SaiSindhu Nallapaneni | Data Analyst

TX, USA | (972) 643 8285 | saisindhunallapaneni1106@gmail.com | Linkedin | Portfolio

SUMMARY

Results-driven **Data Analyst** with **4+ years of experience** delivering high-impact analytics solutions across **regulated industries**, including healthcare. Skilled in designing end-to-end data pipelines, predictive modeling, and interactive dashboards that drive business decisions and operational efficiency. Proficient in Python, SQL, dbt, Apache Spark, and cloud platforms (AWS, Azure), with hands-on expertise in real-time analytics using Kafka, Snowflake, and Power BI. Adept at collaborating with cross-functional teams, ensuring data quality and governance through modern frameworks, and translating complex data into actionable insights. Expert in visualizing insights via Power BI and R Shiny, while ensuring HIPAA/FDA compliance and integrating standards like FHIR and HL7. Demonstrated success in improving process efficiency, optimizing resource allocation, and enabling data-informed strategies across diverse domains.

TECHNICAL SKILLS

Languages: Python, SQL, R

Tools: Pandas, NumPy, Scikit-learn, Plotly Dash, SciPy, Statsmodels, GitHub, Confluence, Notion, Great Expectations, dbt, AWS Glue, AWS Lambda

Databases & ETL: Apache Spark, Apache Airflow, PySpark, Apache Kafka, AWS Glue, dbt, Snowflake, Databricks, SQL, Neo4j, GraphQL

Data Visualization: Power BI, R Shiny, Plotly Dash, Notion, Tableau

Methodologies: A/B Testing, Feature Engineering, Dimensionality Reduction, Dimensional Modeling, Cross-Functional Collaboration, Data Storytelling, Regulatory Compliance (HIPAA, FDA)

Big Data Technologies: Apache Spark, Apache Kafka, PySpark

Cloud Platforms: AWS (S3, Redshift, EC2), Azure (Data Factory, SQL Database)

Data Governance: Great Expectations, Collibra, AWS Lake Formation, FHIR, HL7, LOINC, HEDIS, CMS Star Ratings

Operating Systems: Windows, Linux

PROFESSIONAL EXPERIENCE

Data Analyst | UnitedHealth Group, USA **Feb 2025 – Present**

- Devised real-time analytics pipelines using Apache Kafka and AWS Lambda to monitor patient engagement with chronic care programs, accelerating care coordinator response times and improving adherence to treatment plans.
- Optimized data models in Snowflake and dbt to support cost -benefit analysis of Medicare Advantage enrollees across various care pathways, enabling finance teams to prioritize funding for high-impact interventions.
- Designed stratification logic using Python, SQL, and XGBoost to segment dual-eligible populations based on utilization patterns, which informed targeted care plans and drove a 12% increase in patient satisfaction scores.
- Built HIPAA-compliant patient journey maps by integrating SDOH, claims, and EHR data using Neo4j, GraphQL, AWS Glue, and PySpark to identify care gaps and improve outreach through targeted health coaching.
- Built Excel-based financial models to estimate risk-adjusted revenue projections, integrating with CMS data and applying scenario analysis with what-if tools.
- Benchmarked provider performance using R Shiny dashboards linked to CMS star ratings and HEDIS scores, enabling provider networks to visualize care quality gaps and align incentives accordingly.
- Conducted A/B testing on member communication using Python (SciPy, Statsmodels) and shared insights via Power BI and Notion in cross-functional sessions with clinical leads, driving quality improvements while ensuring compliance through Collibra and AWS Lake Formation.

Data Analyst Intern | UnitedHealth Group, USA **Sep 2024 – Dec 2024**

- Developed and deployed real-time analytics pipelines integrating Apache Kafka and AWS Lambda, enabling live monitoring of patient engagement in chronic care programs and expediting care team interventions.
- Optimized and maintained data models in Snowflake and dbt, supporting financial analysis for Medicare Advantage plans and informing high-impact funding decisions across care pathways.
- Designed population stratification algorithms using Python, SQL, and XGBoost, segmenting dual-eligible members to personalize interventions, resulting in measurable increases in patient satisfaction.
- Constructed HIPAA-compliant patient journey maps by integrating SDOH, claims, and EHR data with Neo4j, GraphQL, AWS Glue, and PySpark, which identified care gaps and informed targeted health coaching initiatives.
- Led A/B testing and advanced statistical analysis of member communications using Python (SciPy, Statsmodels); communicated insights in Power BI and Notion to drive continuous quality improvement and regulatory compliance.

