Thanmaya Sri Sigireddi

Data Analyst

✉ sigireddithanmayasri@gmail.com   📞 +1 (716) 465-9667   in LinkedIn   ○ Github   🔗 Portfolio

PROFILE

Data Analyst with strong analytical and statistical expertise, specializing in integrating, transforming, and analyzing large-scale datasets using Python, SQL, AWS, and Snowflake. Experienced in automating ETL pipelines and developing BI dashboards with Power BI and Tableau to drive data-informed decisions. Proficient in exploratory data analysis (EDA), data visualization, data storytelling, and data science. Skilled in time series forecasting and applying machine learning frameworks (Matplotlib, Scikit-learn, TensorFlow) and models (Random Forest, XGBoost, Large Language Models (LLMS)) for classification, clustering, and regression to deliver actionable insights through predictive modeling. Experienced in data warehousing and engineering to build efficient, robust business solutions.

EDUCATION

State University of New York at Buffalo, *Master of Science in Data Science*   Jan 2024 – May 2025 | Buffalo, NY

VNR VJIET, *Bachelor of Technology in Electrical and Electronics Engineering*   Jun 2019 – May 2023 | Hyderabad, India

PROFESSIONAL EXPERIENCE

Nucleus Business Solutions, *Data Analyst (Intern and Full time)*   Oct 2022 – Dec 2023 | Hyderabad, India
- Engineered and automated end-to-end ETL pipelines from SQL Server to Snowflake, reducing manual reporting time by 45 percent, eliminating 12 plus hours per week of manual processing, and enabling real-time analytics for 20 plus stakeholders.
- Developed a robust K-Means clustering model to segment users by browsing patterns, improving marketing targeting accuracy, personalizing campaigns effectively, and increasing user engagement by 18 percent over a detailed three-month evaluation period.
- Scraped competitor pricing data from 50 plus sites using Python (BeautifulSoup, Requests) and integrated with internal SKU/ margin data, enabling pricing optimization that increased profits by 7 percent across high-revenue product categories.
- Built and deployed Power BI dashboards with advanced DAX, drill-throughs, and custom filters. Adopted by 15 plus teams, these dashboards enabled real-time KPI tracking, streamlined workflows, and reduced reliance on manual Excel reports.
- Conducted comprehensive A/B testing on recommendation logic, boosting click-through rates by 12 percent and driving a 6 percent increase in repeat purchases, improving overall customer retention and loyalty metrics significantly.
- Configured AWS IAM and Snowflake RBAC to securely protect sensitive datasets with 100 percent compliance, and streamlined recurring report generation via Excel VBA macros, cutting delivery time from 2 days to under 4 hours consistently.

PROJECTS

Yelp Reviews: Real-Time Data Pipeline and Analytics,
Cloud-Based ELT Workflow, Python-Driven Data Ingestion, and Scalable Insights using Snowflake and AWS S3
- Delivered an end to end real time data pipeline by ingesting 1M plus semi structured Yelp reviews (JSON) into AWS S3 using Python with schema validation, intelligent batching, and robust error handling to ensure data quality and reliability.
- Automated continuous data loading into Snowflake via Snowpipe, leveraging staging tables, VARIANT columns, and optimized views to parse, transform, and enrich raw JSON with business metadata for downstream analytics.
- Designed 3 Snowflake-integrated Power BI dashboards with user-driven slicers, hierarchical drill-downs, and optimized DAX. Enabled near real-time insights under 2-minute latency, helping stakeholders monitor performance with speed and accuracy.
- Implemented centralized monitoring with AWS CloudWatch Logs and Python-based custom logging, capturing 100 percent batch uploads, invalid record detection, and pipeline errors to support fast issue resolution and performance optimization.

HealthReadmit: Patient Readmission Prediction Engine,
End-to-End Predictive Data Science Project with SQL, Python, Power BI, and Streamlit
- Constructed a patient readmission prediction pipeline using SQL Server and Python (Pandas, NumPy, Scikit-learn, XGBoost) for feature engineering and XGBoost Classifier with grid-search tuning, improving model ROC-AUC from 0.78 to 0.86.
- Created a Power BI dashboard using DAX for 10 plus patient demographics, and clinical attributesincorporating slicers and drill through functionality for stakeholder insights in healthcare, cutting manual report creation by 50 percent.
- Deployed a Streamlit-based web application for real-time readmission risk prediction, leveraging a serialized XGBoost model (via Joblib) with a RESTful API-like interface, reducing assessment time from 15 minutes to under 1 minute per patient.
- Performed exploratory data analysis (EDA) and feature engineering, deriving features (like interaction, encodings) with Git/Github workflows to identify 5 high-impact clinical features, boosting model precision by 14 percent.

SKILLS

- **Programming & Scripting:** Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Web Scraping), R, VBA for Excel
- **Databases:** My SQL, PostgreSQL, SQL Server
- **ETL & Warehouses:** Snowflake, Amazon Redshift
- **Data Visualization and Data Reporting:** Power BI (DAX, RLS), Tableau, Excel, Amazon QuickSight
- **Machine Learning:** Regression, Classification, Clustering, Decision Trees, Neural Networks, Time Series
- **Tools:** GitHub, MS Excel (Advanced Functions, Macros, Pivot Tables), MS PowerPoint, MATLAB
- **Data Governance, Security & Compliance:** Role-Based Access Control (RBAC), AWS IAM, Data Lineage & Governance

CERTIFICATES

- Amazon Web Services: AWS Certified Cloud Practitioner ✎

- Google: Data Analytics Professional Certificate ✎