

SUMMARY

Data Scientist & Data Engineer with experience in payer-provider healthcare transparency data and large-scale streaming analytics. Skilled in building ETL pipelines, data quality frameworks, and warehouse models using Snowflake, GCP, and AWS. Adept at transforming complex datasets into dashboards and AI-driven insights that drive business and policy decisions.

SKILLS

- Technical:** Python (Pandas, NumPy, Scikit-learn, Matplotlib), SQL (Snowflake, PostgreSQL), PySpark, ETL/ELT, Airflow, AWS (S3, Lambda, Glue, RDS), GCP (GCS, BigQuery, Dataflow, Dataproc), Docker, GitHub Actions, Tableau, Power BI, Looker
- Analytical & Business:** Data Cleaning & Transformation, Data Quality Validation, Statistical Analysis, Machine Learning (Regression, Clustering), Dashboard & KPI Development, Cost Optimization, Stakeholder Reporting

EXPERIENCE

Data Scientist

Sept 2025- Present

PayerPrice | Boston, MA

- Built automated Snowflake pipelines to clean and standardize payer-provider transparency data, resolving inconsistencies with data quality scorecards and improving throughput efficiency by 79%, ensuring reliable datasets for pricing analysis.
- Processed and analyzed 500GB+ of raw healthcare pricing files to identify unique billing codes and normalize messy provider inputs for downstream analytics.
- Designed a regex-based classification framework to extract and validate CPT, HCPCS, REV, MS-DRG, and APR-DRG codes, significantly improving accuracy of mappings.
- Applied AI similarity scoring to match raw provider descriptions against canonical code tables, implementing confidence tiers to automate high probability matches while flagging ambiguous cases for review.
- Developed Tableau dashboards to visualize cleaned payer-provider datasets, uncover pricing patterns, and monitor mapping coverage, enabling healthcare stakeholders to make data-driven decisions.

Data Analyst Intern

Allagash Brewing Company | Northeastern University

Jan 2025- March 2025

- Developed a real-time truck recommender system using weather forecast APIs and shipping data, reducing transportation costs by 10%.
- Built interactive Power BI dashboards to visualize shipment trends, pallet utilization, and recommendations, enabling data-driven decision-making across logistics and operations teams.
- Implemented clustering (DBSCAN, K-Means) and route optimization techniques (Google OR-Tools, TSP, VRP), improving delivery routing and reducing fuel costs.
- Collaborated cross-functionally to deliver insights and actionable recommendations that directly contributed to the company's cost-saving strategies.

Data Engineer

SM Enterprises | Mumbai, India

December 2019 - August 2023

- Designed and implemented ETL pipelines for terabytes of user play events, subscription records, and catalog metadata, ingesting raw logs into Snowflake and AWS S3 for downstream analytics.
- Built batch and near real-time pipelines using Python, PySpark, and Airflow, reducing data latency by 35% and enabling timely reporting on user engagement.
- Developed fact/dimension warehouse schemas to support listener behavior analysis, churn prediction, and personalized recommendation engines.
- Implemented data quality validation frameworks (schema checks, anomaly detection, AI-based similarity scoring) to ensure consistency across billions of streaming records.
- Partnered with analytics and product teams to deliver Tableau and Power BI dashboards tracking engagement trends, top tracks/artists, and subscription churn metrics.
- Automated infrastructure provisioning and pipeline orchestration with Docker and Boto3, improving scalability and reducing manual maintenance.

PROJECTS

Healthcare Natural Language to SQL Assistant

- Developed a full-stack healthcare analytics assistant enabling natural language querying of hospital datasets using Gemini 2.0 Flash and LangChain agents.
- Processed payer-provider cost data in Snowflake and created a Pinecone vector database (MiniLM embeddings) for semantic retrieval.
- Engineered multi-agent pipelines to interpret user queries, generate and execute SQL, and return insights with automated visualizations.
- Built a React frontend and containerized backend (FastAPI, Docker), ensuring scalability and easy deployment.

EDUCATION

Northeastern University | GPA: 3.92

Boston, MA

Master of Science in Data Analytics (Applied Machine Intelligence)

September 2023 – March 2025