

Atharva Upadhyay

Data Engineer | Azure Data Integration | ETL/ELT | Advanced SQL | Python

atharva.upadhyay2144@gmail.com | +91 88275-18047 | linkedin.com/in/atharva-upadhyay | github.com/AtharvaUpadhyay

Professional Summary

Data Engineer experienced in building production ETL and data-integration workflows for large UK property, planning, title, lease, address, and geospatial datasets. Strong hands-on background in Python, advanced PostgreSQL/PostGIS, Azure Blob Storage, API ingestion, data quality, schema design, performance troubleshooting, and analytics-ready data delivery.

Experience

Data Engineer, Nimbus Property Systems Ltd, Warwick, UK

Jan 2025 – Aug 2026

- Engineered and maintained Python and SQL ETL/ELT pipelines integrating planning applications, Land Registry/title data, UPRNs, leases, address datasets, APIs, WFS/WMS services, ArcGIS endpoints, files, and Azure Blob Storage into PostgreSQL/PostGIS.
- Developed advanced SQL transformations using complex joins, CTEs, spatial predicates, conditional matching, deduplication, schema-aware updates, and geometry validation to support production analytics and business-critical property datasets.
- Built an Azure Blob Storage extraction workflow for **5,288 HTML deal reports**, parsing semi-structured documents into **27 standardized commercial-property fields** and loading structured outputs into PostgreSQL for downstream analysis.
- Automated a high-volume Scottish title-shape pipeline for **11,000+ records** using Python, OpenCV, PostGIS, WMS imagery, and EPSG:27700 processing; applied area, topology, colour, contour, duplicate, and failure-validation rules before database insertion.
- Designed planning-title enrichment logic that reused validated existing titles first, then generated controlled postcode/UPRN-based fallback titles and geometries only when source relationships and spatial data passed validation checks.
- Implemented lease/title reconciliation and duplicate-handling workflows using TitleId, UPRN, date-window clustering, rent/floor-area fallbacks, and controlled dummy-title creation to preserve downstream referential consistency.
- Built repeatable geospatial ingestion workflows across ArcGIS FeatureServer, WFS/WMS, Cadcorp/GeognoSIS, and MapInfo services, standardizing coordinate systems and producing validated MultiPolygon outputs for downstream spatial analysis.
- Diagnosed production data issues including UUID/type mismatches, invalid or collection geometries, duplicate records, missing relationships, failed image extraction, and slow SQL paths; used processing logs and QA tables to refine rules and improve reliability.

Data Analyst Intern, Careervira Education Private Limited

Apr 2024 – Aug 2024

- Built batch ingestion pipelines from **50+ data sources**, including Azure Blob Storage, to support scalable data integration and processing workflows.
- Developed SQL-based transformations with partitioning, validation rules, data-quality checks, and anomaly detection; automated recurring workflows to reduce manual processing time by **40%**.

Research Analyst, Essmart Global

Jan 2024 – Mar 2024

- Developed predictive models on **100K+ records**, applying statistical modeling and feature engineering to improve targeting accuracy by **15%**.
- Built scalable data-processing workflows and analytical reports using Python, NumPy, and Power BI to support research and decision-making.

Skills

Azure & Data Integration: Azure Storage Accounts, Azure Blob Storage, Azure Storage Blob SDK, ETL/ELT, batch pipelines, API ingestion, semi-structured data processing

Advanced SQL & Data Platforms: PostgreSQL, PostGIS, complex joins, CTEs, query optimization, schema design, relational modeling, spatial SQL, data reconciliation

Python & Processing: Python 3.12, NumPy, psycpg2, SQLAlchemy, OpenCV, regex parsing, automation, geospatial processing, file/web-service ingestion

Data Quality & Operations: validation rules, anomaly detection, deduplication, referential integrity, topology checks, QA tables, processing logs, production troubleshooting, root-cause analysis

Analytics & Reporting: statistical modeling, feature engineering, Power BI, analytics-ready datasets

Education

Devi Ahilya Vishwavidyalaya, Indore

Nov 2021 – May 2026

M.Tech Dual Degree in Artificial Intelligence and Data Science

CGPA: 8.0 / 10