

Taylor Haaff

(260) 494-8867 taylorhaaff@gmail.com [LinkedIn](#) [GitHub](#)

TECHNICAL SKILLS

Languages & Tools: Python (Pandas, Polars, NumPy, Scikit-Learn, PyTorch), SQL, R, DBT, Jupyter, Git/GitHub

Infrastructure & Platforms: BigQuery, Google Cloud, AWS, Segment, PostHog, Statsig, Hex, Tableau, Dagster

Methods: A/B experimentation (Bayesian & Frequentist), hypothesis testing, power analysis, variance reduction, causal inference, cohort & retention analysis, applied ML (feature engineering, hyperparameter tuning, model evaluation)

AI & LLM: MCP server & tool design, custom Claude plugin/skill dev, agent loop design, structured outputs, LLM evaluation frameworks (golden datasets, sampling strategy, human-in-the-loop annotation, preference data, statistical validity checks), RAG architecture.

WORK EXPERIENCE

Clerk - San Francisco, CA, 2025 - Present

Data Scientist - Data & Growth

- Architected and owned the core data modeling layer in DBT and BigQuery, serving as the single source of truth across growth, product, and finance. Consolidated 20+ ad-hoc queries into reusable models. Cut query time from minutes to seconds.
- Designed and shipped event-driven, multi-source ingestion and validation pipelines (Segment, PostHog, BigQuery) across 8+ heterogeneous sources. These form the measurement and experimentation backbone for product, growth, and engineering.
- Led rigorous A/B experimentation across activation, engagement, conversion using Bayesian and Frequentist methods (power, sequential, sampling strategy, variance reduction, statistical validity checks). Defined company-wide experimentation and results reporting standards and decision criteria.
- Drove cohort, retention, funnel, and lifecycle analyses which helped define OKRs. Performed EDA surfacing trends, correlations, and anomalous drop-off patterns. Translated insights into roadmap changes that shaped product strategy.
- Ran observability, capacity planning, validation, and cost optimization for the data platform. Instrumented pipeline monitoring, made quantitative storage/exposure/data-volume tradeoffs that held costs flat QoQ through a 3x DAU increase.
- Designed, built, and shipped internal LLM agentic automations end-to-end with the Claude API using MCP servers, custom tools/function calling, and multi-step agent loops. Defined evaluation methodology including golden datasets, sampling strategy, and statistical validity checks to ensure trustworthy, reproducible results across model iterations and use cases.

Hotplate Inc. - San Francisco, CA, 2023 - 2025

Founding Software Engineer - Data Science

- Architected the consumer-marketplace data and metrics platform from zero as one of the first four hires, including ingestion, schemas, transformation layers, metrics, and analytics dashboards. Scaled to 2,400+ active users and \$50M+ GMV.
- Partnered with engineering and product to design and implement user-event telemetry, KPIs, instrumentation contracts, and evaluation frameworks for feature launches. Turned product hypotheses into concrete eval sets, then measured whether shipped changes moved the targeted metric.
- Built time-series forecasting, regression, and anomaly-detection models that served as the primary decision-making interface for leadership, delivering actionable projections on user growth, revenue trajectory, and operational planning.
- Translated ambiguous business questions into structured and rigorous analytical frameworks. Worked directly with non-technical leadership to size problems, identify levers, and communicate cohesive narratives that shaped company strategy.

Progressive Insurance - San Francisco, CA, 2022 - 2023

Data Analyst II - Staffing Analytics

- Built and optimized ML pipelines with built-in evaluation and anomaly-detection layers (Python, random forests, XGBoost) for claims volume forecasting at enterprise scale for 19,000+ claims representatives across a portfolio valued at \$300B.
- Developed reporting pipelines and dashboards (Tableau, Dash) that served as the primary analytical interface for senior stakeholders, translating complex workforce data into recommendations that drove resource allocation decisions at scale.

Physicians Health Plan - Fort Wayne, IN, 2020 - 2021

Business Analyst - Analytics

- Designed repeatable anomaly-detection pipelines in Python and R, applying statistical models (logistic regression, decision trees) to unstructured healthcare claims data to improve fraud identification consistency across the analytics team.
- Led three cross-departmental data initiatives end-to-end, building time series forecasting pipelines (ARIMA, sliding window KS testing) in R and Python that improved operational planning accuracy and established shared data practices across teams.

Indiana University School of Optometry Vision Development Lab - Bloomington, IN, 2018 - 2020

Faculty Research Associate

- Designed and built systems integrating computer vision, infrared depth mapping, and novel sensor hardware to capture behavioral data in one of the world's only head-mounted infant eye-tracking studies. Pipelines operated in a resource-constrained research environment with no existing tooling or precedent.
- Analyzed large, noisy sensor datasets using cross-correlation, factor analysis, and regression to extract reliable signal, applying rigorous statistical methods in a research environment.

EDUCATION

Indiana University - Bachelor's in Cognitive Psychology, 2013 - 2017

Purdue University - Pre-Medical Post Baccalaureate, 2021-2022