Karthik Adari

Dallas, TX | kvkadari@gmail.com | (469) 556-8866 | GitHub | Portfolio | LinkedIn

SKILLS

Data Engineering: Python, R, SQL, ETL/ELT, Snowflake, Cost Optimization, Data Pipelines, Data Modeling, Data Quality, Monitoring/SLAs, Airflow, Databricks, PySpark, Hadoop (HDFS), Hive
Cloud: GCP(BigQuery, Cloud Run, Cloud Functions, Compute Engine, Cloud Storage, Firestore, Cloud Logging, IAM), AWS (Redshift, S3, AWS Glue, Lambda)
Databases & Tools: MongoDB, MySQL, Docker, Git, CI/CD (GitHub Actions), FastAPI

EXPERIENCE

Founder | GCP Data Engineer & AI Developer   Jan 2025 – Present
Career in Fortune (AI-Powered Job Platform)   *Dallas, TX*
- Led end-to-end architecture of a GCP-native backend (Cloud Run, Firestore, BigQuery) to power job ingestion, search, and analytics workflows.
- Engineered 30+ production APIs (FastAPI, Cloud Functions) with secure auth and real-time NoSQL data management for platform features.
- Orchestrated automated ingestion of 100K+ job postings using Playwright + Compute Engine, sustaining 95%+ pipeline reliability.

Data Engineer | Machine Learning Specialist   May 2024 – Jan 2025
InnovatEV Global   *Dallas, TX*
- Built Airflow ETL pipelines to ingest/transform high-volume datasets; added Python data quality checks and managed MongoDB schema evolution.
- Reduced data issues by 30%+ and cut manual rework by 10+ hours/week through automated validation and pipeline hardening.
- Containerized services with Docker and deployed on GKE as REST APIs; automated CI/CD with GitHub Actions and added Cloud Logging, reducing release cycles by 60%.

Data Analyst   Jan 2021 – Dec 2021
Corizo   *Remote, India*
- Automated recurring analytics workflows with scripts, reducing manual effort by 300+ hours annually and improving team throughput.
- Delivered Tableau KPI dashboards and improved reporting accuracy by 20% through pipeline optimization and cross-functional documentation alignment.

PROJECTS

HDFS-Based ETL and Reporting on NYC Taxi Dataset   Jan 2025 – May 2025
Python, Hadoop HDFS, Hive
- Ingested and transformed 6.4M+ NYC Yellow Taxi records by converting Parquet to CSV, cleaning data, and loading into HDFS for scalable querying.
- Created Hive tables and executed analytical SQL to compare Jan 2024 vs Jan 2025 trip volume, fares, and speeds, surfacing seasonal patterns.

Credit Default Prediction (AmEx Dataset)   Jan 2024 – May 2024
Python, SQL, XGBoost, Neural Networks
- Curated and processed 5.5M+ records (190 features) with optimized SQL feature pipelines; achieved 96% accuracy on a sub-10% default-rate dataset.
- Optimized performance via GridSearch and leveraged XGBoost feature insights to refine a Neural Network, reducing false positives by 25%.

EDUCATION

The University of Texas at Dallas   Aug 2023 – May 2025
M.S. in Business Analytics & Artificial Intelligence   *GPA: 3.70/4.00*
Indian Institute of Technology Madras   Jan 2021 – May 2023
Diploma in Programming & Data Science   *GPA: 8.10/10.00*
Malaviya National Institute of Technology Jaipur   Aug 2019 – May 2023
B.Tech in Metallurgical & Materials Engineering   *GPA: 7.46/10.00*

CERTIFICATIONS

Google Cloud Professional Data Engineer
Salesforce Certified Associate