# Thanmaya Sri Sigireddi

*Data Analyst*

sigireddithanmayasri@gmail.com +1 (716) 465-9667 Texas 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 (Scikit-learn, TensorFlow) and models (Random Forest, XGBoost, Large Language Models) 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 | Hyderbad, India

- Engineered end-to-end ETL pipelines to process live sales, inventory, and user interaction data from SQL Server into Snowflake. Automated data integration workflows to support real-time reporting and reduce dependencies on manual processes.
- Formulated K-Means clustering model using Scikit-learn to segment users based on session-level browsing patterns such as product views, and cart activity. These insights were simplified to drive marketing strategies and improve customer engagement.
- Scraped real-time competitor pricing data using Python (BeautifulSoup, Requests) and integrated it with internal SKU and margin data. The resulting dataset was used by business teams to monitor market trends and optimize pricing decisions across categories.
- Designed Power BI dashboards with DAX measures, dynamic filters, and drill-down capabilities to visualize performance metrics for sales, product mix, and customer segments. These dashboards were used daily by teams for decision-making.
- Conducted A/B testing on product recommendation logic, comparing targeting approaches based on customer clusters. Analyzed test outcomes across user cohorts and presented insights to stakeholders through visual reports and business-friendly summaries.
- Implemented enterprise-grade data governance by configuring AWS IAM and role-based access control (RBAC) in Snowflake. Ensured secure, role-specific data access and compliance with internal client policies across departments and geographies.
- Streamlined recurring report generation and delivery using Excel VBA macros and integrated scheduling logic. This reduced operational bottlenecks, standardized output formats, and allowed analysts to focus on strategic analysis instead of repetitive tasks.

## PROJECTS

**Yelp Reviews: Sentiment & Data Analysis Pipeline,**

*Cloud-Based ELT workflow, Python-Driven Sentiment Analysis, and Scalable Insights using Snowflake and AWS S3*

- Initiated a Python-based data ingestion pipeline to preprocess and split a 5GB deeply nested Yelp JSON dataset into smaller chunks, enabling parallelized and upload to AWS S3 for scalable cloud storage mproving ingestion efficiency by 3 times.
- Leveraged AWS S3 as a durable and highly available data lake to store both raw and intermediate datasets, seamlessly integrating with Snowflake external stage for automated, high throughout data ingestion and subsequent warehousing.
- Developed Snowflake ELT workflows leveraging the variant data type for semi-structured data storage and lateral flatten functions to normalize complex nested JSON structures into relational tables optimized for data modeling and analytical querying.
- Authored Python User-Defined Functions (UDFs) within Snowflake to perform sentiment analysis using TextBlob, and authored complex multi-step SQL queries to derive business insights, sentiment trends, and review behavior patterns.

**HealthReadmit: Patient Readmission Prediction Engine,**

*End-to-End Predictive Data Science Project with SQL, Python, Power BI, and Streamlit*

- Constructed a data science pipeline for patient readmission prediction using SQL Server Management Studio and Python (Pandas, NumPy, Scikit-learn, XGBoost) for feature engineering and XGBoost Classifier with grid-search tuning.
- Created a Power BI dashboard using DAX for patient demographics, clinical attributes, and model performance metrics (ROC-AUC, precision-recall), incorporating slicers and drill-through functionality for stakeholder insights in healthcare.
- 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, ensuring scalable and user-friendly delivery of probabilistic outputs.
- Performed exploratory data analysis (EDA) and feature engineering, deriving features (e.g., interaction, encodings) to boost XGBoost model performance, with Git/GitHub workflows ensuring version control and reproducibility.

## SKILLS

- **Programming & Scripting:** Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Web Scraping), R, VBA for Excel
- **Databases:** PostgreSQL, SQL Server
- **ETL & Warehouses:** Snowflake, Amazon Redshift
- **Data Visualization and Data Reporting:** Tableau, Power BI (DAX, RLS), 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