

# Shauna Heron

Data Scientist | Clinical & Health Analytics | R & Python

sheron@laurentian.ca • linkedin.com/in/shauna-heron-1885247 • github.com/shaunaheron2

## Professional Summary

I'm a data scientist who thrives on the parts of the job most avoid: cleaning and linking messy EHR records, extracting data from disparate sources (structured and unstructured), and turning high-dimensional datasets into reproducible pipelines. Having spent the last three years as a sole data lead, I've built everything from automated clinical reports and RAG systems to dynamic workload forecasting models. I've proven I can build the plane while flying it; now, I'm looking to join a collaborative data science team where I can continue to scale my skills and grow as an engineer.

## Technical Skills

- **Languages:** R (tidyverse, tidymodels, ggplot2, leaflet, cansim), Python (pandas, scikit-learn, LangChain, spaCy, PyTorch, BERTopic), SQL.
- **Statistics & Modeling:** Supervised & unsupervised ML; Cluster Analysis (K-means, HDB-SCAN); Tree-based models (XGBoost, Random Forest); Neural Networks (PyTorch, Transformers); GLMs & Mixed-effects models; Bayesian hierarchical modeling; SHAP (interpretability); Fairness Auditing.
- **Data Engineering:** ETL scripting, record linkage, unstructured data extraction (PDF, .ics), SQLite.
- **Applied AI:** RAG (ChromaDB, Ollama), NLP, Sentiment Analysis, Human-Robot Interaction.
- **Automation & DevOps:** Quarto, GitHub Actions, Git/Version Control, Bash, Neovim.
- **Operating Systems:** Proficient in **Linux (Ubuntu/Debian)**, **macOS**, and **Windows**; experienced in managing cross-platform development environments and local server configurations.

## Professional Experience

Data Scientist | Compass Child & Youth Mental Health

Sudbury, ON | 2023–Present

*Sole data lead for a mental health agency serving thousands of clients across multiple programs.*

- **ETL & Unstructured Data Extraction:** Built pipelines to ingest structured longitudinal EHR data. Developed custom Python/R scripts to extract and organize data from non-standard sources, including clinical PDFs, websites and .ics calendar files using spaCy and other libraries.
- **Statistical Analysis & High-Stakes Reporting:** Conduct routine and ad-hoc statistical modeling and analysis on clinical outcomes and high-risk metrics (e.g., suicide statistics). Deliver results through reproducible Quarto reports (HTML/PDF) used by executive leadership.
- **Reproducible Reporting & Automation:** Architected a **Quarto** ecosystem that automates the generation of a variety of weekly and quarterly reports (HTML/PDF), replacing manual spreadsheet workflows with reproducible, code-based audits.

- **Population Benchmarking:** Utilized the **cansim** and **statcanR** packages to integrate Statistics Canada data with internal metrics. Compared agency-level clinical outcomes against population-level benchmarks to provide context for executive reporting.
- **Geospatial Analysis:** Developed interactive maps using **Leaflet** to visualize program utilization and client demographics, identifying service gaps and geographic trends for program managers.
- **Applied NLP:** Implemented **BERTopic** for context and sentiment analysis on unstructured survey data, identifying key themes in staff and client feedback.
- **Predictive Modeling & Fairness:** Developed interpretable ML models (XGBoost, PyTorch) to forecast clinician workload; applied SHAP-based explanations and fairness auditing to ensure equitable workload distribution.

### Technical Systems & Web Development (Prior Experience)

*Canada & Netherlands | 2000–2022*

- Managed front-end development, UI/UX design, and technical systems support across multiple agencies and in-house teams.
- Maintained full-stack environments, providing a foundation in systems architecture, cross-platform troubleshooting, and hardware-software integration that now informs my data engineering workflows.

### Education

#### MA, Experimental Psychology (Computational Science Focus)

*Laurentian University | Thesis submitted (Defense pending)*

- **Thesis:** *Beyond Counting Clients — Developing a Fair Measure of Clinical Workload with Machine Learning.*
  - **Dynamic Workload Forecasting:** Engineered a longitudinal forecasting model on a large EHR dataset ( $n > 1000$ ) to predict client-related clinician workload 28 days into the future. Developed a rolling-window framework to generate direct-hour estimates at weekly intervals, accounting for dynamic assessment scores and client contact history.
- **MIMIC-III Data Mining Project:** Independently architected a full ETL and modeling pipeline for the high-dimensional MIMIC-III dataset. Engineered features from raw ICU vitals and labs to phenotype pneumonia types via unsupervised learning, validating clusters against clinical outcomes.
- **Awards:** Ontario Graduate Scholarship (2025), IAMGOLD President’s Innovation Fund (2025).

#### BSc Honours, Applied Science / Psychology Specialization

*Laurentian University | Magna Cum Laude*

### Selected Projects

#### Human-Robot Interaction Framework (2025–2026)

*Python, Gemini API, LangChain, Deepgram*

- Built a responsive framework for the Misty II robotics platform integrating real-time NLP and affect detection for autonomous social interaction. arXiv:2603.00154.

### **Youth Hub Tracking System (2026)**

*SQLite, R, Quarto*

- Designed a QR-based check-in system and relational database to track longitudinal program utilization at a community hub.