

# STEPHEN NICHOLSON

DATA ENGINEER   📍 GLASGOW, UNITED KINGDOM   📞 07535244018

◦ DETAILS ◦

Glasgow  
United Kingdom  
07535244018  
[stephen.nicholson1@hotmail.co.uk](mailto:stephen.nicholson1@hotmail.co.uk)

◦ LINKS ◦

[linkedin](#)  
[Personal Page](#)  
[GitHub](#)

◦ SKILLS ◦

Python (Programming Language)  
Object-Oriented Programming (OOP)  
SQL (Programming Language)  
Software Development  
Extract Transform Load (ETL)  
Data Science  
Algorithms  
Machine Learning (ML)  
Artificial Intelligence (AI)  
Docker (Software)  
Graph Database  
DevOps  
Amazon Web Services  
Microsoft Azure  
Statistics  
Data Analytics  
Data Warehousing  
Apache Spark  
Agile Software Development  
Application Development  
Information Governance  
Information security  
Customer Engagement

◦ LANGUAGES ◦

English



EMPLOYMENT HISTORY

**Data Engineer at Smart Data Foundry, Edinburgh**

August 2023 — Present

- Led the development of a high-performance data pipeline using Apache Arrow and Polars in Python, reducing processing times by 65% and ensuring reliable quarterly operations.
- Optimised data management processes across 4 data partners by implementing scripts in Python for automated data transformation, reducing manual data processing time and enabling insights extraction.
- Implemented automated workflows using Airflow, resulting in a 40% reduction in manual tasks and a 30% increase in project delivery efficiency.
- Developed Python data pipelines on AWS to power R Shiny dashboards, enabling cross-functional teams to gain actionable insights on the cost-of-living crisis, resulting in a 25% decrease in operational costs.

**Machine Learning Engineer at NatWest Group, Edinburgh**

July 2022 — August 2023

- Led the development of sophisticated ETL pipelines using Apache Spark and Python, driving enhancements to a critical fraud detection model that resulted in a 20% increase in accuracy and performance metrics.
- Executed the implementation of MLOps best practices, orchestrating deployment of new Python-built machine learning models across on-premises Apache Spark, Hadoop, and AWS Sagemaker and EMR resources, optimising performance and scalability.
- Engineered and deployed advanced Python/PySpark pipelines on Hadoop infrastructure to analyse customer data; boosted bank's GDPR compliance by determining customer relationship end dates, ensuring regulatory adherence.

**Data Engineer at NatWest Group, Edinburgh**

November 2020 — July 2022

- Created high-performance Python/PySpark pipelines in Apache Spark, integrating customer data into Neo4j graph database, providing a tool for fraud analysts that detected fraudsters and saved £7 million for the bank.
- Implemented Neo4j GDS library and algorithms like Louvain, enhancing fraud detection by identifying fraudulent communities.
- Engineered 20+ features with data scientists to enhance predictive modelling accuracy for customer lifetime value and politically exposed persons; achieved 15% increase in model effectiveness and reduced false positives by 25%.
- Orchestrated and delivered comprehensive training on Python, Git, Linux, Docker, and CI/CD to 100+ employees and built reusable Python packages, enhancing technical skills and developer productivity across the organisation.

**Graduate Automation Engineer at Momena Group Global, Glasgow**

May 2020 — November 2020

- Developed a Python app to automate previously manual bot statistics reporting, saving engineers 5 hours per week and boosting team productivity by 25%.
- Consulted with clients to understand and translate RPA bot requirements for 12 bots already in production, ensuring that the bots continued to meet client specifications and maintain optimal performance.
- Resolved incidents and deployed enhancements for Automation Anywhere and UiPath bots, reducing bot downtime by 30% and improving overall bot performance by 20%.

- Wrote detailed documentation for 12 bots, including Design Documents, Test Plans, and Change Requests, reducing maintenance and troubleshooting time by 15%.

#### **Data Scientist at The Scottish Government, Edinburgh**

June 2019 — August 2019

- Authored innovative R scripts to develop a predictive linear panel regression model using historical Scottish company data, delivering actionable insights for informed decision-making and performance metrics.
- Introduced a new way of working to a team of 6 statisticians to manipulate and evaluate data with the R programming language.
- Trained statisticians in R programming, enabling the government to save millions of pounds annually on SAS software license fees.



### **EDUCATION**

#### **Big Data, University of Stirling, Stirling**

January 2018 — January 2019

Master of Science (M.Sc.) with Distinction

#### **Computer Science, The University of Stirling, Stirling**

January 2014 — January 2018

Bachelor of Science (BSc)

#### **Nanodegree, Data Engineering, Udacity**