

RISHAV SEN

Irvine, California | (949) 378-3965 | Rishavsen28@gmail.com | linkedin.com/in/rishav-sen111 | github.com/VahsirRush

SUMMARY

Data Analyst and quantitative researcher who works daily in Python, SQL, and R to design ETL pipelines, build statistical and machine learning models, and turn messy, multi-source data into reliable, decision-ready systems. Experience spans consolidating data from 20+ source systems into centralized databases, building transformer-based and factor-based quantitative models, and supporting scalable data warehousing with Databricks, PostgreSQL/PostGIS, AWS, and Docker, with a consistent focus on data accuracy and automation over manual effort.

TECHNICAL SKILLS

- **Programming:** Python, SQL, R, C, HTML/CSS
- **Quantitative Analytics:** Factor modeling (Barra-style risk models), transformer architectures (MASTER replication), backtesting, cost-aware performance attribution
- **Data Engineering:** ETL pipeline development, data warehousing, Databricks, ArcGIS, PostgreSQL/PostGIS
- **Cloud & Infrastructure:** AWS, Docker
- **Automation:** Power Automate, workflow automation
- **Statistical Methods:** Change-point detection (CUSUM), ablation analysis, time-series analysis
- **Certifications:** Databricks
- **AI:** Agentic workflows, Claude Suite, Qwen, Deepseek, GPT Suite

EDUCATION

Bachelor of Science in Business Administration, Business Analytics, University of San Francisco May 2025
• Relevant Coursework: Data Visualization, Data Wrangling & Digital Analytics, Advanced Business Applications of Analytics, Data Mining, Machine Learning, Data Science

WORK EXPERIENCE

Data Analyst / Payroll Analyst | Sonrava Health June 2025 - Present
• Designed and automated ETL workflows to aggregate unstructured HR data from **20+ source systems** into a centralized database, replacing manual processes for an organization of 2,000+ employees.
• Built and maintained Power BI dashboards, using **SQL**-aggregated data, visualizing core HR KPIs for executive leadership, enabling data-driven decision-making across roughly 600 offices.
• Engineered data pipeline architecture supporting long-standing HR workflows, improving data consistency and reducing manual reporting time.
• Managed payroll processing and doctor bonus calculations for the Northern California region, ensuring accuracy across compensation data systems.

Regional Patient Care Coordinator | Western Dental and Orthodontics Aug 2024 - May 2025
• Coordinated patient care data across multiple dental offices while maintaining HIPAA-compliant handling of sensitive records.

Marketing Analytics Intern | RippleMatch Jan 2023 - May 2023
• Analyzed engagement data from 800+ monthly active users and tracked outreach performance metrics to guide marketing and product strategy decisions.

PERSONAL PROJECTS

MASTER Transformer Replication & Cost-Aware Backtest Engine Dec 2025 - Sep 2026
• Ported the MASTER (Market-Guided Stock Transformer, AAAI 2024) architecture to US equities on a point-in-time backtest engine, benchmarked against ridge and LightGBM baselines with full ablation analysis.
• Built turnover-aware, net-of-cost performance reporting and gate-activation analysis to test the model for regime-conditional factor timing.
Tools/Methods: Python, PyTorch, pandas, scikit-learn, LightGBM, yfinance, SEC XBRL filings, point-in-time backtesting, ablation analysis, Barra-style factor decomposition

CRISPR Guide RNA Efficiency Prediction (CNN-Attention, Calibrated Uncertainty) May 2026 - Sep 2026
• Built a convolutional attention model in PyTorch predicting CRISPR-Cas9 guide RNA on-target efficiency, benchmarked against a reproduced Rule Set 2 (GBT) baseline on the Doench et al. 2016 dataset, with attention-weighted pooling for position-level interpretability.
• Implemented split conformal prediction to generate calibrated uncertainty intervals on every prediction, addressing a gap in most public efficiency models that report point estimates without confidence bounds.
Tools/Methods: Python, PyTorch, scikit-learn, pandas, Doench et al. 2016 dataset, convolutional neural networks, self-attention, split conformal prediction, baseline benchmarking