

Skills

Programming Languages: PostgreSQL, MySQL, MSSQL, Python (Pandas, NumPy, Matplotlib)

Reporting Software: Excel, Tableau, Knime Analytics Platform, Google Analytics

Customer Relationship Management Software: SalesPro, Salesforce

Remote Access Software: Crash FTP, Splashtop Remote Desktop,

Cloud Software: AWS(S3, Athena, Glue, Redshift), Databricks, Microsoft Azure, DBT, Snowflake

Project Management: SCRUM, Agile, ASANA Project Management Software, Jira, GitHub, Git

Project Experience

Retail Sales BI End to End Project

- Designed and built a Medallion-architecture data pipeline (Bronze/Silver/Gold) in Snowflake using dbt Cloud, ingesting 420K+ retail transaction records from AWS S3 via external stages and COPY INTO.
- Implemented SCD Type 1 and Type 2 dimensional modeling (dbt snapshots) to track historical store attribute changes, enforcing data quality with custom dbt schema tests and a generate_schema_name macro for environment isolation.
- Built a Python visualization layer connecting directly to Snowflake via snowflake-connector-python, producing 10 production-quality matplotlib/seaborn reports without a third-party BI tool.
- Automated raw-to-cloud data ingestion using Python/boto3, resolving file-path and NULL-handling edge cases (e.g., "NA" string markdown values) at the Snowflake COPY INTO layer.

Healthcare Staffing Analytics End to End Project

- Engineered an AWS data lake pipeline (S3/Glue/Redshift) processing the full 780MB CMS Nursing Home Provider Data Catalog (22 files, 1.3M+ staffing records, 14,564 facilities) using PySpark for scale.
- Identified and resolved a production data quality issue (UTF-8 decode failure) in a 1.3M-row government dataset, and validated a 99.9% join match rate across the staffing and provider reference tables using CMS Certification Numbers.
- Designed and built 5 Gold-layer metric tables (nurse-to-patient ratio, occupancy rate, contracted-staffing reliance, compliance risk score, quality rating distribution) joining staffing, ownership, citation, and penalty data sources.
- Delivered a 6-page Streamlit dashboard with facility-level drill-down, directly answering stakeholder questions on the relationship between nurse staffing levels and facility occupancy/quality outcomes.

Customer Lifetime Value & Restaurant Analytics End to End Project

- Designed an AWS-only (no Snowflake/dbt) data platform extracting from a SQL Server source via Glue JDBC, implementing 100% of transformation logic in PySpark per strict enterprise tooling constraints.
- Built a daily-evolving Customer Lifetime Value model using Spark window functions over a 200K+ record order dataset, enabling true day-over-day CLV trend tracking rather than static aggregate reporting.
- Implemented RFM customer segmentation and discount-effectiveness analysis in PySpark, and architected production-grade fault tolerance using Step Functions retry policies and SQS dead-letter queues.
- Delivered a CI/CD pipeline via GitHub Actions for automated linting, testing, and deployment of Glue ETL scripts, alongside 6 Streamlit dashboards covering segmentation, churn, loyalty, and location performance.

Real-time Marketing, CRM & Scheduling Platform End to End Project

- Architected and built a real-time, event-driven data platform on AWS (API Gateway, Lambda, SQS, DynamoDB) ingesting live CRM and scheduling webhooks plus a paginated third-party analytics API, fully event-driven rather than batch.
- Implemented a 10-minute SQS delay-queue mechanism to coordinate asynchronous lead-enrichment timing, with signature-verified webhook authentication and idempotent, deterministic S3 key design to safely handle webhook redelivery.
- Validated system reliability through a 7-day continuous production run, including a deliberately engineered failure-and-recovery test and duplicate-event idempotency test, with CloudWatch-based monitoring and SQS dead-letter queue alerting.
- Delivered real-time Slack notifications enriched with cross-system lead/owner data and built 6+ Streamlit dashboards correlating marketing video engagement, scheduling, and CRM lead generation across three independent data sources.

Experience

Amazon

Boston, MA

Data Associate II

May 2025 – Present

- Partner with AGI-DS (Artificial General Intelligence – Data Services) team to enhance Alexa and Nova Services by validating, annotating, and auditing datasets in order to meet quality standards.
- Selected to review peer work to ensure compliance with company standards after two months of tenure.
- Established and participated in the site's IT committee, improving communication with corporate IT and reducing downtime caused by technical issues.
- Ranked in the top 5% of Data Associates for accuracy and task completion speed.
- Help to establish the GROW Initiative on site which is dedicated to streamlining communication processes and improving employee retention throughout the multiple AGI-DS sites worldwide
- Lead a Pizza team to change organizational policies to allow Data Associates to access training materials outside of production areas. Allowing Data Associates worldwide access to free career upskilling.

Decision Science LLC

Newburyport, MA

Data Analyst

Jul 2023 – Mar 2024

- Designed database schema, tables, and SFTP server credentials for new Experian-approved clients.
- Built and optimized Knime workflows to process large datasets (40,000–120,000 records) using PostgreSQL queries.
- Automated secure data delivery via SFTP, reducing manual processing time.
- Extracted CRM data and produced client reports using PostgreSQL to track direct mail campaign performance using unique identifiers like response rate and average response time.

LGC Group

Manchester, NH

Lead Data and Systems Analyst

Aug 2022 – Jul 2023

- Extracted and transformed CMS data to produce reports and visualizations for management teams using PostgreSQL and Tableau.
- Developed dashboards to track Production, QC, and Warehouse team output, increasing operational visibility for roughly 50 employees.
- Enhanced training programs to improve tool usage and efficiency across 15 employees, creating a ~97% increase in productivity on client specific tasks.
- Maintained historical KPI datasets and delivered on-demand data snapshots to support departmental decision-making.
- Coordinated with international analytics teams to align reporting standards across 7 locations on site in order to increase visibility into daily productivity for each department.

Education

Master of Science, Computer Science – Southern New Hampshire University – 2026-Present

Master of Science, Information Technology – Southern New Hampshire University – 2021 – 2024 (GPA:3.3)