

Sahithi Vangala

Data Analyst | +1 (812) 650-8866 | sahithivangala.1808@gmail.com | LinkedIn | GitHub | Open to Relocation

EDUCATION

Master of Science in Data Science Indiana University Bloomington (GPA: 3.7/4) <ul style="list-style-type: none">Coursework: Machine Learning, Statistics, Data Mining, Cloud Computing(AWS)	August 2023 - May 2025 Indiana, USA
Bachelor of Technology in ECE Jawaharlal Nehru Technological University (GPA: 3.7/4)	August 2019 - May 2023 Hyderabad, India

EXPERIENCE

Data Analyst Project 990 Inc. <ul style="list-style-type: none">Architected an ETL pipeline using SQL, Python, processing 2.5M+ records, improving data throughput by 30%, ensuring data integrity and enabling stakeholders to track KPIs.Designed Tableau dashboards that highlighted key KPIs, trends, and funding disparities across 50 states, enabling leadership to make data-driven funding decisions.Performed segmentation analysis of grant distribution, uncovering disparities and supporting targeted funding recommendations.	May 2024 - July 2025 Bloomington, USA
Research Data Analyst Indiana University <ul style="list-style-type: none">Conducted EDA on USDA datasets using Python (Pandas, NumPy), identifying critical data quality issues and coverage gaps across agricultural indicators.Crafted comprehensive Power BI dashboards using DAX calculations, visualizing 20+ years of crop and economic time-series data, providing stakeholders with actionable agricultural insights.Partnered with product and engineering teams to align dashboard design and KPI tracking with business requirements, improving feature optimization.	August 2024 - December 2024 Bloomington, USA
Marketing Data Analyst EFORGE NexGen INNOVATIONS Pvt. Ltd. <ul style="list-style-type: none">Analyzed campaign and customer data using Excel (PivotTables, Power Query) to identify behavioral trends, increasing acquisition by 15%.Mapped customer journeys and identified drop-off points across conversion funnels, leading to a 12% uplift in lead-to-customer conversion.	April 2022 - April 2023 Hyderabad, India

PROJECTS

Customer Segmentation Pipeline <i>Snowflake, dbt, Clustering</i> <ul style="list-style-type: none">Built Recency-Frequency-Monetary (RFM) metrics by transforming 20K+ customer transactions in Snowflake using dbt models, standardizing business logic and ensuring data reliability.Implemented KMeans clustering to segment customers into cohorts, supporting marketing strategies projected to increase engagement by 12%.Formulated a Tableau dashboard visualizing RFM segments and campaign impact, improving stakeholder insight into 5K+ customer records.
E-Commerce Conversion Analysis Using A/B Testing <i>A/B Testing, Python(Pandas, SciPy), Looker</i> <ul style="list-style-type: none">Produced A/B test results with actionable insights to product and engineering teams, influencing UI changes that improved user engagement by 12%.Executed funnel analysis on 25K+ sessions to pinpoint drop-offs, recommending journey optimizations that improved engagement metrics by 10%.Built a cross-functional Looker dashboard tracking conversion KPIs across A/B variants, used by 3+ teams to monitor performance and make data-backed decisions.
Emergency Room Patient Visit Analytics <i>PowerBI, SQL(CTEs, Window Functions, Stored Procedures)</i> <ul style="list-style-type: none">Crafted interactive Power BI dashboard using DAX and advanced SQL queries (CTEs, window functions) to analyze 4,600+ ER visits, tracking patient flow KPIs and operational metrics.Implemented SQL stored procedures for real-time patient flow analysis, reducing manual reporting by 30%.Automated reporting workflows with SQL triggers and data validation, enabling real-time data-driven decisions for healthcare operations.
Stock Market Real-Time Data Analysis <i>AWS (S3,Glue,Redshift), Apache Spark, Kafka, Tableau</i> <ul style="list-style-type: none">Devised real-time streaming pipeline using Apache Kafka to ingest and process 1M+ daily stock messages, implementing automated alerting and data validation.Deployed scalable AWS infrastructure (S3, Glue, Redshift) with Apache Spark for batch processing and Tableau for analytics visualization, reducing overall data latency by 8ms.

SKILLS

Programming & Data Analysis: Python (Pandas, spaCy, Plotly), R, SQL, DAX, Git.
Data Visualization & Reporting: Power BI, Tableau, Looker, Google Analytics, Ad Hoc Analysis.
Statistical & ML Techniques: A/B Testing, Hypothesis Testing, Regression, Classification, Clustering, Time-Series Forecasting, Feature Engineering, Predictive Modeling, Anomaly Detection.
Data Engineering: Azure Data Factory, MySQL, PostgreSQL, BigQuery, Snowflake, Databricks, dbt, Apache Airflow, Apache Spark, Kafka, ETL Pipelines, Alteryx, SSIS, T-SQL.
Cloud Platforms: AWS (S3, Glue, Athena, Redshift), GCP, Azure.
Business Tools & Product Analytics: Excel(PivotTables, Power Query), Access, Salesforce, Amplitude, Mixpanel, Power Automate, Jira, Confluence, Agile (Scrum/Kanban), ERP Systems (SAP, Oracle), SDLC.

CERTIFICATIONS

Microsoft Power BI Data Analyst Associate [Link]
Business Analysis Foundations LinkedIn [Link]
AWS Cloud Practitioner [Link]