

Emily Hough-Kovacs

+1 (603) 484-2534 | San Francisco, CA | emilyhoughkovacs@gmail.com
www.emilyhk.com | linkedin.com/in/emilyhoughkovacs

Summary

Senior data scientist with 10+ years turning messy, large-scale product data into rigorous measurement and evidence-based decisions at Reddit, Meta, and Spotify. Defines source-of-truth metrics that teams rely on, runs observational analyses that quantify real-world impact, and translates ambiguous problems into measurable ones for technical and non-technical audiences. Driven by AI safety, and set on bringing that same measurement rigor to how frontier models behave. Building depth in interpretability and alignment.

Skills & Tools

Languages & Libraries: Python (pandas, NumPy, scikit-learn, statsmodels, matplotlib), SQL (expert)

Statistics & Measurement: statistical modeling, observational impact analysis, metric design & source-of-truth definitions, forecasting, model evaluation

Data Platforms & Pipelines: GCP BigQuery, PostgreSQL, Hive, HDFS, Airflow, ETL, data pipelines

AI / ML Tooling: LLM APIs (OpenAI, Anthropic), embeddings & vector search

Communication: Tableau, dashboards, stakeholder-facing analysis for technical and non-technical audiences

Experience

Reddit — Senior Data Scientist, Ads Marketplace March 2026 – Present

- Own data science for the identity matching and attribution domain, standing up its measurement foundation (metric taxonomy, pipelines, reporting) in a system that has no true measurement today; validated engineering's opportunity-sizing against source data.
- Built two Airflow ETL pipelines powering an internal dashboard for ads-engineering data amendments and make-goods during account incident remediation, keeping billed and reported figures consistent while cutting manual ticket overhead.
- Built a supply-and-demand monitoring dashboard surfacing contextual ad-to-post relevance signals across the marketplace, informing inventory and pacing decisions.

Meta — Senior Data Scientist, Reality Labs July 2024 – August 2025

- Built and evaluated a random forest ensemble predicting user purchase intent, assessing performance tradeoffs.
- Built long-range forecasts for Lifestyle and Entertainment teams' key metrics, informing roadmap and resourcing decisions.
- Ran observational analyses of MR vs. fully immersive VR effectiveness, quantifying the impact of physical space on user engagement.

Spotify — Data Scientist May 2016 – May 2021; Aug 2022 – March 2024

- Pioneered the metrics suite for the creator side of the music marketplace, adopted as the source of truth across creator-facing teams.
- Built monitoring for time-to-delivery of podcast content, surfacing supply-chain bottlenecks for podcast partnerships.
- Produced creator-side insights for Spotify Wrapped '22 and '23 (top editorial tracks, artists, albums), surfaced in Spotify's public, consumer-facing product.

Capitol Records, Universal Music Group — Data Scientist November 2021 – June 2022

- Designed, tested, and implemented a metric to measure catalogue consumption for legacy artists.
- Harnessed Shazam data to identify emerging cities in international markets — observational analysis informing A&R investment.

Tumblr — Product Analyst September 2014 – October 2015

- Extracted business insights for a platform serving 250M+ blogs and 60M+ posts/day; automated reporting and built a Flask app to run and email ad-hoc Hive jobs.

Education

Barnard College, Columbia University — B.A., Mathematics / Computer Science May 2014

Projects & Writing

Build Your Own LLM — writeup from a hands-on workshop building a GPT-style model from scratch. emilyhk.com/llm-workshop

Semantle — word-guessing game (Python, gensim, FastAPI): word2vec embeddings + cosine similarity.

Playlstr — playlist generator (Python, MongoDB): NMF over Rap Genius song metadata surfaces similar songs.