

David Freitag

New York, NY | (262) 271-5365 | dkfreitag@gmail.com

davidfreitag.com | linkedin.com/in/davidkfreitag | github.com/dkfreitag

EXPERIENCE

Senior Data Engineer

February 2023 - Present

American Family Insurance

- Infrastructure Lead on a high-priority AI pilot initiative to improve underwriting risk review accuracy and speed. Responsible for scaling the infrastructure of the AI-powered minimum-viable-product application to additional operating companies, lines of business, and sources of structured and unstructured data. Tech: Python/SQL, GitLab CI/CD, Terraform, GCP Pub/Sub, Docker/Cloud Run, BigQuery, Pydantic AI, LangGraph, Kubernetes.
- Architected and built multiple ETL pipelines to refresh terabyte-size datasets on demand for data science modeling. These datasets powered the models used to re-price AmFam's Auto and Property insurance products in all 50 states. Tech: BigQuery SQL, Airflow, Docker, Kubernetes.
- Improved, automated, and extended a vendor file transfer process for preparing and transferring data securely. Scaled the pipeline to prep and encrypt TB of data in parallel, improving runtimes from days to minutes. Tech: Google Cloud Storage, GCP Secret Manager, GNU Privacy Guard (PGP encryption), Google Batch.

Data Engineer

August 2021 - February 2023

American Family Insurance

- Built ETL pipelines to refresh monthly/quarterly datasets for data science modeling and product analytics. Tech: BigQuery SQL, Airflow, Docker, Kubernetes.

Data Science Intern

March 2021 - July 2021

Pitney Bowes

- Trained, tested, and tuned gradient-boosted machine learning models (XGBoost, LightGBM) to improve the accuracy of forecasting shipping container journey times within the Pitney Bowes network.

EDUCATION

Baruch College, The City University of New York

May 2021

Master of Science, Information Systems, Data Analytics

GPA: 4.0

Honors: Beta Gamma Sigma Honor Society, Beta Alpha Psi Honor Society

PROJECTS

[Build Your First Serverless Data Engineering Project](#) - (4.8 stars, 100+ paid students)

- Intro level course on serverless data engineering in AWS that I designed, marketed, and taught. Tech: REST API's, AWS Lambda, S3, Athena, Kinesis, Glue, Grafana.

Citibike Data Platform - (in progress) - End-to-end data pipeline to ingest data from various Citibike API endpoints, do streaming and batch data processing, and make the data available in a serving layer for both analytics and data visualization. Tech (so far): Kafka, Docker, EC2, Lambda, Glue, Athena, S3, Spark, GitHub Actions, Terraform.