Data Analyst | CitiusTech, India **May 2020 - Apr 2023**

- Analyzed diverse healthcare datasets using Python (Pandas, NumPy) and Apache Spark to uncover data quality issues and standardization gaps across AI models, facilitating a 20% increase in Medictiv model interoperability.
- Engineered ETL pipelines using SQL, dbt, and Apache Airflow to aggregate model metadata from federated clinical systems, ensuring accurate and consistent data ingestion to support the open AI model directory.
- Developed dynamic dashboards using Excel PivotTables, slicers, and charts to summarize claims and patient utilization data for executive stakeholders.
- Visualized AI model performance metrics and patient cohort segmentation using Power BI and Plotly Dash, enabling data scientists and clinicians to compare efficacy across models and make evidence-based decisions.
- Synthesized clinical, claims, and genomic data using Snowflake and Databricks, driving integration of multimodal datasets to enhance the transparency and explain ability of models listed on Medictiv.
- Validated data pipelines and model evaluation outputs by designing automated data profiling scripts in Python and Great Expectations, reducing manual QA efforts by 35% and boosting model reliability in production.
- Collaborated with various teams to map healthcare ontologies using FHIR, HL7, and LOINC standards, improving semantic consistency and enriching metadata quality across 100+ AI model entries.
- Curated datasets for AI model benchmarking using feature engineering and dimensionality reduction in Scikit-learn, enabling deployment of 15 models in Medictiv's directory, while documenting workflows and metadata lineage in Confluence and GitHub to ensure HIPAA- and FDA-compliant governance.

EDUCATION & CERTIFICATIONS

Master's in computer science: Auburn University at Montgomery, AL, USA **May 2023 - Dec 2024**

- Google Data Analytics Professional Certificate – Google

Bachelor of Engineering in ECE: Panimalar Engineering College, Anna University, India **May 2017 - Apr 2021**

- Python Programming Foundation– Apponix Technologies
- Data Science with Python Workshop– Genesys Academy
- Python Programming for Everybody – University of Michigan, Coursera

PROJECTS

Academic Research Project & Published Paper – IJSART Journal **Apr 2021**

- Published research in IJSART on designing a CNN-based brain tumor classifier using segmented MRI data, combining traditional image preprocessing in MATLAB with deep learning techniques to improve diagnostic accuracy.

Predictive Modeling for Healthcare Claims Cost Optimization **Nov 2024**

- Developed a predictive model using scikit-learn to forecast high-cost claims from Medicare Advantage datasets by integrating SDOH, EHR, and claims data.
- Engineered features like comorbidity flags, provider type, and utilization frequency to improve prediction accuracy (AUC: 0.87)
- Automated ETL workflows using dbt and AWS Glue to ingest and transform CMS and payer data into Snowflake for downstream analytics.
- Created Power BI dashboards for actuarial and clinical teams to monitor projected vs. actual costs across care pathways, supporting value-based care planning.

Oil Field Equipment Monitoring & Downtime Optimization Dashboard **Apr 2025**

- Built an automated analytics dashboard using Seeq and Power BI to track critical equipment performance (compressors, pumps) using AVEVA PI System time-series data.
- Applied statistical anomaly detection and rolling-window averages to identify leading indicators of equipment failure.
- Correlated asset downtime with external variables (temperature, shift timing, operator logs), uncovering patterns that led to a 15% reduction in unplanned outages.
- Integrated historical SCADA logs via Python and SQL pipelines into a centralized Azure SQL database to support root cause analysis.
- Enabled field engineers and operations managers to make data-driven maintenance decisions, improving asset utilization rates.

Deloitte Australia - Data Analytics Job Simulation – Forage **Jul 2025**

- Completed a Deloitte job simulation involving data analysis and forensic technology
- Created a data dashboard using Tableau
- Used Excel to classify data and draw business conclusions