

SUMMARY

Experienced Data Analyst with 4+ years of success supporting clinical research through advanced data management, statistical modeling, and analytics.

- Specialized in curating longitudinal registries, harmonizing real-world data, and automating quality control processes to drive accurate reporting and health surveillance.
- Proficient in SQL, R, Python, Stata, and Git, with hands-on experience in EHR data extraction, clinical systems integration, and developing secure, scalable pipelines across research environments.
- Experienced in configuring server tasks, managing data dictionaries, validating datasets, and executing analyses aligned with IRB, HIPAA, and NIH compliance standards.
- Adept at working with interdisciplinary teams, mentoring junior scientists, and delivering visualizations, documentation, and statistical outputs that inform research publications and funding proposals.

EDUCATION

<b>Master of Science in Business Analytics and Information Systems</b>	<b>May 2024</b>
University of South Florida, Tampa FL USA	CGPA -3.94/4
<b>Bachelor of Technology in Electronics &amp; Communication Engineering</b>	<b>May 2019</b>
SR Engineering college, India	CGPA -9.14/10

SKILLS

- **Programming & Scripting:** Python, R, SQL, C++, Java, Bash, Unix Shell, SAS
- **Data Science & Statistical Tools:** Scikit-learn, TensorFlow, Bioconductor, MATLAB, NumPy, pandas, SciPy, Statsmodels
- **Data Engineering & Processing:** Apache Airflow, HL7 Integration, REST APIs, JSON, Parquet, ETL Design
- **Databases & Warehousing:** PostgreSQL, SQL Server, MySQL, Snowflake, Oracle, REDCap API
- **Biomedical & Research Systems:** EHR Systems (Epic, Cerner), Clinical Registries, NIH/UCSF Standards
- **Cloud & HPC:** Azure (Synapse, Data Factory, Key Vault, Databricks), AWS (S3, Lambda, RedShift, EC2, Glue, Athena), GCP (BigQuery, DataFlow, Pub/Sub), Linux HPC Clusters, Kubernetes, Docker
- **Machine Learning & AI:** Keras, PyTorch, XGBoost, Hugging Face Transformers, LangChain, NLTK, SpaCy, Reinforcement Learning, NLP, Time Series Forecasting, MLOps (CI/CD, Model Registry)
- **Visualization & Reporting:** Tableau, Power BI, Streamlit, Dash, Matplotlib, Seaborn, Plotly
- **Project & Agile Tools:** JIRA, Confluence, Trello, Scrum/Kanban, Reproducibility & Documentation Standards
- **Version Control & DevOps:** Git, GitHub, GitLab, Azure DevOps, CI/CD Pipelines
- **Professional Competencies:** Data Governance, Research Collaboration, Data Quality, Cross-Disciplinary Communication

PROFESSIONAL EXPERIENCE

<b>Data Engineer   CVS Health   USA</b>	<b>January 2024 - Present</b>
<ul style="list-style-type: none"><li>• Developed multiple scalable ETL pipelines using <b>Databricks</b>, <b>Kubeflow</b>, and <b>Snowflake</b>, seamlessly integrating with <b>Azure Cloud</b>, <b>Azure Data Lake Storage (ADLS)</b>, <b>AKS</b>, and <b>SFTP</b> to/from JDA servers.</li><li>• Designed and implemented an end-to-end pipeline using <b>Snowpark</b>, leveraging <b>Snowflake warehouse compute</b> exclusively to run and score ML models, eliminating the need for tools like Airflow and Databricks, and reducing compute costs by <b>20%</b>.</li><li>• Built reusable <b>Argo workflow templates</b> and developed node-specific ETL jobs in <b>Kubeflow</b>, optimizing resource allocation and cutting infrastructure costs by <b>50%</b>.</li><li>• Contributed to a <b>\$10 million business impact</b>, including <b>\$500K</b> in value from enhanced precision tracking enabled by Kubeflow over Databricks.</li><li>• Improved workload efficiency in <b>Azure Kubernetes Service (AKS)</b> through advanced workflow automation and strategic resource provisioning.</li><li>• Reduced unnecessary node computations by optimizing model input features, enhancing overall ML pipeline performance.</li><li>• Integrated <b>Snowflake stored procedures</b> and <b>vectorized UDFs</b> for complex data transformations, improving analytical capabilities.</li><li>• Engineered robust ETL pipelines in <b>Databricks</b> to migrate data from <b>Oracle to Snowflake</b>, addressing schema mismatches and preserving data integrity.</li><li>• Automated and orchestrated workflows using <b>Apache Airflow</b> and fine-tuned processing with <b>PySpark</b>, improving runtime performance.</li><li>• Diagnosed and resolved critical issues in data workflows, ensuring accurate event processing and stabilizing key business operations.</li></ul>	

- Improved data pipeline runtimes by **20%** through **Apache Spark** performance tuning and efficient debugging practices.

