

David Freitag

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

LinkedIn: [linkedin.com/in/davidkfreitag](https://www.linkedin.com/in/davidkfreitag)

GitHub: github.com/dkfreitag

EXPERIENCE

Senior Data Engineer | American Family Insurance

February 2023 - Present

- Project leader: led a team of three engineers to build an ETL pipeline using Google BigQuery, Airflow, Docker, and Kubernetes to refresh AmFam's largest auto modeling dataset, reducing the time-to-market cycle by months.
- Improved automation of multiple vendor file transfer processes using Python, GCP Storage API, Google Secret Manager, and GNU Privacy Guard. Decreased person-hours per transfer from 2 hours to 5 minutes per transfer.
- Mentored a Rotation Analyst in technical skills including Python, BigQuery/SQL, Git/GitLab, Airflow, Terraform, and various services in Google Cloud Platform (GCP) and Amazon Web Services (AWS).

Data Engineer | American Family Insurance

August 2021 - February 2023

- Led my team's Data Quality Initiative. Created best practices and guided implementation of automated processes for monitoring data quality. Automated identification of data quality issues monthly across 15+ datasets.
- Decreased runtime for an insurance re-rating process from 4 days to 1.5 hours by distributing the workload across multiple Kubernetes pods orchestrated with Airflow.
- Developed a fully automated ETL pipeline with Python, SQL, Airflow and GitLab CI/CD to produce a monthly product analysis dataset with 600+ attributes.
- Built a process for ad-hoc AWS Athena to GCP BigQuery data transfers using Airflow. Transferred terabytes of data using this process.

Data Science Intern | Pitney Bowes

March 2021 - July 2021

- Analyzed global container scan data using Snowflake SQL, Python, and statistical testing to provide stakeholders with visibility into the Pitney Bowes network.
- Trained, tested, and tuned a machine learning model using LightGBM to predict cycle time for container journeys between locations within the Pitney Bowes network to improve container arrival forecasts.

EDUCATION

Baruch College, The City University of New York

May 2021

Master of Science, Information Systems, Data Analytics | GPA: 4.0

Relevant Coursework: Data Mining for Business Analytics, Big Data Technologies, Data Warehousing

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

PROJECTS

[Streaming Reddit Posts, Comments, and Stock/Crypto Ticker Data into BigQuery](#)

- **Task:** Capture streaming Reddit posts, comments, and stock/crypto prices and store them for interactive querying
- **Technologies:** BigQuery Storage API, Protobuf, Python (yfinance, PRAW, pandas, argparse, dotenv), Docker

[Chicago Crime Data Warehouse and Business Intelligence Suite](#)

- **Task:** Create a Data Warehouse and a Business Intelligence application to provide insight on geolocation, predictors, and leading indicators of crime in Chicago
- **Technologies:** Google BigQuery, SQL, ETL pipeline creation in Python, Tableau