

# 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 specializing in bridging the gap between 'notebook data science' and production-grade scalable infrastructure. I architect end-to-end systems—from raw ingestion to deploying predictive models that drive revenue. I bring a software engineering mindset to data teams, championing CI/CD, unit testing, and modular design to untangle complexity and build maintainable tools.

## EXPERIENCE

### Data Scientist (Lead: ML, Data Engineering, MLOps) - Ed Pioneers Fellow

Sep 2024 — Oct 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, moving the org from manual spreadsheets to automated warehousing.
- **Predictive ML & Risk Modeling:** Developed and deployed explainable ML models (Logistic Regression, Deep Learning) to predict staff turnover. Engineered a 'Risk Tolerance' configuration allowing non-technical leadership to adjust precision/recall thresholds based on quarterly hiring capacity.
- **Modern Data Stack Architecture:** Architected and built a scalable platform on Google Cloud Platform (GCP). Orchestrated ELT pipelines using Dagster, dbt, and dlt to ingest data from disparate SIS (School Information Systems) 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. Built a custom automated data quality framework to catch inconsistencies between systems.
- **API Integration & Interoperability:** Solved complex data interoperability challenges by designing integrations between Schoolmint and PowerSchool. Reverse-engineered undocumented APIs to create a unified data model for cross-functional analysis.
- **Data Democratization:** Updated the organization's data security policy and designed data literacy training modules, empowering school leaders to access real-time attendance and academic metrics without technical bottlenecks.
- **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.
- **Stakeholder Management:** Partnered with leadership on high-impact, data-driven solutions; presented findings to stakeholders including the Board of Directors.

### Data Scientist (ML, Data Engineering, MLOps)

Aug 2021 — Feb 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 (deal closure probability) 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 (S3, Athena, EMR). Implemented 'SQL Push-down' strategies and asynchronous DAGs, reducing data processing latency by 75% and scaling to multi-terabyte datasets.
- **Scalable Data Modeling:** Architected scalable Star Schema data models and optimized ETL/ELT processes, ensuring data readiness for LLM integration.
- **Engineering Best Practices:** Championed the adoption of CI/CD pipelines (GitHub Actions), unit testing (pytest), and Agile methodologies within the data science team.
- **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.
- **Technical Consulting:** Acted as a technical consultant for enterprise customers, diagnosing complex data discrepancies and proposing architectural solutions for data integration.
- **Cross-Functional Leadership:** Collaborated seamlessly with Engineering, Product, and Support teams to spec out large-scale infrastructure restructuring.

## Data Science Research Team Lead

UC Berkeley School of Public Health

- **Leadership:** Led data science components for mixed-methods studies on equity and public health (specifically violence against homeless youth). Managed a team of undergraduates.
- **Unstructured Data:** Analyzed diverse unstructured and non-traditional datasets (qualitative interviews, geospatial data, text corpora, hand-drawn maps) requiring the development of novel data processing approaches.
- **Geospatial Analysis:** Performed geospatial analysis (ArcGIS) to identify and visualize spatial violence patterns for non-technical stakeholders.
- **Visualization:** Created interactive dashboards (Tableau, Plotly) to communicate findings to stakeholders.
- **Interdisciplinary Collaboration:** Collaborated across disciplines (public health, psychology, sociology) ensuring ethical, robust research.

## Full Stack Engineer

May 2019 — Aug 2019

Los Medanos College

- **Full Stack Development:** Independently designed and developed a web application (Python, JavaScript, PHP, PostgreSQL) to guide student program selection.
- **Data Acquisition:** Built a database of 500+ university programs via web scraping (Beautiful Soup) and rigorous data cleaning.
- **Stakeholder Consulting:** Consulted with college stakeholders (district, counselors) to define user needs.
- **UX/UI:** Created an intuitive management dashboard for non-technical staff.

## EDUCATION

### University of California, Berkeley

Aug 2019 — May 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).