

Mutya Vaibhavi

Bay Area | <https://github.com/vaibhavimutya> | vaibhavii@gmail.com | <https://vaibhavimutya.github.io>

SUMMARY

Data Engineer with 3+ years building cloud-native data pipelines and analytics infrastructure across healthcare and HR-tech domains. Skilled in Azure (ADF, Databricks), dbt, Airflow, SQL data modeling, and Python ETL. Proven track record shipping pipelines ingesting 100k+ records, modeling dimensional warehouses, and powering BI/ML downstream systems.

SKILLS

Languages: Python (pandas, NumPy, PySpark), SQL, Java, Bash

Data Processing & Orchestration: Apache Airflow, dbt, Apache Spark/PySpark, Azure Data Factory

Cloud & DevOps: Azure (Databricks, ADF, Synapse), Docker, Git, CI/CD

Data Warehousing & Modeling: Snowflake, Star/Snowflake schemas, Dimensional Modeling, ETL/ELT

Databases: PostgreSQL, MySQL, SQL Server

BI & Visualization: Power BI, Tableau, DAX, Excel (PivotTables, VLOOKUP)

ML & NLP: scikit-learn, BERT, LangChain, RAG, spaCy

PROFESSIONAL EXPERIENCE

Epsoft Technologies Inc.

Data Engineer, Dallas, TX

June 2025 – December 2025

- Architected MySQL and PostgreSQL data models for candidate profiles, verification records, and recruiter logs; built stored procedures and aggregation layers reducing query response time by 40% across **Trustume's AI-driven platform**.
- Built event-driven ingestion pipelines and webhook services (Java, Spring Boot) feeding recruiter dashboards and AI matching models, reducing data latency by 35% and improving feature freshness for downstream ML scoring.
- Containerized and orchestrated data services on Azure using Docker, improving deployment consistency and cutting environment-related pipeline failures by 50% across nationwide recruitment infrastructure.
- Designed normalized dimensional data models for Trustume's ATS, enabling structured storage and feature extraction across candidate profiles, skill assessments, and verifications while reducing data redundancy by 30%.

Association of American Medical Colleges

Research Data Analyst, Washington, DC

July 2023 – March 2025

- Architected an Azure pipeline (ADF, Databricks, dbt) ingesting 100k+ NIH grant records; applied BERT NLP to resolve affiliations, surfacing 60% of funding directed to medical schools with a 2-4% YoY decline via Power BI.
- Migrated legacy Excel/VBA reporting to a modern Power BI layer built on dbt-modeled LCME survey data spanning 150+ medical schools, cutting benchmark report generation time by 80% and eliminating manual data prep.
- Engineered an end-to-end faculty compensation data pipeline (Python) integrating AAMC, LCME, and BLS sources; modeled regression analyses detecting salary inequities by gender and rank, served to finance leadership via Power BI dashboards.

UCR, Bourns College of Engineering

Software Engineer, Riverside, CA

February 2023 – May 2023

- Built Python pipelines (pandas, numpy, scikit-learn) to clean and process large-scale mass spectrometry datasets; executed GNPS molecular networking workflows for compound identification.

PROJECTS

E-Commerce Sales Dashboard | Python, dbt, Snowflake, Airflow, Power BI

Ingested an e-commerce dataset into Snowflake, modeled fact/dimension tables using dbt and data quality tests, scheduled daily refresh with Apache Airflow, and built a Power BI dashboard tracking revenue, customer segments, and return rates.

Job Market Analyzer | Python, HuggingFace, spaCy, Claude

Python pipeline that parses PDF resumes via Claude Sonnet and aggregates job postings from multiple APIs, using HuggingFace NER + spaCy to extract skills/seniority/visa requirements, then ranks job markets by a heuristic fit score.

EDUCATION

Master of Science in Artificial Intelligence

University of the Cumberland

Expected August 2027

Master of Science in Computer Science

University of California, Riverside

September 2021 – December 2022

Bachelor of Technology in Computer Science Engineering

GITAM University, Hyderabad

June 2016 – July 2020