SAHITHI MADAS

Phoenix, AZ | +1-518-599-8103 | madas.sahithi@gmail.com | linkedin.com/in/sahithi-madas/ | https://sahithim.netlify.app/

SUMMARY

Data Analyst with 2+ years of experience working at the nexus of data systems and informatics. Specialized in data visualization, reporting, and improving KPI tracking through cloud-based pipelines, interactive dashboards, and AI-powered compliance tools. Adept with AWS, Python, R, Tableau, Power BI, SQL, Excel, and LLMs, with a strong foundation in KPIs, data governance, and secure data infrastructure.

EDUCATION

University at Albany, State University of New York | *Master's, Data Science*   **Aug 2023 - May 2025**
 • **Award & Achievements :** Dean's Merit Scholarship | Presenter at UAlbany Showcase 2025

Andhra University College of Engineering | *Bachelor's, Computer Science*   **Aug 2019 - May 2023**

PROFESSIONAL EXPERIENCE

Albany County Local Government - Division of Information Services   **Sep 2024 - May 2025**
Data Engineer Intern, Graduate Fellow   *Albany, NY*
 • Designed and implemented a secure, HIPAA-compliant data lake solution on **AWS (S3, Glue, Lambda, Redshift, CloudWatch)** to centralize analytical infrastructure across departments reducing latency of inter-agency (ACDOH, NYSDOH, NYCDOH) data sharing by 60% and developing **Python-based ETL workflows** that included real-time monitoring hooks using AWS CloudWatch and integrated CI/CD.
 • Audited **IAM access** policies and monitored encryption and lifecycle events with AWS CloudTrail, improving system-wide compliance.
 • Simulated real-time health alert ingestion using **Kafka streams** from EMS and hospital reports with **SQL** transformations applied to normalize incoming datasets before downstream analytics.
 • Managed data cataloging and lineage tracing using **AWS Glue** and Lake Formation, improving traceability of sensitive health data.
 • Conducted compliance mapping against ISO/IEC 27001, HIPAA/HITECH, NIST 800-53 frameworks by building interactive **Tableau dashboards** using **SQL** data querying for 30+ policy documents showing clause-to-control mapping, audit status, and implementation lifecycle.
 • Performed **actual vs expected data reconciliation** across IT policies to support **internal compliance tracking** and stakeholder reporting.

New York State Department of Health - Department of Environmental Protection   **May 2024 - Aug 2024**
Data Analyst Intern   *Glens Falls, NY*
 • Designed 15+ dynamic **Tableau dashboards** and integrated **SQL** data queries to visualize environmental health KPIs, including air quality index, water pollution levels, and risk zones across 60+ counties, enabling data-driven policy planning and accelerating public health reporting by 40%.
 • Automated **data ingestion** and transformation pipelines using **Python and SQL**, cleansing over 10,000 public health records and eliminating 1200+ inconsistencies and implemented **anomaly detection** using **Python's scikit-learn** to flag non-compliance patterns.
 • Standardized environmental codes and identified trend outliers using **Microsoft Excel** tools like conditional formatting, **pivot tables, Power Query, and DAX formulas**, which enhanced data clarity and facilitated better decision-making.

University at Albany Libraries - Digital Scholarship Lab   **Feb 2024 - May 2025**
Graduate Student Assistant   *Albany, NY*
 • Developed interactive dashboards using **Power BI (PowerQuery)** and **Tableau** to analyze digital resource usage, enhancing research initiatives.
 • Automated **Python/R** workflows and **SQL** for **statistical analysis** and visual reporting, streamlining the research data lifecycle.
 • Advised over 100 researchers and students on **data visualization strategy** and insight extraction, enhancing data literacy.

PROJECT

ComplyAI
 • Develped an **AI-powered compliance automation tool** to analyze Organization's policies against major information security frameworks and standards such as HIPAA, ISO, NIST, etc. - reducing manual audit time by over 80%.
 • Built **Gap Analysis reports** and a **chatbot** using **LLM framework (LangChain & Ollama)** that comparing policies to 150+ controls.
 • Integrated a RAG pipeline, enabling real-time, file-specific chatbots with 70% improved comprehension with document ingestion and segmentation system using **PyMuPDF + LangChain** loaders to pre-process multi-page PDFs with versioning support.
 • Benchmarked chatbot performance across GPT-3.5 & Ollama using a custom latency-monitoring suite to track response time across CPU/GPU.
 • Enabled document-level audit traceability, version control tagging, and metadata tagging (e.g., policy type, compliance scope).

Life Expectancy Simulation via Thresholding Robust Regression
 • Applied custom thresholding techniques to **detect and neutralize inconsistent records**, improving the model's capacity to deliver reliable life expectancy forecasts across 90+ countries with varied socio-economic variables and conditions.
 • Improved forecast accuracy by 28% over standard L1 and L2 baselines using **scikit-learn's Ridge, Lasso**, and custom **PyTorch** models.
 • Applied **PCA** for dimensionality reduction, **KNN** imputation (null handling), and polynomial feature expansion to boost model performance.

AutoPrice Predictor
 • Built a predictive analytics tool for used cars' resale prices with a high-precision **Regression model**, achieving 79.20% test accuracy.
 • Engineered end-to-end data pipeline with **PySpark, Pandas, PyTorch, TensorFlow, scikit-learn, XGBoost pipelines, Matplotlib, NumPy**, and advanced encoding **techniques for cleaning**, feature transformation, **handling outliers/missing values** across 10K+ cars.
 • Implemented logging system using **Python Logging + Prometheus** for model performance and refresh tracking.

SKILLS

 • **Programming Languages**: Python, Java, R, C++, C#, Go, HTML & CSS, JavaScript, Node.js, React, Angular, D3.js, Flask
 • **Databases**: PostgreSQL, MySQL, NoSQL (MongoDB), Oracle, AWS RDS, AWS S3, AWS Lambda, Azure Blob, GCP
 • **Data Analysis & Visualization Tools**: Tableau, Microsoft Excel (Power Query, Pivot, VBA), Power BI, Seaborn, SAS, STATA
 • **Software Tools and IDE**: Oracle Cloud, Azure, AWS (Glue, RDS, S3, Redshift, CloudWatch, IAM, Lambda), Kubernetes, Docker, Linux, Kafka, Helm, API, Databricks, Computer Vision, Airflow, TensorFlow
 • **Monitoring & CI/CD**: DataDog, Splunk, Databricks, Prometheus, GitHub Action, CircleCI, TravisCI, DevOps, Pytest, Junit
 • **Soft Skills**: Multi-tasking, Cross-Functional Teams, Strong Problem-Solving Skills, Documentation, Compliance Mapping

CERTIFICATES

 • Google Data Analytics Professional Certificate | Google   • AWS Serverless Assessment Badge | AWS