Jack Stehn

Machine Learning Engineer | Data Engineer | Data Scientist

San Francisco, California • jack@ste.hn • www.linkedin.com/in/jack-stehn • github.com/jstehn • ste.hn

PROFESSIONAL SUMMARY

High-agency Data Professional with a unique background blending rigorous social science research with production-grade engineering. I specialize in bridging the gap between 'notebook data science' and scalable infrastructure. I architect end-to-end data systems—from raw ingestion to deploying predictive models—and bring a software engineering mindset to data teams.

EXPERIENCE

Data Scientist (Lead: ML, Data Engineering, MLOps) - Ed Pioneers Fellow   Aug 2024 — Sep 2025

Caliber Public Schools

- **Strategic Leadership (Solo Data Lead):** Owned the full data lifecycle (DS, DE, ML) as the sole data scientist. Partnered directly with C-suite and department heads to navigate a 'zero-to-one' environment.
- **Predictive ML & Risk Modeling:** Developed and deployed explainable ML models to predict staff turnover. Engineered a 'Risk Tolerance' configuration allowing non-technical leadership to adjust precision/recall thresholds.
- **Modern Data Stack Architecture:** Architected a scalable platform on GCP. Orchestrated ELT pipelines using Dagster, dbt, and dlt to ingest data from disparate SIS and HR platforms.
- **Engineering Maturity (ROI):** Engineered a comprehensive People Team data pipeline, reducing manual consistency checks from months of collective annual work to seconds.
- **Survey Design & Enrichment:** Leveraged data to enrich staff satisfaction surveys with demographic, work location, role, grade level, and tenure data, enabling highly segmented and actionable insights.
- **Data Democratization:** Updated the organization's data security policy and designed data literacy training modules, empowering school leaders to access real-time metrics without technical bottlenecks.

Data Scientist (ML, Data Engineering, MLOps)   Jul 2021 — Jan 2023

SetSail

- **Business Impact:** Contributed to product enhancements that achieved 33% faster ramp times, 16% higher revenue, and 15x ROI for customers.
- **Production ML (Revenue):** Developed and deployed production ML models for Propensity Scoring and Churn Modeling. Leveraged NLP on unstructured email metadata to identify sales signals.
- **Pipeline Architecture (AWS):** Led a critical overhaul of the AWS data infrastructure. Implemented 'SQL Push-down' strategies and asynchronous DAGs, reducing data processing latency by 75%.
- **Causal Analysis:** Performed deep causal inference studies to isolate specific sales behaviors that drive outcomes, influencing the product roadmap to focus on 'high-leverage' user actions.
- **Engineering Best Practices:** Championed the adoption of CI/CD pipelines (GitHub Actions), unit testing (pytest), and Agile methodologies.

Data Science Research Team Lead   Aug 2020 — Apr 2021

UC Berkeley School of Public Health

- **Leadership:** Led data science components for mixed-methods studies on equity and public health. Managed a team of undergraduates.
- **Unstructured Data:** Analyzed diverse unstructured and non-traditional datasets requiring the development of novel data processing approaches.
- **Geospatial Analysis:** Performed geospatial analysis (ArcGIS) to identify and visualize spatial patterns for non-technical stakeholders.
- **Visualization:** Created interactive dashboards (Tableau, Plotly) to communicate findings to stakeholders.

EDUCATION

University of California, Berkeley   Jul 2019 — Apr 2021

Bachelor of Arts in Data Science (Domain Emphasis: Quantitative Social Science)

GPA: 4.00/4.00

Highest Distinction (Summa cum laude). Outstanding Data Science Undergraduate Award (Top of Class).