Sahithi Vangala

Bloomington, Indiana

📞 812-650-8866 | ✉ svangal@iu.edu | 🔗 linkedin.com/in/sahithi-vangala/ | ⭘ github.com/sahithi-1881

SUMMARY

Data Analyst with 2 years of experience in SQL, Python, Power BI, and AWS, specializing in data pipelines, dashboards, and process optimization. Proven ability to translate complex data into actionable insights that support business decisions across healthcare, marketing, and nonprofit sectors.

EXPERIENCE

Data Analyst — May 2024 - Present
Project 990 — *Bloomington, USA*
- Engineered an ETL pipeline with SQL & Python, processing 2.5M+ records, improving data throughput by 30%.
- Designed Tableau geographic heat maps for healthcare philanthropy, identifying 50+ under-served regions for resource allocation.
- Devised NLP features using spaCy to analyze 160,000 non-profit mission statements, extracting race and ethnicity insights for diversity initiatives.
- Created a Tableau dashboard to visualize education philanthropy deserts in 50+ regions, facilitating targeted educational funding to improve graduation rates.

Research Assistant — August 2024 - December 2024
Indiana University — *Bloomington, USA*
- Developed interactive Power BI dashboards visualizing 20+ years of crop and economic data for local farmers, improving actionable insights and operational efficiency by 25%.
- Implemented NLP-based news feeds using OpenAI's GPT-4, summarizing real-time agricultural updates from 10+ sources, boosting user engagement by 30%.
- Curated a centralized data catalog with 100+ metadata entries, enhancing data accessibility and usability for diverse stakeholders across agricultural communities.

Marketing Data Analyst — October 2022 - April 2023
EFORGE NexGen INNOVATIONS Pvt. Ltd. — *Hyderabad, India*
- Analyzed market data in Excel to identify key consumer trends, driving a 15% increase in customer acquisition.
- Designed Power BI dashboards to track 20+ marketing KPIs, optimizing resource allocation and boosting campaign ROI by 25%.

PROJECTS

Stock Market Real-Time Data Analysis Using Kafka
- Streamlined real-time analysis of over 1 million daily stock market messages using Apache Kafka, enhancing market decision-making speed.
- Optimized data workflows with AWS S3, Athena, and Glue, reducing processing times by 25%.
- Enhanced data simulation and ingestion accuracy using Python and Boto3 in high-stakes trading scenarios.

Emergency Room Patient Visit Analytics
- Deployed an interactive Power BI dashboard to analyze trends from 4,600+ ER visits, improving healthcare response strategies.
- Constructed interactive dashboards leveraging DAX to monitor key performance indicators for ER visits improving data accuracy by 15%.
- Optimized reporting processes by 30%, enabling data-driven decision-making for healthcare administrators.

CarePal: AI-Driven Health and Wellness Chatbot
- Developed and deployed CarePal, an AI-driven health chatbot leveraging Llama 2 and Pinecone, achieving 90% user satisfaction.
- Indexed 10K+ WHO/CDC health guidelines using ML pipelines, improving response accuracy by 35%.

Phishing Website Detection Using Machine Learning Algorithms
- Established an XGBoost-based AI system with a 94% accuracy rate in detecting phishing websites.
- Analyzed and identified key phishing patterns from 10,000+ sites, enhancing cybersecurity measures.

EDUCATION

Indiana University — August 2023 - May 2025
Master of Science in Data Science (GPA : 3.7/4) — *Bloomington, USA*

Jawaharlal Nehru Technological University — August 2019 - May 2023
Bachelor of Technology in ECE (GPA : 3.7/4) — *Hyderabad, India*

SKILLS

Programming Languages & Tools: Python (Pandas, NumPy, Scikit-learn, Matplotlib), R, SQL, DBT, Power BI, Tableau, Looker, CRM(Salesforce), Microsoft Office Suite (Word, Excel, PowerPoint, Outlook, Teams).
Machine Learning & Data Science Techniques: A/B Testing, Statistical Modeling, Hypothesis Testing, Regression Analysis, LLM Fine-tuning, Time-Series Analysis, ETL Pipelines, Feature Engineering.
Databases & Cloud Computing: MySQL, PostgreSQL, BigQuery, Snowflake, AWS (S3, RDS), GCP.
Data Engineering: ETL pipelines, Apache Hadoop, Apache Spark, Apache Airflow.
Project Management Tools: Agile (Scrum, Kanban), JIRA, Confluence, GitHub, CI/CD pipelines.

CERTIFICATIONS

Microsoft Power BI Data Analyst Associate [Link]
PwC Switzerland Power BI Job Simulation (Forage) [Link]
Business Analysis Foundations LinkedIn [Link]
AWS Cloud Practitioner [Link]

LEADERSHIP

- **Marketing Head, IETE CVR** – Led a team to organize technical workshops and the largest annual technical fest, driving measurable improvements in student engagement and event participation.