

Ahmed Ikram | Data Engineering & Analytics

Dundee, Scotland | +44 7421 996048 | ahmedikram30@gmail.com | [LinkedIn](#) | [GitHub](#) | [Online Portfolio](#)

PROFILE

3rd-year Computer Science student at the University of Dundee on track for a First-Class degree, seeking a Data Engineering/Analytics placement or internship. Focused on building resilient data pipelines, scalable ETL workflows, and analytics-ready data models on AWS. Experienced in handling messy, real-world data through robust ingestion, validation, and orchestration - turning raw data into usable insight.

EDUCATION

High School of Dundee

Aug 2017 - June 2023

- National 5 Computer Science, Maths, English, Economics, Physics, Chemistry, Biology (A)
- Higher Computer Science, Maths and English (A), Higher Chemistry (B) and Higher Physics (C)
- Advanced Higher Computer Science (C)

University of Dundee - BSc (Hons) Computer Science (Data Science and AI)

Sep 2023 - Present

- Expected Graduation: June 2027 (with year-long placement, June 2028)

PROJECTS

ATM Log Aggregation, Analysis & Diagnostics Platform | Python, FastAPI, SQLite, APScheduler, React, RAG

- Led backend/data engineering for a 7-person team, delivering a production grade system for NCR Atleos
- Designed a 7-table data model and a resilient ingestion pipeline with dead-letter routing for schema drift and a retry-with-backoff mechanism handling concurrent writes
- Built a local RAG-based diagnostic assistant (LangChain, ChromaDB, Ollama) enabling natural language querying of ATM log data; implemented semantic chunking via LangChain's SemanticChunker with nomic-embed-text embeddings, inference served locally by llama3.1:8b - architected as air-gapped so no log data leaves the local network, aligned with banking-grade data privacy requirements
- Worked from a formal requirements specification - translating 12 functional and 8 non-functional requirements and user stories into technical design decisions

W3C Web Logs ETL Pipeline | Apache Airflow, Python, Power BI, Power Automate, AWS RDS

- Designed an Airflow DAG orchestrating incremental ETL of over 90 raw web logs totalling 156,000 log records into a 9-dimension Star Schema on AWS RDS PostgreSQL - nine dimension tasks fan out simultaneously post-staging, with the fact table load fanning back in once all nine complete, making phase three 8x faster
- Engineered ingestion handling two distinct IIS column formats (14 and 18 columns) detected dynamically from hash fields headers; incremental filename deduplication makes reruns fully safe
- Resolved data quality issues during transformation: IP-to-geolocation enrichment via ip-api.com batch API (100 IPs/request) with rate-limit-aware 1.5s pausing and AWS RDS reconnection with 3-retry logic after idle drops during long API phases; left joins with -1 surrogate key fallback throughout ensure zero records are silently dropped from the fact table
- Built Power BI reports across 7 pages surfacing traffic trends, P95 response times via DAX, geographic distribution across 78 countries, and bot vs. human traffic split (38% crawler); automated dataset refresh via Power Automate with a failure-handling flow sending success/failure email notifications

Haggis Data Mining & Predictive Modelling Notebook | Python, Scikit-learn, pandas, matplotlib, Jupyter Notebook

- Ran a complete ML pipeline from EDA through feature engineering to model evaluation; engineered and validated 2 novel ratio-based features using sklearn pipelines (ColumnTransformer, RobustScaler) that ranked among the top-3 predictors across all models, whilst preventing data leakage
- Trained and compared 7 classifiers (XGBoost, Random Forest, Logistic Regression and others) achieving ~90% accuracy and tuned hyperparameters via GridSearchCV with 5-fold cross-validation
- Applied K-Means and DBSCAN clustering and built a Linear Regression model ($R^2=0.756$) and validated assumptions via residual diagnostics and standardised coefficient analysis

DevSync Project Tracker with GitHub Integration | React, Flask, Socket.io, GitHub API, Pytest, AWS, Docker

- Designed a PostgreSQL schema on AWS RDS managing complex relationships across users, projects, and GitHub entities; enforced data integrity via SQLAlchemy with auto-migrations on every deployment
- Architected production AWS infrastructure: ECS Fargate behind an ALB, RDS isolated in a private subnet with least-privilege security groups ensuring zero public database exposure, React frontend on S3 + CloudFront
- Enforced quality gates via GitHub Actions CI/CD: 271 Pytest tests at 80% line coverage and 270 Jest tests at 90% line coverage gate every PR, with OIDC federation eliminating static AWS credentials

PROJECTS

Global Superstore Business Intelligence & KPI Interactive Dashboard | Excel

- Analysed 51,000 sales records across regions, categories, and time periods; identified that aggressive discount strategies were directly suppressing net profit in the Furniture segment and surfaced this as one of many actionable KPI recommendations
- Built an interactive slicer-driven dashboard enabling non-technical stakeholders to independently explore the data without requiring analytical support; presented findings as actionable business recommendations and a recorded walkthrough

TECHNICAL SKILLS

Languages	Python, SQL, JavaScript, Bash, Typescript
Data Engineering	Apache Airflow, ETL/ELT, Data Modelling (Star Schema), Pipeline Design
Databases	PostgreSQL, MongoDB, MySQL, SQLite, NoSQL
Analytics & ML	Scikit-learn, PyTorch, Pandas, NumPy, Matplotlib, Power BI
Cloud & DevOps	AWS (EC2, S3, RDS, IAM, SageMaker), Docker, GitHub Actions, Linux, Azure
Tools	Git, Jupyter Notebook, REST APIs

CERTIFICATES

<u>AWS Academy Graduate - Machine Learning Foundations</u>	Feb 2025
<u>AWS Academy Graduate - Cloud Foundations</u>	Mar 2024
<u>Foundations of Azure AI: Concepts, Capabilities, and Implementation</u>	Oct 2024

EXPERIENCE

Beaulieu Mini Market, Dundee, Scotland

Retail Assistant

September 2023 - Present

- Diagnosed and resolved Point-of-Sale and stock system failures rapidly to minimise service disruption

Zulus Peri Peri Grill, Dundee, Scotland

Front of House Worker

June 2021 - September 2023

- Trained 5 new staff members; managed competing priorities across dine-in, takeaway, and delivery orders under sustained pressure

TRANSFERRABLE SKILLS

Pipeline Design & Data Modelling

Design normalised schemas and ingestion architectures with reliability, schema drift tolerance, and edge cases as first-class concerns

Data Quality & Validation

Apply multi-stage validation - row count assertions, null checks, surrogate key fallbacks - to ensure no records are silently dropped or corrupted downstream

Root Cause Analysis

Systematically isolate pipeline failures and data issues; experienced diagnosing silent data loss, concurrency bugs, and API rate-limit failures under time pressure

ETL/ELT Orchestration

Translate raw, messy data sources into reliable analytical foundations using orchestrated, scheduled, and idempotent workflows

Monitoring & Troubleshooting

Build operational awareness into pipelines, failure notifications, retry logic, and dead-letter routing, so no failure goes unnoticed

Stakeholder Communication

Translate pipeline outputs and analytical findings into plain business language for non-technical audiences and project stakeholders

Cross-functional Collaboration

Led backend and data engineering workstreams in a 7-person Agile team; experienced coordinating across frontend, ML, and analysis sub-teams

Requirements, Documentation & Detail

Translate functional and non-functional requirements into schema contracts, API specifications, and data retention policies; document systems to a standard others can maintain and verify against formal acceptance criteria