

# Soham Mane

 SohamMane812 |  Soham |  Portfolio |  sohammane01@gmail.com |  8578918703

## SUMMARY

---

Data Analyst & Engineer with 4+ years delivering measurable impact from contributing to \$13.6M in annual cost savings to building pipelines serving 100M+ users in healthcare and OTT domains using GCP, Snowflake.

## SKILLS

---

**Technical:** Python (Pandas, NumPy, PySpark), SQL (Snowflake, BigQuery, PostgreSQL), Airflow, Docker, AWS (S3, Glue, Lambda), GCP (BigQuery, Dataproc, Dataflow), Tableau, Power BI, GitHub

**Analytical:** ETL/ELT Pipeline Development, Risk Adjustment Modeling (CMS-HCC V28), Data Validation, Statistical Analysis, Machine Learning (Regression, Clustering), Dashboard & KPI Development

## WORK EXPERIENCE

---

**Data Science Intern | PayerPrice, Boston MA** Sept 2025 – Oct 2025

- Cleaned and mapped 500GB+ of hospital charge data to NPI provider records using Snowflake pipelines, improving processing speed by 79% and enabling accurate provider-level pricing comparisons for clients.
- Designed a regex and Snowflake AI.SIMILARITY scoring framework to classify and validate ICD-10 billing codes, automating high-confidence matches and reducing manual review burden by 40%.
- Developed Tableau dashboards visualizing provider-level pricing patterns, enabling healthcare clients to benchmark costs and make informed network decisions.
- Delivered confidence-tiered code normalization standards across healthcare analytics workflows, improving mapping accuracy across insurance plans.

**Data Analyst Intern | Allagash Brewing Company** Jan 2025 – Mar 2025

- Developed a real-time truck recommender using weather APIs, clustering, and Python, contributing to \$13.6M in annual transportation cost savings alongside the logistics team.
- Applied DBSCAN and K-Means clustering with Google OR-Tools route optimization, reducing fuel costs by 30% across distribution operations.
- Built Power BI dashboards tracking shipment trends and pallet utilization, enabling logistics and supply chain managers to make faster operational decisions.

**Data Engineer | SM Enterprises (OTT & Media Analytics), Mumbai** Dec 2019 – Aug 2023

- Built Python, PySpark, and Airflow ETL pipelines ingesting terabytes of OTT content and user behavior data into Snowflake and AWS S3, reducing data latency by 35% and enabling near real-time reporting.
- Delivered warehouse schemas powering recommendation engines serving 100M+ users, improving content watch time and reducing subscriber drop-off rates.
- Implemented data quality frameworks with schema checks and anomaly detection across billions of records, ensuring pipeline reliability for analytics teams.

## PROJECTS

---

**ACO Risk Adjustment Dashboard – CMS-HCC V28 Analytics Platform | [Link to Website](#)**

- Built a 7-step BigQuery SQL pipeline implementing CMS-HCC V28 for 15,000 Medicare beneficiaries – ICD-10 mapping (8,019 codes), hierarchy trumping (109 pairs), coefficient scoring across 7 cohorts, and 11 interaction.
- Deployed a live dashboard with population risk stratification, monthly GHCC trends, and coding gap detection to surface undercoded patients for care management prioritization.

**HospRates – Healthcare Price Transparency Analytics Platform | [Link to Website](#)**

- Built a GCP ETL pipeline processing 100GB+ of U.S. hospital pricing data, normalizing ICD-10 codes across 80 payers and 231 insurance plans for procedure-level cost benchmarking.
- Compared standardized hospital charges against Medicare benchmark rates, enabling data-driven pricing transparency across procedures.

## EDUCATION

---

**Northeastern University | GPA: 3.92** September 2023 – March 2025

Master of Science in Data Analytics, **Specialization:** Applied Machine Intelligence