
Cameron Poole

MSc Data Science | Python · SQL · statistical modelling · data quality · applied analytics

London | 07930 991269 | cameron.edek.poole@gmail.com | linkedin.com/in/cam-edek-poole | github.com/cedekpoole | cedekpoole-website.vercel.app

PROFILE

Philosophy graduate and MSc Data Science student moving into applied data roles in London. I am interested in how messy and ambiguous information can be cleaned, structured, analysed and turned into something meaningful. Through my MSc, I have worked across Python, SQL, statistics, machine learning, data warehousing and applied data projects, with particular interest in statistical evaluation, data quality and decision support. My dissertation develops a supplementary compliance-risk score for Environment Agency data, testing whether multi-year compliance history can improve serious-breach prioritisation beyond the annual A–F band.

TECHNICAL & PROFESSIONAL SKILLS

Programming, analysis & modelling	Python, R, SQL, pandas, NumPy, SciPy, scikit-learn, statistical testing, linear regression, logistic regression, classification metrics, model evaluation, feature engineering
Databases, data modelling & SQL	PostgreSQL, Oracle SQL, BigQuery, ER modelling, star schemas, fact/dimension tables, grain
Data workflows & quality	dbt, Azure Blob Storage, GitHub Actions, API ingestion, pytest, data cleaning, validation checks, missing-value handling, reproducible analysis
Visualisation & development	Tableau, Looker Studio/Data Studio, matplotlib, seaborn, Highcharts, Git, GitHub, VS Code, JavaScript, TypeScript, React.js, HTML/CSS, REST APIs
Professional strengths	Structured thinking, written communication, attention to detail, teamwork, problem-solving

SELECTED PROJECTS

MSc Dissertation - Environment Agency Compliance Risk Analysis

Kingston University | Python, SQL, statistical evaluation | In progress, 2026

- Built a reproducible permit-year panel from 2015–2024 EA compliance and breach data, resolving changing permit IDs and auditing data-quality issues.
- Engineered longitudinal features including prior breaches, band history, volatility and sustained good-compliance signals to assess next-year serious-breach risk.
- Developing an interpretable supplementary risk score to test whether compliance history improves prioritisation beyond the annual A–F band, using logistic regression and rare-event ranking metrics.

TfL London Transport Analytics Mart

Personal portfolio | Python, Azure Blob Storage, BigQuery, dbt, GitHub Actions, Looker Studio | 2026

- Built an end-to-end analytics pipeline around TfL Tube line-status data, including Python API ingestion and raw JSON preservation using Azure Blob Storage, plus BigQuery loading, dbt modelling and scheduled GitHub Actions automation.
- Modelled nested API snapshots into staging, fact and mart layers, including line-level status counts and daily reliability marts with dbt tests.
- Added lightweight BI output and analysis SQL for disrupted lines, worst daily reliability and latest line status.

Statistical Modelling in R

MSc project | R, regression, logistic regression, model diagnostics, classification metrics | 2026

- Built and validated linear regression models in R using stepwise selection, residual diagnostics and VIF checks.
- Developed logistic regression models for clinical fracture classification, evaluated with confusion-matrix metrics including sensitivity and specificity.

ADDITIONAL WORK

- **Rail Punctuality Data Warehouse** - Oracle SQL star schema integrating train journey and daily weather data at journey grain, cleaned and loaded via staged ETL with Tableau reporting on delay patterns.
- **UK Carbon Intensity Pipeline** - Current Python ingestion and pytest project exploring public carbon intensity and generation-mix data, with planned Snowflake/dbt extension.

EXPERIENCE

Front-End Software Development Intern

Goodwright Ltd. | London / Remote | Sept 2023 - Dec 2023

- Built interactive biological data visualisation components in React.js and Highcharts for scientist-facing software.
- Developed PCA visualisations with 2D/3D plots and scree plots to support sequencing data quality review.
- Built an interactive RNA-seq volcano plot using log2 fold-change and adjusted p-value thresholds.
- Translated complex biological data into clearer, responsive interfaces for scientific users.

Web Designer & Developer

Edek Design | London / Remote | Aug 2024 - Aug 2025

- Created a self-initiated design and web project while exploring a move deeper into technology.
- Designed and built the brand, website, visual identity and portfolio concept independently, developing practical skills in front-end technologies, responsive design, UI structure and visual design.
- The project helped clarify that I was more drawn to the analytical and technical side of digital work, contributing to a move into data.

Store Colleague

Holland & Barrett | Twickenham | Aug 2023 - Aug 2025

- Worked in a customer-facing retail role covering service, operations, stock control, product placement and team delivery.

EDUCATION

MSc Data Science

Kingston University | Expected Sept 2026

Focus: Python, SQL, statistics, machine learning, data warehousing, data quality, applied modelling and operational data analysis.

Module marks: Data Analytics & Visualisation 99; Databases & Data Management 96; Machine Learning & Deep Learning 82; Applied Data Programming 81.

BA Philosophy - First Class Honours

Newcastle University | 2017 - 2021

Developed analytical reasoning, evidence evaluation, close reading and written communication.

Front-End Web Development Certificate

Trilogy Education | Nov 2022 - Mar 2023

Scored 100% on an intensive bootcamp covering HTML, CSS, JavaScript, React, Git, APIs and testing.