

MALAV PATEL

Data Engineer

malavpatel4217@gmail.com | +1 (774) 528-7940 | github.com/Malav4217 | linkedin.com/in/-malav-patel | patelmalav.com | Edison, NJ

PROFESSIONAL EXPERIENCE

Data Analyst Intern — AtliQ Technologies

Jan 2024 – Apr 2024

Vadodara, India · Tools: SQL, Power BI, DAX

- Conducted in-depth analysis of user behavior and store performance data; identified actionable insights that **boosted sales by 14.21%** and supported fact-based decision-making across business teams
- Evaluated two years of sales data using SQL to surface KPIs including total profit, average basket size, and spending patterns — generating financial recommendations to enhance user experience
- Identified **top 3 employees contributing 27% of total sales**; recommended performance-based incentive model to increase productivity
- Prepared and delivered weekly and final reports, distilling technical insights for non-technical stakeholders

Data Scientist Intern — Muljibhai Patel Urological Hospital

May 2023 – Jun 2023

Nadiad, India · Tools: Python, Keras, NumPy, Pandas

- Conducted genome research under Chief Scientist Sachchida Nand Pandey, focusing on pharmacogenomics to optimize personalized medication dosing for kidney transplant patients
- Processed and analyzed **256 GB of unstructured genomic data**; utilized Manhattan plots to identify genetic variants impacting drug response
- Built ML models (Keras, NumPy, Pandas) to predict optimal dosing for transplant patients, improving medication precision

PROJECTS

Retail Intelligence & Data Platform [github](#)

Feb 2026

Microsoft Fabric · PySpark · Delta Lake · Spark SQL · Power BI Direct Lake · Data Factory

- Engineered end-to-end Medallion Architecture (Bronze → Silver → Gold) in Microsoft Fabric using PySpark, ingesting **676K+ raw records** across 8+ years of legacy retail POS data into a high-fidelity Star Schema
- Implemented Smart Revenue rules to differentiate regulatory fees from promotional discounts, achieving **100% reconciliation** with physical store records; promoted 351,275 rows to Silver layer
- Built fully automated Data Factory pipeline: file-drop trigger → notebooks → Azure Blob archive → cleanup; Power BI Direct Lake delivers sub-second dashboard queries

NYC MTA Bus Reliability Tracker [github](#)

Jan 2026

Python · DuckDB · Streamlit · Docker · pytest · Apache Airflow · MTA Bus Time API

- Built real-time pipeline analyzing **926K+ bus arrivals** across 4 NYC routes; detected ghost buses, schedule delays, and bus bunching with 88% false positive reduction through multi-layer filtering
- Implemented Haversine-based bunching detection and ghost bus forward-tracking algorithm; validated with **40 pytest tests** (76% coverage, 2.66s runtime)
- Deployed via Docker Compose with zero-conflict read replica pattern; Airflow DAG production-ready

Real-Time Traffic Data Analytics [github](#)

Nov 2024

Azure Functions · Event Hub · Blob Storage · KQL · Power BI · BeautifulSoup · Pandas

- Automated pipeline collecting traffic & weather data every 10 minutes from HERE Traffic API — processing **1,440+ data points daily** with zero manual effort
- Streamed data through Azure Event Hub into KQL Database; built interactive Power BI dashboard visualizing congestion patterns across New York, Boston, and California

EDUCATION

University of Massachusetts Dartmouth — Dartmouth, MA

May 2026

Master of Science in Computer Science GPA: 3.78

Courses: Advanced Machine Learning, Big Data Analytics, Theoretical Computer Science, Database Design

Charotar University of Science & Technology (CHARUSAT) — India

May 2024

Bachelor of Technology in Information Technology GPA: 9.07/10

Courses: Data Structures & Algorithms, Data Analytics & Visualization, Statistical & Numerical Techniques

TECHNICAL SKILLS

Languages: Python, SQL, C++, Java, C

Data Engineering: PySpark, Delta Lake, Microsoft Fabric, Medallion Architecture, ETL/ELT Design, Apache Spark, Hadoop, Data Factory Pipelines

Cloud & Streaming: Azure Functions, Azure Event Hub, Azure Blob Storage, KQL (Kusto Query Language), AWS, BigQuery

Databases: MySQL, DuckDB, MongoDB, Star Schema Design

Visualization & BI: Power BI, DAX, Power Query, Matplotlib, Seaborn, Plotly

ML & Science: TensorFlow, Keras, Scikit-learn, PyTorch, Pandas, NumPy

Tools: Docker, Git, Jupyter Notebooks, pytest, Apache Airflow

AWARDS & CERTIFICATIONS

Best Data Extraction Accuracy & Best Collaboration and Teamwork — Intelligent Data Discovery Hackathon, Oct 2024

Grand Finale Selection — Azadi ka Amrit Mahotsav Hackathon, Oct 2022

Google Data Analytics Professional Certificate *[verify]*

Machine Learning Specialization — DeepLearning.AI *[verify]*