Shauna Heron

Data Scientist · Clinical & Health Analytics · R and Python

Sudbury, ON · Remote only
sheron@laurentian.ca · linkedin · personal website · github

Summary

Data scientist with five years of applied experience turning messy, real-world data into production-quality analytics. Currently the sole data lead for a community mental health agency, where I build predictive models, reproducible reports, and interpretable machine-learning workflows that clinicians and executives rely on daily. Graduate training in experimental psychology and machine learning underpins the statistical depth of the work. I look for roles where the problem is genuinely hard, the data is genuinely messy, and the work matters.

Technical Skills

Languages: R (daily driver), Python, SQL, Bash
R ecosystem: tidyverse, tidymodels, ggplot2, lme4, Shiny, Quarto
Python ecosystem: pandas, scikit-learn, PyTorch, Hugging Face, spaCy, LangChain
Statistics & modelling: Generalized linear and mixed-effects models; Bayesian hierarchical models; supervised and unsupervised machine learning; model interpretability (SHAP); fairness auditing
Data: Longitudinal electronic health record (EHR) analytics; feature engineering; cohort construction; SQL/SQLite
Reporting & delivery: Quarto (HTML, PDF, DOCX); Shiny dashboards; Power BI; Git-based reproducible workflows
Applied AI: Retrieval-augmented generation (RAG); local LLM integration (Ollama); vector databases (ChromaDB)
Tools & OS: Positron, RStudio, VS Code, Jupyter, Neovim · Linux, macOS, Windows

Professional Experience

Data Analyst — Office of Systems Performance
Compass Child & Youth Mental Health, Sudbury, ON · 2023–Present

Sole data analyst for a mental health agency of 100+ staff serving thousands of clients annually across counselling, walk-in, and supervised-access programs.

- Own the full analytics pipeline for a longitudinal EHR dataset spanning 2013 to present — extraction, cleaning, linkage, feature engineering, and modelling — in R and Python.
- Built and maintain reproducible Quarto-based reporting that produces weekly, monthly, and quarterly reports used by leadership, program managers, and clinical teams.
- Developed interpretable machine-learning models in R to forecast clinician workload from EHR data; operational dashboard integration in progress.

- Designed and deployed an internal retrieval-augmented-generation (RAG) system over agency policy documents, reducing time-to-answer for staff policy questions.
- Implemented two validated organizational-health instruments for annual workplace monitoring; analyze year-over-year results and present findings to executive leadership.
- Translate complex analytical findings for non-technical audiences — clinicians, program managers, and executives — through briefings, reports, and ad-hoc analysis.

Prior technical roles — Netherlands & Canada

Over a decade of earlier technical work including web development, front-end and UI work, and technical systems support across multiple agencies and in-house teams.

Education

MA, Experimental Psychology *(Computational Science focus)* — Laurentian University, 2023–2026 *(thesis submitted, pending oral defense)*

Granted special permission to complete graduate-level Computational Science coursework (Artificial Intelligence & Machine Learning, Data Mining).

Thesis: Beyond Counting Clients — Developing a Fair Measure of Clinical Workload with Machine Learning. Applied interpretable ML to longitudinal EHR data from a community mental health agency; used multilevel variance decomposition to quantify how much workload is actually predictable from client-level intake data (answer: less than most complexity-weighted caseload tools assume). Supervisors: Dr. Michael Emond, Dr. Kalpdrum Passi.

BSc Honours, Applied Science / Psychology Specialization — Laurentian University, 2019–2022 *Magna Cum Laude.*

Awards: Ontario Graduate Scholarship (2025); IAMGOLD President's Innovation Fund (2025).

Selected Projects

Youth Hub Visit Tracking System *(deployment Feb 2026)* — Designed a QR-based check-in and analytics system backed by a relational SQLite database, for longitudinal tracking of youth program utilization at a community hub.

Policy Document RAG Assistant — Production RAG system over agency policy documents; local LLM inference, vector store, retrieval evaluation.

Other

Published preprint in autonomous human–robot interaction (affect-adaptive dialogue systems on the Misty-II platform; under peer review). arxiv.org/abs/2603.00154