

Experienced data scientist with 5+ years in Python and statistical analysis, specializing in life sciences, statistics, and the finance sector. Skilled in handling large datasets, machine learning, and risk analysis. Proven track record in delivering actionable insights and automated data pipelines and stakeholder reports. Eager to contribute expertise to dynamic teams and further professional growth.

PROFESSIONAL EXPERIENCE

Data Scientist NatWest Group

 Mar 2023 - Present Edinburgh

- Automated profitability scenario reporting, reducing human error + hours of excel work
- Implemented version control for production codebase, improving team standards
- Delivered improved polynomial fit of early-repayments, improving product profitability forecasting
- Extracted insights from transactional datasets to inform pricing optimisation

Junior Data Scientist Strike

 Sep 2021 – Feb 2023 Colchester


- Developed and deployed production forecasting model to estimate future product profitability.
- Cleaned and documented large financial datasets
- Contributed to customer value model development based on signup data.
- Implemented and supported DAGs using Apache Airflow to run scheduled scripts generating forecasts.
- Performed A/B multivariate testing for product selection.
- Generated reports and figures to answer fast questions from stakeholders
- Stakeholder meeting lead.

Data Science Consultant Pivigo

 May 2021 - July 2021 London

- EDA + data mining
- Led team of 5
- Delivered actionable insights to company

Postdoctoral researcher University of Glasgow

 Aug 2019 - May 2021 Glasgow

- Teaching data analysis and biostatistics.
- Implemented data cleaning, storage and statistical analysis pipelines for large image and video datasets.
- Published in high-impact journal (Lachaud et al., 2022)

Biomedical consultant Elanco (Eli Lilly)

 2016 – 2018 (ad hoc) Glasgow

- Created and delivered client data report.

TECHNICAL SKILLS

- **Languages:** Python, SQL
- **Tools:** AWS SageMaker, Jupyter,
- **Database:** Redshift, AWS Athena
- **Cloud/VC:** Apache Airflow, Bitbucket, GitHub
- **Packages:** pandas, numpy, scikit-learn, scipy, statsmodels, tensorflow, selenium, s3fs

PROJECTS

Detection and analysis of electrical spikes in cardiac cells via fluorescent video assay (Ph.D. Thesis):

- Used signal-filtering algorithms to remove signal noise and preserve information. Analysis of duration, amplitude and frequency using Python and other techniques. Advanced statistical analyses and presentation using SciPy and Graphpad.

UK property market trend analysis through web scraping and time-series forecasting:


- Used web-scraping and automation (selenium) to scrape, collect and clean data from various real-estate agencies to estimate near-future price (ARIMA).

Regression-based revenue forecasting (Strike):

- Used past behaviour to predict cohort maturity plateaus for 1 year+ using linear regression, exponential smoothing, and linear averaging. Successfully integrated model in company's total revenue forecast.


EDUCATION (University of Glasgow)

Ph.D. Cardiac Electrophysiology

 2015/19

- Statistical modelling (Monte Carlo Simulation)
- Python | Ruby | Time-series analysis
- Scientific report presentations

Master of Science Translational Medicine

 2014/15

- Scientific writing | presentations
- Statistics (Power calculations)

Bachelor of Science Neuroscience (Hons)

 2009/12

- Statistics | Image analysis | Critical writing