

Jacob Ferraiolo

Data Engineering | Machine Learning | Data Science

804-929-8620 | ferraioloj@gmail.com | Dallas, TX | [linkedin.com/in/jacob-ferraiolo/](https://www.linkedin.com/in/jacob-ferraiolo/)

Summary

Senior data engineer with a product lens, specializing in architecting and leading the development of scalable data platforms, streaming pipelines, and ML models that solve ambiguous business problems and unlock growth.

Experience

Nerdwallet

Senior Data Engineer | October 2023 - Current

- **Reduced front-end instrumentation effort 70%** by leading a 4-engineer team that migrated shopping-funnel events from custom collectors to Segment and redesigned the warehouse star schema.
- Owned and drove the adoption of an org-wide **Monte Carlo data quality framework**, building 12 core Airflow DAGs and an on-call runbook that cut data QA tickets from 12 → 3 per month.
- Created the department's first XGBoost machine learning model to estimate click value **improving performance marketing return on ad spend by 5%**.
- Built an LLM-powered sentiment summarizer for user reviews; **lifted average SERP rank by +1.5 positions** (monitored via GSC), directly impacting organic traffic acquisition.
- Partnered with the Travel PM to define KPIs for newsletter performance; built a Snowflake data mart and Looker dashboard covering ≈3k pages, enabling A/B tests that **lifted sign-ups 5% in 2 months**.

Data Engineer II | July 2022 - October 2023

- Owned NerdWallet's transaction-data ingestion platform—200+ partner integrations serving 10 embedded DEs—and **cut new-partner onboarding time 30%** (10 → 7 days) by automating schema drift checks and template-based pipeline scaffolding.
- Drove the re-architecture of a monolithic pipeline into 8 domain-scoped Airflow DAGs following an org restructure, enabling full ownership for embedded DEs and eliminating cross-domain outages.
- **Eliminated 100% of manual weekly reporting** for the Podcast team by building an automated Airflow ETL pipeline, saving 260 analyst hours annually and reducing decision latency from 3 days to under 24 hours.

Mission Lane

Data Engineer | April 2021 - July 2022

- Designed and implemented real-time data models and streaming pipelines (Spark, Python, Snowpipe) for 3 new product launches, **enabling day-one analytics and reporting capabilities**.
- Used Snowflake, Cloud Storage, GCP Cloud Functions, and Python to connect to external marketing APIs and **design data pipelines to ingest and store external marketing performance data** allowing Marketing to optimize spend

Southwest Airlines

Associate Data Scientist | February 2020 - April 2021

- Automated ETL for 10+ critical data sources using Python and Airflow, **freeing 40 analyst-hours per day** and eliminating human error in key operational reports.

UDig

Data Scientist | July 2019 - February 2020

- Built a TF-IDF and cosine similarity model that matched 15k research papers to 3k reviewers, **automating a core business process** and replacing a legacy manual lookup tool.

Elephant Insurance

Pricing Analyst / MIS Developer | July 2017 - July 2019

- Designed MSSQL/SSIS pipelines and an XGBoost model that added \$100K ARR in marketing lift.

Education

James Madison University / Harrisonburg, VA — B.A. Economics 3.8