

# GAURI JAYDEEP PATOLE

Boston, MA • (857)308 6256 • [patole.g@northeastern.edu](mailto:patole.g@northeastern.edu) • [LinkedIn](#) • [GitHub](#) • [Portfolio](#)

## EDUCATION

### Northeastern University

*Master of Science in Information Systems, ML Engineering & Data Science*

**Sep 2022 - May 2024**

*Boston, MA*

### University of Mumbai

*Master's in Mathematics & Bachelor's in Mathematics*

**Sep 2014 – May 2019**

*Mumbai, India*

## PROFESSIONAL EXPERIENCE

### Data Analyst

*Freelance*

**Sep 2021 - July 2022**

*Mumbai, India*

- Implemented a scalable AWS Big Data pipeline with NiFi for Covid-19 data ingestion, transforming JSON to CSV.
- Utilized PySpark for data processing, created Hive external tables, deployed to a data lake, and visualized with Tableau.
- Developed Batch and Real-time analytical pipelines using an eCommerce dataset to simulate user interactions.
- Used Kinesis, Glue, S3, and QuickSight for Batch processing, and Lambda, CloudWatch, and SNS for DDoS detection.

### International Baccalaureate Math Tutor

*Freelance*

**July 2019 - May 2021**

*Mumbai, India*

- Mentored over 100 students through mathematical concepts for International Baccalaureate, IGCSE and MYP curricula.
- Taught advanced mathematical subjects including AP Calculus, probability, and statistics, SAT and ACT ensuring comprehensive understanding and application of concepts.

## CERTIFICATIONS

### AWS Certified Solutions Architect – Associate

- Validation Number: QVLFJC7K8MQ415CP

**Oct 2023 – Oct 2026**

## ACADEMIC PROJECTS

### Real-Time Twitter Feed Dashboard

**Apr 2024 - May 2024**

- Used API, Azure Storage, Azure Queue, Snowpipe, and Snowflake for near real-time Twitter feed ingestion.
- Implemented Snowpipe with Azure IAM for seamless data transfer from Azure Blob Storage to Snowflake.
- Created Snowflake dashboards with 30-minute refresh intervals for Twitter feed metrics and sentiment analysis.
- Leveraged Azure Queue Storage for scalable and efficient data ingestion and workflow management.

### Power BI Spotify Dashboard

**Sep 2022 - Dec 2023**

- Conducted analysis of Spotify's top 2023 songs using Python, DAX enriching the dataset with streaming stats.
- Integrated DENEb to create dynamic visualizations, heatmaps for pattern identification, for deeper data insights.
- Implemented HTML for dynamic display of album covers, enhancing the visual appeal and interactivity of data.

### NYC Motor Vehicle Collision Analysis

**Sep 2022 - Dec 2023**

- Structured 2.05 million rows of raw data to clean, scale, and leverage SQL and Tableau answering business questions.
- Designed a model using ER studio with 10+ facts and dimensional entities, overcoming discrepancies within the data.
- Created ETL pipelines using Alteryx and Airflow, for 40% reduction in data load times for Azure, MySQL, and Postgres.
- Extracted data from Google Cloud- BigQuery and processed in Talend, resulting in a 30% reduction in loading time.

## TECHNICAL SKILLS

- Languages: Python, R, SQL, Java (SprigBoot), HTML, C++, Spark.
- Tools & Databases: Alteryx, Talend, Tableau, Power BI, Snowflake, Airflow, TensorFlow, PyTorch, Machine Learning.