Jack Stehn

Data Scientist

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

---

PROFESSIONAL SUMMARY

Data Scientist expertly blending rigorous social science research with production-grade engineering. I specialize in the full ML lifecycle—translating complex data into actionable insights and deploying explainable predictive models that drive measurable business outcomes. I am focused on bridging the gap between statistical foundations and scalable, ethical applications.

EXPERIENCE

Data Scientist (Lead: ML, Data Engineering, MLOps) - Ed Pioneers Fellow   Sep 2024 — Oct 2025
Caliber Public Schools

- **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.
- **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.
- **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:** Designed data literacy training modules, empowering school leaders to access real-time attendance and academic metrics without technical bottlenecks.
- **Stakeholder Management:** Partnered with leadership on high-impact, data-driven solutions; presented findings to stakeholders including the Board of Directors.

Data Scientist (ML, NLP, Causal Inference)   Aug 2021 — Feb 2023
SetSail

- **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.
- **Business Impact:** Contributed to product enhancements that achieved 33% faster ramp times, 16% higher revenue, and 15x ROI for customers.
- **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.
- **Scalable Data Modeling:** Architected scalable Star Schema data models and optimized ETL/ELT processes, ensuring data readiness for LLM integration.
- **Technical Consulting:** Acted as a technical consultant for enterprise customers, diagnosing complex data discrepancies and proposing architectural solutions.

Data Science Research Team Lead   Sep 2020 — May 2021
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.

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).