

Vani Sai Kumar Komarraju

513-400-2992 | k.vanisaikumar@gmail.com | linkedin.com/in/vanisai | Portfolio

EXPERIENCE

Data Engineer

Infosys

August 2022 – July 2023

Hyderabad, India

- Designed and maintained scalable data pipelines on AWS using AWS Lambda, Amazon S3, and Amazon Redshift to support efficient ingestion and processing of provider, patient, and claims data across multiple healthcare systems.
- Built real-time data streaming solutions using Amazon Kinesis and Apache Kafka to enable timely operational monitoring across healthcare workflows, and utilized AWS Athena for ad-hoc querying of clinical and operational data stored in S3, enabling fast insights into customer behavior and operational metrics.
- Developed interactive healthcare dashboards using Power BI, visualizing KPIs related to claims processing, provider performance, and patient engagement to assist executive decision-making.
- Implemented relational databases using Amazon RDS (PostgreSQL, MySQL) and Amazon Redshift to store and query large-scale provider and claims data, and processed high-volume healthcare transactions using Apache Spark on Databricks and dbt, improving reporting accuracy and enabling timely, data-driven clinical insights.
- Utilized Microsoft SQL Server, Oracle, and Teradata for advanced SQL development, building complex stored procedures and views to support regulatory reporting, risk stratification, and healthcare analytics.
- Integrated Power BI with AWS Redshift, SQL Server, and Snowflake to deliver real-time performance tracking and compliance reporting, and containerized healthcare microservices using Docker, orchestrating deployments with Kubernetes to enhance scalability, resilience, and operational efficiency.

SKILLS

Data & Analytics: SQL, Python, Power BI, Tableau, Excel (Pivot Tables, VLOOKUP), EDA, A/B Testing, Hypothesis Testing, Data Cleaning, Feature Engineering

Data Engineering: Airflow, dbt, Snowflake, PostgreSQL, Apache Spark, PySpark, Data Warehousing, ETL Pipelines, Git, Docker

Business & Product Analytics: KPI Development, Dashboard Design, Product Usage Tracking, Funnel Analysis, User Segmentation, Stakeholder Reporting, Experiment Analysis

Cloud & DevOps: AWS (S3, EC2, Lambda), Azure, CI/CD Pipelines, Shell Scripting, GitHub Actions, Cost Optimization, Resource Monitoring

Machine Learning & MLOps: Scikit-learn, XGBoost, Random Forest, Clustering (KMeans, Hierarchical), FastAPI, Flask, Model Deployment, MLOps Concepts, Reproducibility

Collaboration & Tools: Agile/Scrum, JIRA, Confluence, Smartsheet, Notion, Slack, Documentation & Versioning

PROJECTS

Predictive Modeling for Patient Length of Stay using ML — *Python, Scikit-learn, TensorFlow, Streamlit, XGBoost, LightGBM, SHAP, SMOTE*

Sentiment Analysis of Product Reviews using Python and NLP | *Python, Pandas, NumPy, Scikit-learn, Matplotlib, Power BI, NLTK, Gemini AI, EDA*

State Farm Distracted Driver Detection | *Python, TensorFlow, PyTorch, Data Augmentation, Keras, Flask, CNNs, OpenCV*

Student Dropout Prediction | *Python, Pandas, Scikit-learn, NumPy, Matplotlib, Seaborn, Logistic Regression, K-Means, RMSE*

Data Warehouse and Analytics Project using SQL Server | *SQL, ETL, Star Schema, Data Modeling, SSMS, Power BI, dbt, Medallion Architecture*

EDUCATION

University of Cincinnati

Master of Engineering in Data Science, GPA 3.7

Cincinnati, OH

May 2025