

SAI CHARAN REDDY GANKIDI

[LinkedIn](#) | [Github](#) | [Portfolio](#) | saicharangankidi17@gmail.com | +353 894522666 | Dublin, Ireland

PROFILE SUMMARY

Data Analyst with 2 years turning large-scale data into business action: built Python + SQL ELT pipelines (Airflow, dbt) that ingest 10M-row/day logs from AWS S3 into Snowflake/Redshift and produced Power BI & Tableau dashboards that cut month-end reporting time 40% and improved forecast accuracy 15%. Skilled in Excel, statistical testing, and stakeholder storytelling.

EDUCATION

National College of Ireland

Master of Science in Artificial Intelligence; Percentage: 90%

Dublin, Ireland

Jan 2024 – Jan 2025

GITAM University

Bachelor of Technology in Computer Science Engineering; Percentage: 80%

Bengaluru, India

July 2018 – July 2022

SKILLS SUMMARY

Programming & Query — SQL (Snowflake, Redshift, BigQuery) • Python (Pandas, NumPy) • Excel (PivotTables, Power Query)

Cloud & Storage — AWS S3 • Azure Blob Storage • Google Cloud Storage

Data Engineering & Orchestration — AWS Glue (Spark) • Apache Airflow / Cloud Composer • Kafka (basics) • dbt Core

Warehousing / Lakehouse — Snowflake • Amazon Redshift • Google BigQuery • Databricks Delta Lake

BI & Visualisation — Power BI • Tableau • Looker Studio • Statistics / A-B Testing

Version Control & CI/CD — Git • GitHub Actions • Azure DevOps Pipelines

WORK EXPERIENCE

Business Analyst Intern | SRCOSMOS LLC | Chicago, Illinois

April 2024 – Present

- * Led deep-dive **SQL** analysis on 10M + engagement rows in **Snowflake**, boosting revenue-forecast accuracy **15 %**.
- * Built & maintain 8 interactive **Power BI** / Excel KPI dashboards for Product, Finance, Design & Marketing; month-end reporting time cut **40 %**.
- * Optimised complex window-function queries (CTEs, partitions) — report runtime dropped from 12 min → 90 sec.
- * Partner with cross-functional stakeholders to turn business questions into scalable **dbt** models and **Airflow**-scheduled refreshes.
- * Present executive insights that shaped 2 roadmap OKRs and a €200 k marketing-budget shift.

Data Engineer | ADAA IT Solutions Private Limited | Bengaluru, India

April 2022 – December 2023

- * Designed **Python (Pandas, NumPy)** + **SQL** ETL pipelines orchestrated by **Apache Airflow**; automated ingest from 10 + APIs & CSV feeds.
- * Loaded raw data into **AWS Redshift** and transformed it with **dbt Core**; query latency improved **65 %**.
- * Managed MySQL & PostgreSQL OLTP databases; scripted schema migrations under **Git** + GitHub Actions CI.
- * Deployed serverless enrichment on **AWS Lambda**, integrated with **S3** tiered storage; pipeline cost down **30 %**.
- * Implemented data-quality tests (Great Expectations) raising accuracy rate from 88 % → 99 %.
- * Built Tableau & Power BI performance dashboards adopted by 3 departments, accelerating decision-making **20 %**.

PROJECTS

Diabetes Early Disease Prediction | Python, Pandas, Seaborn, Scikit-Learn, XGBoost | [LINK](#)

July 2024

- Built a machine learning pipeline to predict early onset of diabetes using real-world patient health data.
- Conducted extensive exploratory data analysis (EDA) and feature selection to identify key health indicators.
- Addressed class imbalance, scaling, and preprocessing to improve model robustness and generalization.
- Trained and evaluated multiple models (Logistic Regression, Random Forest, SVM), achieving 95% prediction accuracy.
- Reduced prediction error rates by 5% through hyperparameter tuning and model optimization.

Patient Case Similarity Engine | Python, Scikit-Learn, TF-IDF, KMeans, NLTK | [LINK](#)

May 2024

- Developed a recommendation system to suggest similar patient case studies based on diagnosis and symptom overlap.
- Applied text preprocessing, TF-IDF, and dimensionality reduction to enhance semantic similarity detection.
- Implemented KMeans clustering to group semantically similar patient cases, improving retrieval relevance by 30%.
- Strengthened diagnostic support for healthcare professionals through faster and more accurate case study retrieval.
- Developed scalable, cloud-based data pipelines that improved data processing efficiency by 30%.

CERTIFICATIONS

- Google Data Analytics Professional Certificate — [Google](#)
- Microsoft Power BI Data Analyst Associate — [Microsoft](#)
- IBM Data Analyst Professional Certificate — [IBM](#)
- Microsoft Data Visualization Professional Certificate — [Microsoft](#)
- Advanced SQL Certification — [HackerRank](#)