

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 developing data products and shaping product decisions at Reddit, Meta, and Spotify. Expert in SQL, statistics, and large-scale analytics, with a track record of partnering across product, engineering, and business teams to build ETL pipelines, design metrics frameworks, and deliver ML-driven insights.

Skills & Tools

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

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

Statistics & Analytics: statistical modeling, metric design, Tableau

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

Engineering: Git, GitHub, Flask

Experience

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

- Own data science for the identity matching and attribution domain within ads engineering; leading opportunity sizing with engineering to augment the identity lookup service and identity graph with third-party advertiser data.
- Built and deployed two Airflow ETL pipelines powering an internal dashboard that enables ads engineering to execute data amendments and make-goods for advertiser account incident remediation, reducing manual ticket overhead.
- Built supply-and-demand 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

- Fine-tuned random forest ensemble model to predict user purchase intent.
- Built long-range forecasts for Lifestyle and Entertainment teams' key metrics, informing roadmap and resourcing decisions.
- Evaluated MR vs. fully immersive VR effectiveness for the fitness use case and quantified impact of physical space on user engagement.

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

- Pioneered metrics suite for the creator side of the music marketplace, adopted as source of truth across creator-facing teams.
- Evaluated and monitored time-to-delivery of podcast content, surfacing supply-chain bottlenecks for podcast partnerships.
- Built and maintained Tableau dashboards tracking performance of internal and external editorial tools.
- Produced creator-side insights for Spotify Wrapped '22 and '23 (top editorial tracks, artists, albums).

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

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

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 execute and email ad-hoc Hive jobs.

Education

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

Projects

Semantle — semantic word-guessing game (Python, gensim, FastAPI). Backend using pre-trained word2vec embeddings and cosine similarity to score player guesses against a target word.

Playlistr — smart playlist generator (Python, MongoDB). Non-negative matrix factorization and pairwise dot product to surface songs with similar artistic influences; song metadata from Rap Genius API.