

SARAH WESOLOWSKI

Location: York, UK

Linkedin: <https://www.linkedin.com/in/sarahcwesolowski/>

Data science leader in UK finance with a PhD in physics and 10+ years of experience delivering innovative, practical insights and models across different industries from healthcare to banking. Currently lead a mortgage behavioural data science team at Yorkshire Building Society in Leeds, using individual customer, product, and economic data to build predictive models and insight dashboards that support risk-based mortgage pricing and feed directly into Treasury risk mitigation strategy.

BUSINESS SKILLS

- Working with customer and markets data.
- People management.
- Partnering with business stakeholders.
- Project management and delivery.
- Reprioritising when business needs change.
- Distilling technical information clearly for senior leadership and stakeholders.
- Staying up to date with AI and Data Science news and academic literature.

TECHNICAL SKILLS

Languages: Python, SQL, SAS, C++, Matlab, etc.
Version control: Git (Gitlab and Github).
Data Visualisation: Tableau, plotly, seaborn, dash, streamlit, etc.
Data and ML: pandas, pyspark, mlflow, sklearn, xgboost, etc. and custom methods from scratch.
Cloud: Azure Databricks, Docker.
Mathematical: Machine learning, statistics, Bayesian methods, MCMC, feature engineering.
Data: Data pipelines, ETL, unifying disparate data sources.

PROFESSIONAL EXPERIENCE

Lead Data Science Manager, Yorkshire Building Society

Leeds, UK

(January 2024 – Present)

Line manager for a team of 4 data scientists working to analyse and predict customer mortgage behaviour for the entire mortgage lifecycle, ranging from expected sales volumes up to customer retention after their mortgage deal ends. Team products feed into mortgage pricing, financial planning, and hedging against economic shocks when customers tend to leave mortgage deals early or move to different products.

- **Led delivery for a high-profile mortgage prepayment forecast and monitoring model**, replacing outdated assumptions and working closely with stakeholders to achieve deployment. This has resulted in nearly **£2 billion** of new interest rate hedging deals transacted in the Treasury, protecting against future interest rate shocks.
- **Built a mortgage price elasticity model** for acquisitions of new mortgage business, engineering complex features based on mortgage market competitiveness of the Society and other customer level features. Tested challenger models and chose a random forest model for interpretability and performance within business tolerances.
- **Moved to an improved central data source** for mortgages from 2024 to mid 2025, streamlining the team's analytics products, **saving work in the amount of 1 FTE annually**. Managed stakeholder expectations and timeline as the change occurred, because outputs changed significantly from the previous data source.
- **Recruited 2 new data scientists** to build up the team in 2024, with both employees delivering major analysis work and exceeding performance expectations in their first year.

Senior Data Scientist, University of Pennsylvania School of Medicine

Philadelphia, Pennsylvania, USA

(February 2022 – December 2023)

Senior data scientist managing the Herman Lab project portfolio and mentoring 2 junior colleagues, leading the lab's analytics transition into an Azure Databricks framework. Developed new statistical methods, got them implemented in clinical studies, and reported deliverables to funding agencies.

- **Built a machine learning model framework** for targeting patients for disease screening using electronic health records data and delivered quarterly reports to the funding agency. Customised setup was needed beyond out-of-the-box ML methods due to ethical and regulatory requirements. Used customised genetic algorithm ("Feat") to create an interpretable and performing model against a traditional logistic regression model with custom features.
- **Wrangled complex health data sources** from multiple hospital systems into easily-usable views in Databricks and using pyspark to target patients who would benefit from disease screening.
- **Optimised codebase for disease screening prospective study**, working with junior data manager to target patients for primary aldosteronism screening in the local health system, with weekly queries and dashboard reporting for the clinical research coordinator to enrol patients.

Research and Lecturer, Salisbury University Department of Mathematical Sciences

Salisbury, Maryland, USA

(August 2017 – January 2022)

Performed and led independent research projects, developed and taught statistics and data science curriculum, mentored student research, and guided client-facing statistical analysis projects.

- **Published 4 scientific articles with international teams** in top physics journals while managing a challenging full-time teaching and mentoring workload. Projects focussed on nuclear physics models using Bayesian methods.
- **Coordinated and mentored team analytics projects** where student groups worked with clients such as the United States Navy and NASA to analyse organisational logistics data and produce reports for the clients.
- **Developed new data science programme** by both creating cutting-edge curriculum for upper-level courses and coordinating budget with 6 university departments and administrators.

EDUCATION

PhD in Physics, The Ohio State University

Columbus, Ohio, USA

Thesis title: "Bayesian Methods for Effective Field Theories"

Bachelor's of Science in Physics, Saint Vincent College

Latrobe, Pennsylvania, USA

LANGUAGES

English (native), Japanese (N3)