysis and data structures.