**Tools & Tech:** Python, SQL, Snowflake, Databricks, Snowpark, Kubeflow, Azure (ADLS, AKS), Oracle, Apache Airflow, PySpark, Argo Workflows, SFTP, Git, Bash, Linux, Vectorized UDFs, Stored Procedures

#### Data Engineer | Accenture | Hyderabad, India

August 2020 - July 2022

- Developed custom **Python** and **SQL** scripts and pipelines to extract and harmonize **EHR** data from multiple clinical systems, supporting HIV registry development.
- Designed a research-ready **relational data model** for integrating **CFAR** clinical research data, enabling efficient statistical analysis and cohort tracking.
- Performed statistical modeling and ad hoc analysis using **Python (Pandas, NumPy)** and **R (tidyverse, dplyr)**, contributing to multi-site collaborations.
- Collaborated on research publications, preparing reproducible **analytics code** and documentation aligned with **IRB** and **data governance** protocols.
- Implemented **SQL views** and **stored procedures** to automate reporting of **quality assurance** metrics and site-level engagement indicators.
- Developed and maintained a comprehensive **data dictionary** and **metadata repository** to ensure transparency and consistency across multi-source datasets used in **HIV/AIDS research** and **clinical trial** analyses.

**Tools & Tech:** SQL, Python, R, PostgreSQL, Microsoft SQL Server, GitLab, Confluence, Bash, HIPAA, IRB Standards

#### Data Analyst | Tech Mahindra | Hyderabad, India

May 2019 - July 2020

- Led end-to-end **data analysis initiatives on AWS**, supporting insights for over **100 enterprise applications** with scalable and cost-efficient analytics solutions.
- Built and maintained **data pipelines** using **AWS Glue, S3, and Redshift**, transforming and processing **5+ TB of data daily** for real-time reporting.
- Designed and delivered **interactive dashboards** in **AWS QuickSight, Power BI, and Tableau**, supporting strategic decisions across sales, operations, and leadership.
- Wrote and optimized complex **SQL queries** in **Amazon Redshift** to extract, clean, and analyze large datasets, identifying trends that improved efficiency by **15%**.
- Conducted advanced analysis using **Python** and **Pandas**, including **anomaly detection, A/B testing, and predictive modeling** to support data-driven marketing and operational strategies.
- Automated **ETL workflows** and data validation using **Python** and **AWS services**, improving reliability and reducing manual intervention by **30%**.
- Partnered with business stakeholders to gather reporting requirements and deliver **custom analytics solutions**, while creating **data dictionaries** to support self-service BI.

**Tools & Tech:** SQL, Python, AWS, Power BI, Tableau, Amazon Redshift, Git, Cloud Monitoring, RBAC, A/B Testing, Anomaly Detection, Predictive Modeling

#### CERTIFICATIONS

- Business Analysis & Process Management – [\(Link\)](#)
- Working with BigQuery – [\(Link\)](#)
- Database Operations in MariaDB Using Python – [\(Link\)](#)

#### ACADEMIC PROJECTS

- Real time Stock Market Analysis:** Designed and implemented a real-time data pipeline using Python, Apache Kafka, and AWS services. Built a stock market simulation app to generate CSV data, sent to Kafka on Amazon EC2 via Boto3, and stored in Amazon S3. Cataloged data with AWS Glue and analyzed it using Amazon Athena. [\(Link\)](#)
- Made Easy Pharmacy Store Website:** We developed and implemented a user-friendly and visually appealing e-commerce website for a pharmacy store, providing seamless online shopping experience for customers. We used technologies to develop the website HTML, CSS, JavaScript, Flask. [\(Link\)](#)
- Airbnb project:** With SQL Developer, created a database like Airbnb's business workflow using the data-dimensional modeling approach. It helps capture business data and can handle operations smoothly. From the data, we can infer targeted customers and the footfall of the customers. [\(Link\)](#)
- Data visualization Project:** Developed interactive Tableau dashboards to analyze the EPL 2018-19 season, assess the growth of Indian startups, explore Netflix content trends using IMDb data, and evaluate the health impacts of wildfires. [\(Link\)](#)