Farrukh Nauman

Senior Data Scientist | Data Analytics & Quality, SQL/Python, Agents | PhD

farrukh.nauman@inertialrange.com | (+46) 0702984959 | fnauman.com | inertialrange.com
Swedish citizen | LinkedIn: fnauman | Github: fnauman

Summary

Senior data scientist and analytics consultant focused on data quality, SQL/Python analytics, stakeholder reporting, and agent-assisted automation. I work hands-on in production data platforms, turn messy data problems into reliable checks and dashboards, and use coding agents where they improve speed without weakening correctness. Selected results:

- **∼1 week**: Cross-platform data-quality validator from zero to live dashboard (Databricks → Snowflake).
- **5–10× faster, 50%+ cheaper**: SQL pipeline optimization with row-level correctness proof.
- **3,000+ people impacted**: Python/SQL/Power BI analytics for public-sector decision support.
- **IC + lead**: Senior contributor and interim lead for a 4–7 person DS/advanced analytics team.

Target Role Fit

- **Applied analytics & reporting**: dashboards, stakeholder summaries, metric/quality checks, ad-hoc analysis, and decision support.
- **Data quality**: schema checks, row-count reconciliation, key distributions, date-range validation, row-level SQL correctness oracles.
- **Python analytics**: pandas, NumPy, scikit-learn, statistical analysis, time series, anomaly detection, and visualization.
- **Agents**: coding-agent workflows for validation, analytics automation, SQL refactoring, migration checks, and Text-to-SQL evaluation.

Skills & Tech Stack

Analytics & Quality	Data-quality validation, metric checks, dashboarding, reporting automation, row-level reconciliation, anomaly detection, stakeholder narratives
SQL & Platforms	SQL (advanced), Snowflake, Azure Databricks, Spark, Snowpark, Azure Data Factory, warehouse migration validation
Python & Stats	Python (8+ yrs), pandas, NumPy, scikit-learn, SciPy/statistical analysis, time series, predictive modeling, data visualization
Agents & AI	Codex CLI, Cortex Code, Claude Code, LangChain, OpenAI SDK, Text-to-SQL, evaluation datasets, RAG, prompt engineering
BI & Engineering	Power BI, Streamlit, Git, Docker, CI/CD, REST APIs, Azure DevOps, Azure Functions, C/C++, High Performance Computing
Collaboration	Stakeholder management, product/engineering communication, project scoping, documentation, mentoring, independent delivery
Languages	English (Fluent), Swedish (SFI C2), Urdu (Native)

Experience

InertialRange Labs AB - Self-employed — Linköping, Sweden
Principal Consultant - Data Analytics, Quality & Agents — *Aug 2025 - Present*

Engagement: Senior DS / Interim Lead – Data Science & Advanced Analytics (Sep 2025 - Jun 2026):
Enterprise Data Platform Engagement

- Senior IC and interim lead for a 4–7 person DS/analytics team; own analytics roadmap, stakeholder steering, and technical direction across telemetry analytics, data quality, ML, and GenAI automation.
- Coordinate **Databricks → Snowflake** transition planning/execution while maintaining daily Spark/BI workloads over several terabytes of sensor and production telemetry data.
- Built a **cross-platform data-quality validation system** using AI coding agents: automated comparison of schemas, row counts, key distributions, and date ranges across dozens of tables; shipped from zero to live dashboard in ∼1 week. (Case study)
- Connected **Databricks and Snowflake CLIs** through coding agents (Codex CLI / Cortex Code CLI) to orchestrate repeatable validation and reconciliation workflows, replacing multi-day manual runbooks.
- Redesigned a critical daily SQL pipeline from full recompute to **validated incremental processing**: 5–10× speedup with row-level correctness proof, projected to cut compute costs by **50%+**; rejected a faster but incorrect alternative using multi-scale benchmarks and `EXCEPT`-based validation. (Case study)
- **Tech**: Snowflake, Databricks/Spark, Snowpark, SQL, Python, Streamlit, Codex CLI, Cortex Code, LangChain.

