Aniket Patole

Boston, MA |(617)-352-5273 | aniketpatole.works@gmail.com | LinkedIn | Portfolio

EDUCATION

Master of Science in Information Systems   **Graduated May 2025**
Northeastern University, Boston, MA

Bachelor's of Engineering in Computer Science   **Graduated May 2021**
University of Pune, MH, India

WORK EXPERIENCE

Humanitarians AI   **June 2025 - Current**
AI Engineer/Developer
- Built **data preparation & ingestion scripts** to structure large content sets for LLM workflows; added **automated checks** for format/encoding anomalies and basic PII redaction before indexing
- Partnered with product to instrument **usage analytics** (event logs → curated tables) for content performance reporting

Northeastern University   **September 2024 - April 2025**
Graduate Data Engineer – Admissions Systems
- Designed and automated **ETL pipelines** integrating Salesforce, Slate & Five9, reducing reporting turnaround from 3 days to 1 and enabling scalable experimentation with admissions funnel data
- Developed interactive **Tableau dashboards & SQL-driven metrics**, tracking 8K+ applicant records per cycle, insights for leadership
- Partnered with engineers and analysts to define **dimensional data models and governed workflows**, ensuring reproducibility, compliance, and accurate interpretation of key metrics

Accenture, Pune   **August 2021 - July 2023**
Data Engineer - Logistics & Parcels Domain (Royal Mail Group)
- Processed **1M+ weekly parcel transactions** to support global **supply chain analytics** and SLA monitoring in **Teradata and Qlikview**
- Extracted insights on **fraudulent transactions**, **SLA delays**, and anomalous user activity across **10+ data sources**, directly supporting law enforcement fraud investigations and risk management
- Embedded **100+ master data syncs** compliance frameworks, strengthening audit readiness, **data reliability** across logistics workflows
- Automated **CI/CD pipelines** cloud migration to **AWS (Redshift + S3)**, cutting infra costs by 30% while maintaining zero SLA breaches
- Designed and maintained **distributed ingestion** and **transformation services (SQL, JSON, XML)** to enable consistent reporting on logistics performance, returns, and delivery reliability
- Mentored junior engineers on **SQL optimization**, **fraud analytics, experiment analysis**, improving reporting accuracy across the team

ANALYTICS & TECHNICAL PROJECTS

RAG Event Analytics Platform | Python, Airflow, LangGraph, FAISS, Pinecone, AWS
- Automated ingestion of ad event logs with **Airflow** to parse click/install data into **FAISS embeddings stored on Pinecone**
- Designed **multi-hop agent workflows** with **LangGraph + RAG** for anomaly detection in user acquisition & performance insights
- Delivered compliance-ready analytics by combining retrieval pipelines with structured datasets, improving reliability and scalability

Real-Time Order & Logistics Tracking Dashboard | Python, dbt, Databricks, Power BI, PostgreSQL
- Architected a **real-time data pipeline** to process 5M+ order and shipment events daily from partner warehouses and carriers
- Leveraged **dbt** for transformation and **Unity Catalog** for governance, enabling consistent metrics across teams
- Delivered **BI dashboards in Power BI**, tracking partner SLA adherence, delivery delays improving operational visibility by 40%

Product Catalog Normalization & Data Quality Framework | SQL, HiveQL, AWS Glue, Teradata
- Built **data models and dimensional schemas** to normalize product catalogs across diverse partners (brands, retailers, marketplaces)
- Embedded **data quality checks** and anomaly detection into pipelines, reducing invalid SKU attributes and duplicate listings by 25%

Generative AI Research Tool | AWS (EC2, RDS, S3), Python, Vector Search
- **Built modular pipelines** for **vector similarity search** and contextual enrichment of user engagement datasets
- Leveraged RAG to provide **explainable insights from ad activity logs**, ensuring accuracy and reproducibility for downstream analytics
- Partnered with engineers to optimize retrieval workflows, enabling scalable experimentation with **LLM-powered analytics**

TECHNICAL SKILLS

Languages & Analytics: SQL (Joins, Window Functions, CTEs, Optimization), Python (Pandas, NumPy, Matplotlib), PySpark, HiveQL
Fraud & Risk Analytics: Anomaly detection, CTR/fraud metrics, A/B testing, hypothesis testing, engagement & retention analysis
Visualization & BI: Tableau, Power BI (DAX), QuickSight, Looker Studio, Streamlit, Excel (Power Query, Pivot)
ETL & Data Modeling: Airflow, dbt, AWS Glue, DataStage, Dimensional Modeling, Data Quality Frameworks
Cloud & Databases: AWS (Redshift, S3, Glue, Lambda, RDS), GCP (BigQuery), Teradata, PostgreSQL, SQL Server, Oracle, MongoDB
Systems & Infrastructure: Linux/Unix, CI/CD (GitHub Actions, AWS CodePipeline), Terraform
Certifications: AWS Certified Data Engineer – Associate (2025), Databricks Certified Data Engineer – Associate (2025)