

EXPERIENCE

Data Analyst Intern

Allagash Brewing Company | Northeastern University

Jan 2025- March 2025

- Reduced freight cost by 10% by building a dynamic truck recommender system using weather APIs and shipping data, deployed via Python and Power BI.
- Enabled logistics insights through real-time dashboards visualizing shipment trends, pallet utilization, and delivery performance using Power BI and SQL-based modeling.
- Enhanced route efficiency by 15% through clustering (DBSCAN, K-Means) and optimization algorithms (Google OR-Tools), reducing delivery time and fuel costs.

Data Engineer

SM Enterprises | Mumbai, India

January 2020 - August 2023

- Engineered robust data pipelines using Python and Airflow to ingest and transform high-volume marketing and customer engagement data into Snowflake, reducing manual reporting by 80%.
- Integrated Airtable with Snowflake via Python-based connectors and REST APIs, enabling real-time synchronization of campaign metadata for marketing operations teams.
- Built SQL-based transformation layers (dbt-style) to standardize diverse data sources, email campaigns, CRM, web analytics into a unified model for reporting and attribution.
- Led the development of automated quality checks using Python and SQL (null detection, outlier detection), raising data accuracy across marketing datasets by 25%.
- Partnered with analytics and media teams to deliver self-serve reporting dashboards using Snowflake views and downstream tools (Power BI), improving campaign optimization velocity.
- Developed reusable Python modules for log parsing and API extraction, enhancing engineering productivity and shortening onboarding for new team members.
- Documented data models, pipeline workflows, and lineage using Airtable as a source of truth, enabling business users to trace KPIs back to source logic with ease.
- Supported transition to a remote-first engineering culture, mentoring new hires, and introducing sprint rituals and code review practices aligned with Agile.

PROJECTS

Healthcare Natural Language to SQL Assistant

- Built a full-stack analytics assistant enabling natural language querying over Snowflake-hosted hospital data using LLM and LangChain.
- Designed a multi-agent backend with FastAPI and Pinecone for metadata retrieval, SQL generation, execution, and auto-plotting.
- Dockerized and deployed React + FastAPI services, supporting environment-specific API routing and scalable container orchestration via Docker Compose.

OTT Content Analytics Pipeline

- Architected an end-to-end data pipeline using Pub/Sub, Dataflow, BigQuery, and Composer (Airflow) to process and analyze user engagement with OTT content.
- Enabled real-time and batch analytics with PySpark on Dataproc and automated DAG deployment via CI/CD (GitHub Actions), improving pipeline uptime and agility. Designed resilient infrastructure with zone fallback, dependency isolation & proactive data validation, reducing production failures by 40%.

Marketing Campaign Data Hub

- Developed a Snowflake-based reporting layer aggregating campaign performance metrics from Google Ads and Meta using Python ETL.
- Automated nightly data loads using Airflow & scheduled validation tests to ensure accuracy across spend, impressions, and ROI metrics.
- Created reusable Airtable workflows for campaign metadata intake, improving campaign launch coordination and reducing misconfigurations by 30%.

EDUCATION

Northeastern University | GPA: 3.92

Boston, MA

Master of Science in Data Analytics (Applied Machine Intelligence)

September 2023 – March 2025

SKILLS AND TECHNOLOGY

- **Technical Skills:** Python (Pandas, NumPy, Scikit-learn), SQL, Snowflake, Apache Airflow, Airtable, Apache Spark, PySpark, Docker, FastAPI, React, Apache Kafka, GCP (BigQuery, Composer, Dataflow), AWS (Lambda, S3, RDS, Glue), Git, REST APIs, dbt (style workflows), Parquet, Looker, Power BI, Tableau
- **Analytical & Business Skills:** Marketing Attribution, Campaign Analytics, Data Modeling, KPI Standardization, Stakeholder Communication, Agile/Scrum, Data Validation, Dashboard Development, Decision Support, Root Cause Analysis, Process Automation, Documentation