Surya Naraynan Ramessh

Data Science · Data Analytics · Data Platforms

Python · SQL · Power BI · Microsoft Fabric · AWS · Machine Learning

Manchester, UK | +44 7867 230050 | suryanaraynanr1@gmail.com

linkedin.com/in/surya-naraynan | suryanaraynan.github.io

UK right to work, no sponsorship required | Available immediately

PROFILE

Data Scientist with over six years across data science, analytics and software engineering. Builds and explains machine learning and statistical models in Python, R and SQL, designs A/B tests and controlled experiments, and builds the pipelines, warehouses and Power BI reporting that put those results in front of decision makers. Works across the full path from raw source data to a governed dataset to a model or dashboard a team can rely on. MSc Data Science, University of Manchester.

KEY STRENGTHS

- **Machine learning and statistical modelling:** computer vision and regression models, including a 20-point accuracy gain on a biometric model and deployed workforce models explaining ~75% of outcome variance.
- **Experimentation and statistics:** A/B testing, sampling design and validation on a £3.2M government-funded trial with 2,500+ participants and 52 systems.
- **Data platforms and pipelines:** sourcing, cleaning and validating 450,000+ patient records (Microsoft Fabric) and 200,000+ images (AWS) with Python, SQL and PySpark, including star schema design and automated quality checks.
- **Analytics and communication:** Power BI and DAX reporting that replaced 255 legacy Excel reports, with findings presented to executives at eight client organisations.

CORE SKILLS

Machine Learning & AI: regression, classification, clustering (k-means), deep learning (TensorFlow), computer vision (OpenCV), NLP (sentiment analysis, topic modelling, NER), sequential pattern mining, feature engineering, model evaluation and error analysis

Statistics & Experimentation: A/B and hypothesis testing, experimental design, sampling and bias reduction, multivariate and longitudinal analysis, exploratory data analysis

Programming: Python (Pandas, NumPy, Scikit-learn, TensorFlow), R (Tidyverse), SQL (T-SQL, CTEs, window functions), PySpark

Data Platforms & Cloud: AWS (S3, Athena, Redshift, SageMaker), Microsoft Fabric, Azure, Databricks, MongoDB, SSIS, SSMS, Gen2 Dataflows, medallion and star schema modelling, ETL/ELT, data migration

Analytics & Visualisation: Power BI (DAX), Power Query, Tableau, dashboards and reporting for technical and non-technical audiences

Ways of Working: Git, Azure DevOps, Jenkins CI/CD, pytest, Jira, Agile (Scrum), code review, mentoring, GDPR and data governance

EXPERIENCE

IPRS Group (Robertson Cooper and IPRS Health) | Data Scientist *Oct 2025 – Aug 2026*

Occupational health and workforce wellbeing group. Concurrent workstreams across Robertson Cooper and IPRS Health.

Robertson Cooper: Client Workforce Analytics

- Analysed **56,000+** records of employee wellbeing survey data with SQL and Python to uncover trends, benchmarking each client against Robertson Cooper's own benchmark dataset.
- Developed regression models that explained **~75%** of the variation in productivity, wellbeing and staff turnover, helping clients identify the key drivers to prioritise.
- Deployed the models and risk hotspot analysis across **12 health measures** as a reusable Python pipeline that refreshed with each new client dataset, presenting findings to executives at eight client organisations.

IPRS Health: Data Platform and Reporting

- Sourced, cleaned and validated **450,000+ patient records** from **90+ source tables** into a Microsoft Fabric warehouse, with star schema models, Gen2 Dataflow pipelines and quality checks at each layer.
- Migrated 255 legacy Excel reports into a single Power BI framework, redefining KPIs with 8 team leads and saving **40 hours of team effort per week** across eight teams.
- Standardised auditable patient data handling organisation-wide through a layered governance framework of pseudonymisation, row-level security and role-based access control.
- Supported **6 analysts** through SQL and Fabric training and peer review of their work, contributing to continuous learning across the team.

Age Check Certification Scheme | Applied Data Scientist *Jan 2025 – Sep 2025*

Data Scientist (promoted to Applied Data Scientist) *Feb 2024 – Dec 2024*

Independent certification body for age assurance and biometric technology. Worked with regulator, vendor and academic partners.

- Improved biometric computer vision model accuracy from **61% to 81%** by tracing errors to source data and fixing them through deduplication, labelling correction and validation rules in an AWS S3 pipeline.
- Led the data work on a **£3.2M** government-funded Australian evaluation of **52 biometric systems**, designing sampling, validation and **A/B testing** across **2,500+ participants** and analysing results with AWS Athena and Redshift.
- Contributed to scalable data science infrastructure (SQL and PySpark pipelines on AWS S3) for UKRI-funded AI research in deepfake detection, curating **200,000+ image records** across UK and Swiss jurisdictions under GDPR.
- Co-authored the published Evaluation Proposal Report (Feb 2025), setting out the trial methodology, and presented findings to 50+ technical and non-technical stakeholders.

Earthchain | Data Consultant (pro bono, alongside MSc) *Jan 2023 – Jun 2023*

- Designed a weighted scoring model for **1,500+** carbon credit projects, scraped and validated from 3 ESG registries into MongoDB, to support investment decisions; led six people to deliver 7 Power BI dashboards, the **top-ranked outcome** across all teams.

Quinbay Technology | Senior Software Development Engineer in Test *Jun 2019 – Aug 2022*

- Created UI and API test frameworks (pytest, JavaScript, Appium) that increased automation coverage by **80%**, and streamlined releases with Jenkins CI, saving **20+ hours per sprint**.

SELECTED PROJECTS

Golden Dataset Pipeline | github.com/suryanaraynan/golden-dataset-pipeline *2026*

- Built a demo pipeline turning a manual dataset of completed legal matters into a versioned, pseudonymised, quality-checked and repeatable dataset, with evaluation data for measuring whether an AI model improves (Python, SQL, pandera, Parquet, GitHub Actions). Synthetic data only.

Mining User Behaviour Patterns and Trends | MSc Dissertation *Jun 2023 – Sep 2023*

- Combined **1.2M** interaction log entries from 8 dashboards used by 50 people into one analysis-ready dataset (Python, JSON).
- Mined sequential patterns (SPMF) to identify the **7 most common** behaviour patterns, predicted task speed and accuracy from them with Support Vector Regression across 9 usability measures, and ran A/B tests on dashboard designs.

EDUCATION

MSc Data Science (Merit) | University of Manchester *Sep 2022 – Sep 2023*

BTech Information Technology (Distinction) | Anna University *Aug 2015 – May 2019*