

DANIEL HALWELL

SUMMARY Highly motivated Senior Scientist transitioning to a role in Data Science leveraging advanced analytics, machine learning (ML), and artificial intelligence (AI) to solve complex problems and drive innovation. I possess a strong foundation in analytical problem-solving from the pharmaceutical industry, combined with advanced expertise in Python, statistical modeling, REST API development, and the creation of AI-powered applications. Proven ability to manage and analyze large datasets, develop predictive models, automate processes, and lead technical projects. Actively upskilling with certifications from IBM, Google, and DataCamp, and pursuing personal development in cybersecurity through hands-on learning on platforms like *Hack The Box* and *TryHackMe*, to stay at the forefront of data science, AI, and security.

SKILLS & ABILITIES

- **Machine Learning & AI:**
 - Advanced proficiency in Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, PyTorch, Plotly, Streamlit, PyTorch and Jupyter Notebooks).
 - Achieved 4th place in the Kaggle Playground Series S5E3 competition: "Binary Prediction with a Rainfall Dataset" in 2025.
 - Familiar with a broad range of ML algorithms including regression, clustering, bagging, boosting, and neural networks; and model evaluation techniques (confusion matrix, ROC-AUC, Type I/II errors).
 - Certified as an IBM AI Developer, with experience building AI-powered applications, including development of troubleshooting chatbot using OpenAI and Retrieval Augmented Generation (RAG).
- **Data Analysis:**
 - Skilled in SQL for querying relational databases (e.g., Snowflake) and manipulating large structured/unstructured datasets. Familiar with concepts of non-relational databases.
 - Strong understanding of statistical principles: probability, confidence intervals, hypothesis testing (t-test, ANOVA), central limit theorem.
 - Data visualization expertise (Power BI, Matplotlib, Seaborn, Plotly) for conveying insights.
 - Certified in DataCamp's SQL Associate and Data Literacy Fundamentals, Google Data Analytics Professional Certificate.
- **HPC & Big Data:**
 - Experienced in handling large datasets and have recently begun to leverage high-performance computing resources for efficient data processing and analysis.
 - Developed and deployed REST APIs (e.g., Physical Chemical Data API, Nitrite Database Data API) using Python (e.g., Flask/FastAPI) for data retrieval, service integration, and enabling real-time/batch data exchange.
- **Cybersecurity Knowledge & Application:**
 - Actively developing practical knowledge in cybersecurity principles, penetration testing methodologies, and threat actor tactics through hands-on labs and challenges on Hack The Box and TryHackMe.
- **Technical Skills:**
 - Proficient in Linux with some experience using cloud providers GCP for using Google BigQuery, Amazon Bedrock/Sagemaker for LLM's and ML and Azure for Generative AI (AI Foundry) in a personal capacity. Also familiar and frequently use Git version control.

AstraZeneca | Senior Scientist (2021 – Present)

- **Data Science Initiatives:** Routine use of Python to automate tasks, increasing efficiency and accuracy in data handling. Developed and implemented predictive models using machine learning algorithms (retention time prediction by liquid chromatography) to enhance departmental capabilities. Active member of the internal Coding Network, presenting my work on a regular basis and coaching others. Recently investigated the use of Bayesian frameworks to reduce method development by 50%
- **Data Analysis:** Utilized SQL and Python in conjunction with Snowflake for complex data analysis, including cleaning data for machine learning and retrieving instrument usage metrics.
- **Generative AI Project:** Developed a gas chromatography troubleshooting chatbot using OpenAI for image and text processing with Streamlit for the user interface. Demonstrated the potential for AI to enhance workflows.
- **API Development for Data Services:** Designed, developed, and deployed two key REST APIs (Physical Chemical Data API and Nitrite Database Data API) using the FastAPI Python framework, improving data accessibility and inter-system connectivity.

Sanofi/Recipharm | Analytical Specialist (2016 – 2021)

- **Technical Support & Data Analysis:** Provided technical support for analytical method transfer to cGMP standards. Conducted statistical analysis for equivalency testing using standard frequentist techniques and reviewed technical documentation.
- **Problem-Solving:** Facilitated large, cross-functional investigations using Practical Problem Solving (PPS) methodology to resolve complex technical issues.

Sanofi | Quality Analyst/Senior Quality Analyst (2012 – 2016)

- **Analytical Laboratory:** Focused on analysis of nasal and aerosolized pharmaceutical products. Developed troubleshooting strategies for technical issues, methods, and test results (OOS, OOT).
- **Team Leadership:** Played a key role in planning team activities and acting as a deputy for the line manager.

EDUCATION

- **MChem in Chemistry, Loughborough University (2007 – 2012)**
- **IBM AI Developer Professional Certificate (2024)**
- **DeepLearning.AI TensorFlow Developer (2024)**
- **Microsoft Python Development Specialization (2025)**
- **Docker Foundations Professional Certificate (2025)**
- **Google Data Analytics Professional Certificate (2024)**
- **DataCamp:**
 - Associate Data Science Certificate (2023)
 - SQL Associate (2024)
 - Data Literacy Fundamentals (2024)
 - AI Fundamentals (2024)
- **IBM Generative AI Engineering Professional Certificate (In Progress – 2025)**
- **Google Agile Essentials (In Progress - 2025)**
- **IBM Cybersecurity Analyst (In Progress - 2025)**

PUBLICATIONS

- Vicente Ferreira, Ricardo Lopez, Flavour Science: Proceedings from XIII Weurman Flavour Research Symposium (2014:487-491), Daniel Plyer, Daniel Halwell, et al.
- GitHub - [Personal GitHub](#)
- Medium – [CodeHalwell](#) - Author on the Medium Platform, writing articles on coding.