Thanmaya Sri Sigireddi

Data Analyst

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

PROFILE

Data Analyst with hands on experience integrating, transforming, and analyzing large scale datasets using Python, SQL, AWS, and Snowflake. Proficient in automating ETL pipelines and building BI dashboards using Power BI and Tableau. Skilled in EDA, time series forecasting, and building ML models (Random Forest, XGBoost, LLMs) using classification, clustering, and regression for predictive and prescriptive analytics. Experienced in building KPI dashboards, DAX calculations, drill-down reports, and delivering data-driven strategies.

PROFESSIONAL EXPERIENCE

Nucleus Business Solutions, *Data Analyst*   May 2023 – Dec 2023 | Hyderbad, India
- Created Power BI and Tableau dashboards to track recruitment KPIs, improving visibility and cutting manual reporting by 50%.
- Automated ETL pipelines with PySpark and SQL to unify HR data, improving availability and boosting report accuracy by 30%.
- Modeled attrition risk using Scikit-learn to support hiring strategy and reduce recruitment costs by 20%.
- Optimized DAX queries and streamlined Power BI models, reducing report load time by 35%.

Nucleus Business Solutions, *Data Analyst Intern*   Oct 2022 – Apr 2023 | Hyderabad, India
- Designed Power BI dashboards on 10+ KPIs, reducing offer rejection by 12%.
- Streamlined data retrieval with custom SQL and Python scripts for recurring reports.
- Evaluated hiring trends via ad hoc reports, improving sourcing success by 10%.

EDUCATION

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

VNR Vignana Jyothi Institute of Technology and Engineering,   Jun 2019 – May 2023 | Hyderabad, India
Bachelor of Technology in Electrical and Electronics Engineering

Courses: Mathematics, Statistics, Data Modeling, Deep Learning, LLM, Data Structures and Algorithms, Data Analytics, Data Mining, Predictive Analytics, Business Intelligence and Business Analytics

SKILLS

- **Programming & Scripting:** Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), R, SQL, VBA for Excel, Bash, PowerShell
- **Databases:** MySQL, PostgreSQL
- **ETL & Warehouses:** Snowflake, Amazon Redshift, Apache Hadoop, Apache Spark, Hive
- **Data Visualization and Data Reporting:** Tableau, Power BI (DAX, RLS), Excel Pivot Tables, Charting & Dashboard Design, 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 Practices

PROJECTS

Detox Dial: AI-Powered Digital Wellness App,
Personalized Behavioral Nudging, NLP-based Profiling, and Real-Time User Engagement
- Engineered an NLP-based nudging system that improved trial retention by 30%.
- Deployed Flask APIs for real-time user responses, triggering interventions that addressed key churn drivers during trials.
- Refined personalization scoring logic and backend APIs to reduce trial-phase churn by 25% and enable real-time engagement delivery.

SpecTrack: Laptop Specification Analytics Engine,
Relational Database Modeling, Query Optimization, and Consumer Insights
- Structured a normalized PostgreSQL database for 5,000+ laptops with 20+ features.

- Cleaned and standardized data with Python and SQL, improving integrity by 85%.
- Coded 25+ SQL queries for analytics on price trends, feature filters, and popularity metrics. •
- Boosted query efficiency by 70% via indexing strategies and filter-based tuning.

CryptoCluster: Unsupervised Learning for Market Segmentation,
Time-Series Feature Engineering, K-Means Clustering, and Dimensionality Reduction
- Developed a K-Means model to segment crypto assets by time-series behavior.
- Extracted trend-based features from time-series crypto data to improve clustering reliability and reduce noise.
- Reduced dimensionality by 50% using PCA while retaining 90% of variance.
- Boosted silhouette score by 40% through a preprocessing and denoising pipeline.

CERTIFICATES

- Data Science internship - Plasmid ✎
- Python - Excel Solutions ✎
- AI-Integrated Application - AI for Good Hackathon ✎

- SQL - Excel Solutions ✎
- Microsoft Power BI - Excel Solutions ✎
- Data Science Foundations - Great Learning Academy ✎