

RASOUL SAMEI

Data Scientist

Bergamo, Italy | +39 320 444 6158 | samei@live.com | linkedin.com/in/rsamei | rsamei.github.io

PROFILE

Data scientist with an MA in Economics and Data Analysis and 3+ years applying quantitative and statistical methods to large, complex datasets in health research, financial services and enterprise settings. Currently at the Mario Negri Institute for Pharmacological Research, working alongside statisticians and clinicians on clinical study data and co-author of a 2025 JAMA paper. Experienced in Python, R and SQL; in building validated, reproducible analytical datasets from fragmented multi-source data; and in communicating results to non-technical stakeholders.

CORE COMPETENCIES

- **Statistical & Quantitative Analysis:** Hypothesis testing, regression modelling, time series forecasting, multivariate analysis (PCA, MCA, FAMD), model evaluation and validation, exploratory data analysis.
- **Statistical Software & Programming:** Python (pandas, NumPy, scikit-learn, XGBoost, Matplotlib), R, SQL, SAS (competition experience), Stata-transferable workflow; Git for reproducible research.
- **Data Management & Linkage:** Harmonising and linking fragmented multi-table databases into validated analytical datasets; ETL/ELT pipeline design; data quality and validation; Google BigQuery, PostgreSQL, Databricks, Azure Synapse.
- **Health & Regulated Data:** Occupational and health data standards (HL7 FHIR R4, OMOP Common Data Model); regulated-data handling in clinical research and financial conduct settings.
- **Machine Learning:** Supervised and unsupervised learning, anomaly detection, clustering, feature engineering, NLP; MLOps (MLflow, FastAPI, Docker).
- **Communication:** Technical documentation, stakeholder reporting, Power BI dashboards, peer-reviewed academic writing and conference presentation.
- **Languages:** Persian (native), English (professional working proficiency), Italian and German (basic).

PROFESSIONAL EXPERIENCE

Data Scientist — Istituto di Ricerche Farmacologiche Mario Negri | Ranica, Italy | 2024 – Present

Independent biomedical research institute

- Co-authored a peer-reviewed paper published in JAMA (2025) on the development and validation of the SOFA-2 score, contributing ETL and data processing, statistical analysis and manuscript review.
- Engineered a 900+ line SQL pipeline on Google BigQuery that transformed a highly fragmented, multi-table database into a single validated analytical table, now reused as a standard data asset by other teams and external researchers — directly analogous to harmonising and linking heterogeneous survey and policy data.
- Automated recurring processing and reporting workflows for large study datasets, standardising outputs and cutting manual handling across multiple projects.
- Partnered with statisticians, clinicians and software engineers to translate ambiguous domain requirements into reproducible, documented analytical deliverables.
- Designed an interactive analytics dashboard with an embedded LLM assistant, enabling non-technical researchers to explore study data in natural language.

Data Scientist (Intern) — The Openwork Partnership | Swindon, United Kingdom | 2023 – 2024

Financial services / wealth advisory

- Built an end-to-end unsupervised anomaly detection system in Azure Machine Learning to surface irregular patterns in regulated mortgage advisory conduct data for compliance review; published the methodology as a peer-reviewed conference paper (ICSEM 2025).
- Defined, computed and validated 28+ business KPIs and delivered them through Power BI dashboards for senior supervisors.
- Developed scalable data pipelines (Azure Python SDK v2, Spark SQL) to process and manage large-scale datasets; presented findings to non-technical stakeholders.

Machine Learning Teaching Assistant — University of Bergamo | Bergamo, Italy | 2023

- Led group tutoring sessions on applied machine learning, coaching students through model building, evaluation and end-to-end data analysis workflows.

EDUCATION

- **Master of Arts, Economics and Data Analysis** — Università degli Studi di Bergamo | Bergamo, Italy | 2024
- **Bachelor of Arts, Business Management** — Islamic Azad University of Neyshabur | Neyshabur, Iran | 2015

PUBLICATIONS

- Ranzani OT, et al. (contributing author: Samei R). Development and Validation of the Sequential Organ Failure Assessment (SOFA)-2 Score. *JAMA* (2025). doi:10.1001/jama.2025.20516
- Shalchilar M, Samei R, et al. Assessing the Efficacy of a Sequential General Variational Mode Decomposition-Based Combination Model for US Wind Power Forecasting. *Modeling Earth Systems and Environment* (2025). doi:10.1007/s40808-025-02693-5
- Samei R. Anomaly Detection in Financial Advisory Services: A Machine Learning Approach for Mortgage Conduct of Business Advisers. *Proceedings of ICSEM* (2025).

SELECTED TECHNICAL PROJECTS

- **Standards-Based Health Data Integration:** Mapped complex nested records conforming to HL7 FHIR R4 into OMOP-style conformed staging tables (Bronze/Silver/Gold medallion pipeline on Databricks), served through documented FastAPI endpoints — hands-on experience with the standards-based data harmonisation central to multi-cohort research.
- **NLP-Driven Risk Prediction Platform:** Built an NLP pipeline (spaCy) extracting structured entities from unstructured free-text records, then trained XGBoost classifiers for high-stakes outcome prediction with full experiment tracking (MLflow).
- **Retrieval-Augmented Generation Assistant:** Developed a semantic search and question-answering assistant over a private document corpus (LangChain, PostgreSQL/pgvector) with audit logging for traceable, source-grounded answers.

AWARDS

- **First Place, Stat-Hackathon Competition** — Università degli Studi di Bergamo (2021). Advanced multivariate analytics (PCA, MCA, FAMD) delivered in Python, R and SAS under competition time constraints.