Independent projects (separate from client engagement):

- Built a **Text-to-SQL pipeline against a production ERP database** (hundreds of tables, live data, not a tutorial schema): schema-aware retrieval over a wide schema, a semantic layer linking business vocabulary (e.g. "biggest buyers", "SKUs that stopped selling") to columns and join paths, and intent-level evaluation on results—valid SQL hitting a half-empty column still returns zeros and gets logged as success.
- Open-sourced **ts-agents** (PoC): autonomous time-series agent framework applied to machine activity recognition from sensor data.
- Developed a practitioner framework for **large code migrations with coding agents**: evidence-preservation methodology, behavioral oracles, and `AGENTS.md`-based quality contracts. (Case study)

RISE Research Institutes of Sweden AB — Linköping, Sweden
AI Researcher & Consultant — *Jul 2021 - Aug 2025*

Project Lead: Sustainable Fashion AI Automation (2022-2025: *24 months*): Led two major initiatives: Vinnova: AI for Circular Fashion (**Project Lead**, ∼9M SEK) and CISUTAC (AI Lead, ∼2M SEK).

- Built data-centric automation for quality assessment: custom annotation/review tools, dataset improvement loops, model validation, and stakeholder-facing summaries.
- Used Python analytics on annotation data, metadata, and model outputs to identify data-quality gaps and prioritize collection/improvement work.
- **Impact**: 40% reduction in processing time, 50%+ reduction in data collection costs through synthetic data.
- **Tech**: Python, pandas/NumPy, PyTorch, Vision Transformers, CLIP, Streamlit/Flask, Gradio, Docker, synthetic data.
- **Recognition**: 1 of 5 projects at EU sustainable AI (2023).

Low-Energy IoT Solutions (2022: *4 months*):

- Identified miniROCKET-based time-series methods for industrial edge deployment, enabling real-time sensor analysis with ∼90% lower hardware cost than heavier alternatives.

Additional Analytics & AI Work:

- Delivered hard-brake detection/anomaly analysis for autonomous bus telemetry, factor analysis for traffic congestion, and mentorship for student projects spanning time-series forecasting, image embeddings, and edge AI.

2MNordic IT Consulting AB — Gothenburg, Sweden
Data Scientist & Data Engineer — *Dec 2019 - Jun 2021*

Project: Early Warning System for Student Performance (2020: *6 months*):

- Developed predictive analytics for Helsingborg schools, identifying absence and English/Math grade patterns in 6th grade as leading indicators of 9th grade outcomes.
- Built Python/SQL analysis workflows and Power BI reporting for stakeholder review, enabling early intervention for 10% of the student population and policy recommendations impacting 3,000+ students.
- **Tech**: Python, pandas, NumPy, SQL, Power BI, Azure DevOps, Azure Functions, Data Factory.

Project: Mathematics Assessment Optimization (2021: *4 months*):

- Conducted statistical analysis of digital mathematics test results across 8 schools, identifying misalignments between digital assessment outputs and established grading schemes.
- Produced clear findings for non-technical stakeholders; results informed district-wide grading-policy adjustments and improved assessment accuracy.
- **Tech**: Python, pandas, NumPy, scikit-learn, statistical analysis, data visualization, Azure Notebooks.

Previous Research Positions — 2009–2019

- **Research Fellow, Chalmers University of Technology**: Complex systems modeling, large-scale data analysis, simulation-based inference — Gothenburg, Sweden, 2018–2019
- **Research Scientist, Niels Bohr Institute**: Simulation, forecasting, computational modeling, uncertainty-aware analysis — Copenhagen, Denmark, 2015–2018
- **Research Assistant/PhD Student, Univ. of Rochester**: Data analysis, predictive modeling, high-performance numerical workflows — New York, USA, 2009–2015

Education & Certifications

Microsoft Certified — Azure
Azure Data Engineer Certificate — *2020*

University of Rochester — Rochester, New York (USA)
PhD in Physics and Astronomy — *Oct 2015*
Focus: Complex Systems Modeling, Data Analysis, Computational Fluid Dynamics, High Performance Computing, C/C++