

# Jack Grimes

jack@jackgrimes.net

London, UK

jackgrimes.net

GitHub: [jackgrimes](#)

LinkedIn: [jacketgrimes](#)

Google Scholar: [HZLogpQAAAAJ](#)

## EDUCATION

### IMPERIAL COLLEGE LONDON

Ph.D. Civil and Environmental Engineering, and Infectious Disease Epidemiology

*Thesis: An investigation into the roles of water, sanitation, and hygiene in the control of schistosomes and other helminths*

March 2012 - September 2015

### IMPERIAL COLLEGE LONDON

M.Sc. Civil and Environmental Engineering

September 2010 - September 2011

### UNIVERSITY OF OXFORD

M.A. (Oxon) Physics

September 2007 - September 2010

## SKILLS

### PROGRAMMING

**Very familiar:** Python • R • Git/GitHub  
Matlab/Octave

**Familiar:** SQL • JavaScript • HTML  
CSS •  $\text{\LaTeX}$

### PLATFORMS

Google Cloud Platform • Domino

## COURSES

Bayesian Statistics: From Concept to Data Analysis (Coursera, University of California, Santa Cruz)

Deep Learning Specialization (Coursera, deeplearning.ai)

Machine Learning (Coursera, Stanford University)

## INTERESTS

Cycling • Walking • Travelling • Reading

## EXTENDED CV

A more detailed version of my CV can be found on my [website](#)

## EXPERIENCE

### MORSUM London, UK (remote)

September 2022 - Present: **TECHNICAL LEAD (DATA SCIENCE)**

Leading a squad responsible for product search, recommendations, and substitutions, in an online food marketplace (Ancho.vi)

- Productionising Sentence-BERT model serving and fine-tuning, using Vertex AI
- Data versioning in BigQuery
- Recruitment and line management of a Machine Learning Engineer

March 2021 - August 2022: **SENIOR DATA SCIENTIST**

Building the data science team: defining areas of responsibility, embedding good practices, and recruiting data scientists and engineers

- Driving good software engineering practices: unit tests, code linting and formatting, and sonarcloud and GitHub Actions builds
- MLOps: implementing an initial standardised approach to the model training, registry, loading and serving, using Google Cloud Storage
- Data architecture for data ingestion: redefining data entities

Standardisation of incoming text data: the display name cleaner

- Using text vectorisers and classifiers to fill out abbreviations, and to separate out product descriptions from names
- Embedding solutions within wider business logic code, and productionisation on a Flask API
- Ownership of the product and coordination with stakeholders

Using Git, Google Cloud Platform, MySQL, pandas, PyCharm, pytest, scikit-learn

### LLOYDS BANKING GROUP London, UK

January 2019 - March 2021: **DATA SCIENTIST, APPLIED SCIENCE**

Writing efficient, readable, unit-tested data science code for production, and contributing to the data science community within the bank

- Classification, clustering, network analysis, NLP: identification of incorrectly applied account and customer indicators
- Running the *Modelling Interest Group*, organising regular talks
- *Data Science Graduate Scheme Training Programme*: delivering the *Advanced Modelling* and *Natural Language Processing* sections
- Git, Jenkins, NetworkX, pandas, PyCharm, pytest, scikit-learn, spaCy

August 2017 - December 2018: **DATA SCIENTIST, RISK INNOVATION**

Building and demonstrating proofs of concept to demonstrate potential applications of machine learning in the bank

- Predictive modelling of fraudulent transactions, and of financial difficulty
- Topic modelling of regulatory data
- Face recognition for fraud prevention
- Presenting at internal conferences throughout the UK
- face\_recognition, Flask, Gensim, Git, pandas, SAS VA, scikit-learn

### IMPERIAL COLLEGE LONDON London, UK

January 2016 - July 2017: **POSTDOCTORAL RESEARCHER**

Statistical analyses of the environmental control of tropical diseases

- Mixed effects logistic regression (R and lme4) to compare risk factors, parasitic infection, and overall health, in a clustered dataset ([paper](#))
- Relationships between infection intensity and prevalence (Box-Cox transformations, linear regression, R)