James Sandy

Machine Learning & Data Engineer

✉ mondaysandy3@gmail.com | 🌐 https://jamessandy.github.io/ | 🔗 LinkedIn | 💻 GitHub

GCP-certified Machine Learning Engineer with over six years of experience architecting scalable MLOps infrastructure and end-to-end data solutions. Proven ability to reduce data processing costs, enhance system performance, and drive user engagement through production-grade machine learning. My background in Philosophy provides a unique foundation for my work in AI, emphasizing logical frameworks and ethical considerations in model development and AI Safety research.

Technical Proficiencies

- **Cloud & MLOps:** Google Cloud Platform (GCP), AWS, Azure, Vertex AI, MLOps, CI/CD, Docker, Kubernetes, Terraform

- **Data Engineering & ETL:** Apache Airflow, Kafka, Spark, dbt, Snowflake, BigQuery, Redshift, Data Warehousing, Databricks

- **ML Frameworks & Libraries:** TensorFlow, PyTorch, Keras, Scikit-learn, spaCy, Weights & Biases

- **Programming & Databases:** Python, SQL, Scala, MySQL, PostgreSQL

- **Analytics & Visualization:** Tableau, Looker, Grafana, Power BI

Professional Experience

Machine Learning & Data Consultant | Remote | Sept 2023 – Present

Provided long-term, embedded consulting for Series A fintech and healthcare startups, acting as the lead engineer for their data infrastructure.

- Architected and deployed robust, cloud-native data pipelines that achieved 99.9% data availability for mission-critical fintech and healthcare applications.

- Engineered a high-throughput streaming pipeline using Kafka and GCP Pub/Sub, slashing data processing latency by 45% and enabling real-time fraud detection.

- Optimized ETL workflows with Airflow and dbt, reducing data processing costs by 10% through improved efficiency and resource management.

- Designed and implemented a scalable MLOps infrastructure on GCP (Vertex AI), reducing the model deployment lifecycle from weeks to days.

Data Engineer | Coinstal | April 2021 – Sept 2023

- Engineered and managed a centralized data lake on GCP, integrating over 20 disparate data sources using Airbyte, which became the single source of truth for all analytics.

- Automated 15+ key business intelligence dashboards using dbt and Power BI, eliminating 30 hours of manual reporting effort per week (a 70% reduction).

- Redesigned and optimized data models for the analytics warehouse, resulting in a 30% improvement in query performance for business intelligence tasks.

Machine Learning Engineer | Analog Teams | June 2020 – Feb 2021

- Led a research team in developing an AI-powered documentation tool; the resulting NLP models built with TensorFlow and spaCy automated the summarization of complex codebases.

- Improved documentation accuracy by 15% through rigorous A/B testing of different model architectures and hyperparameter tuning.

Machine Learning Engineer | Shareqube | April 2019 – May 2020

- Developed and deployed a personalized recommendation engine that increased user engagement on a crypto rewards platform by 25%.

- Engineered and maintained a fraud detection model that successfully identified and prevented over $500k in fraudulent crypto transactions in its first six months.

Certifications

- Google Cloud Professional Machine Learning Engineer

- Google Cloud Professional Data Engineer

Education

B.Sc. in Philosophy | University of Uyo | 2016 – 2021

Community & Publications

- **dbt Community Leader:** Founder of the Uyo dbt user group, organizing workshops on data and analytics engineering.

- **Technical Writing:** Portfolio