Vaishnavi Mocherla

San Jose, CA | +1 (530) 407-3384 | vmocherla25@gmail.com | [LinkedIn](#) | [Github](#)

EDUCATION

San Jose State University — San Jose, CA
M.S. in Applied Data Science — *Aug 2023 – May 2025*

EXPERIENCE

Informatica — California, USA
Data Analyst – AI & Cloud Systems — *May 2024 – Dec 2024*

- Developed a **AI-powered analytics agent** using **Azure OpenAI** and **Python** for NL-to-SQL translation, reducing data query turnaround by **60%** for analysts and increasing SQL precision by **20%**.
- Built and maintained **ETL pipelines** using **Informatica IICS** to ingest structured/unstructured data; achieved **99.9% data accuracy** through robust **IDQ validations** and schema harmonization.
- Optimized batch and real-time data processing using **PySpark** and **SQL** in **Databricks**, improving data pipeline throughput and reducing job runtimes by **35%**.
- Executed secure migration of **1 TB+ enterprise data** to **Azure Data Lake** using **Informatica CDAM**; implemented **CI/CD pipelines**, ensuring governance compliance and reducing manual deployment time by **50%**.
- Configured and managed **EMR Clusters** to run **Spark jobs** on **AWS**, optimizing big data workloads with **30% faster** job completion and reduced compute costs via **spot instance tuning**.
- Designed and delivered **Power BI dashboards** with actionable insights on campaign reach, spend, and conversion rates; reduced manual reporting time by **70%** and enabled real-time KPI tracking for marketing and operations teams.

Kanerika Software Pvt Ltd — Hyderabad, India
Data Scientist & ML Research — *Aug 2022– Jun 2023*

- Conducted exploratory data analysis on **100K+ resumes** and job descriptions; built a **BERT-based system** to predict candidate progression up to L3, reducing manual screening effort by **60%** and filtering out **35–40%** low-fit profiles early.
- Designed **Power BI dashboards** to visualize fit scores, model confidence, and hiring pipeline conversion, improving recruiter decision-making and reducing screening turnaround time by **30%**.
- Built a **Python**-based parser to process 10+ resume formats (PDF, DOCX, scanned text), improving data extraction reliability by **70%** and streamlining ingestion for downstream analytics.
- Integrated the parsing and prediction modules into a real-time ETL workflow using **Airflow** and **Spark**; ensured consistent data flow and timely insights for recruiting teams.
- Presented prototype to **stakeholders**, securing **beta approval** and roadmap inclusion for ATS integration.

PROJECTS

Marketing Mix Modeling & Multi-Channel Attribution Platform | *SQL, Snowflake, Tableau, MLflow, Airflow, Scikit-learn*

- Architected **marketing attribution system** analyzing **$500K+ annual ad spend** across 8 digital channels using **multi-touch attribution modeling**, **Bayesian regression**, and **Adstock transformation**.
- Engineered **real-time attribution algorithms** processing **2M+ daily touchpoints** with **Shapley value attribution** and **Markov chain modeling**; integrated **MLflow** for automated A/B testing.
- Delivered **35% ROAS improvement** and **28% CAC reduction** while establishing **executive attribution dashboard** in **Tableau** for strategic decision-making.

Financial Analytics & Automated Anomaly Detection System | *SQL, Snowflake, Power BI, Airflow, Scikit-learn, Prophet*

- Engineered **financial analytics platform** processing **10M+ transaction records** using **isolation forests** and **Prophet models** for anomaly detection and forecasting.
- Identified **$50K revenue leakage** within first month and improved **forecast accuracy by 22%** through automated controls and real-time P&L monitoring.

TECHNICAL SKILLS

Languages: Python, SQL, R, Excel

Statistical Analysis: A/B Testing, Regression Analysis, Time Series Forecasting, Hypothesis Testing, Statistical Modeling

Data Processing & Analytics: Pandas, NumPy, PySpark, Scikit-learn, TensorFlow

BI & Visualization: Power BI, Tableau, Looker, Matplotlib, Seaborn

Data Engineering: Spark, Kafka, Databricks, Snowflake, Hadoop, BigQuery

ETL & Data Workflow: Informatica (IICS, IDQ, CDAM), Airflow, Azure Data Factory, CI/CD

Databases: PostgreSQL, MySQL, MongoDB, SQL Server, HDFS

Cloud & Analytics: Azure, AWS, GCP