MATTHEW PAVER

LinkedIn | GitHub

EMAIL: mattpaver@outlook.com

MOBILE: +44 7771 672205

PROFILE

Experienced data and technology specialist with a background in analytics, modelling and solution design. Over six years of experience developing tools and systems that enhance forecasting, reporting and decision making. Skilled in Python, data engineering and machine learning, with hands-on experience in generative and graph-based applications. Confident working directly with clients across defence, engineering and construction sectors, combining technical depth with clear communication and problem-solving skills.

EDUCATION

BSc (Hons) Computer Science with Artificial Intelligence — University of Liverpool

Level 4 Project Data Analyst Apprenticeship — Projecting Success

SKILLS SUMMARY

Languages: Python, SQL, C, C++, Java, CSS, HTML, Cypher

Data Science & Engineering: LLMs, RAG, Prompt Engineering, Embeddings, GANs, LangChain, GraphRag

Frameworks: Pandas, Numpy, Scikit-Learn, Matplotlib, Node.js, Tensorflow,

Tools: Power BI, MySQL, Azure, Databricks, Neo4j

Platforms: Visual Studio Code, Jupyter Notebook

WORK EXPERIENCE

Solutions Architect | Projecting Success | Jul 2025 – October 2025

- Helped design and support the ongoing development of Project Brain, a knowledge graph platform designed to connect complex data documents and insights to use case visibility and planning.
- Contributed to the design and testing of the data pipeline and ontology used to structure information in Neo4j.
- Assisted with documentation and preparation for integration with internal tools such as Marvin.
- **Data Protection Officer (DPO)**: Led the organisation through Cyber Essentials and Cyber Essentials Plus certifications and supported ISO27001 compliance.
- Worked with stakeholders to translate project requirements into practical, compliant data solutions.
- Supported the design and rollout of intelligent data frameworks that streamline automation and integrate predictive models across client workflows in defence, engineering and construction.

Data Scientist | Projecting Success | Jun 2022 – Jul 2025

- Developed ML models (ARIMA, SARIMA, GANs) for predictive schedule stability, change-point detection and milestone risk analysis.
- Implemented forecasting pipelines in Python and Databricks to improve accuracy of project outcome predictions.
- Delivered proof-of-concept models for engineering clients, including Rolls-Royce and Baker Hughes, helping demonstrate the value of predictive analytics in project delivery.
- Produced research and technical summaries that informed decisions on high-value client projects.
- Worked closely with stakeholders to define analytical requirements, explain results and ensure outputs were actionable and transparent.

Data Analyst | Projecting Success | Sept 2019– Jun 2022

- Designed and delivered Power BI dashboards to track KPIs, risks and project performance metrics, reducing manual reporting time by 70%.
- Automated reporting processes and data validation tasks using Python to improve consistency and speed.
- Provided technical support and training to more than 200 apprentices, helping teams build stronger data literacy and reporting capability.
- Contributed to Cyber Essentials and ISO27001 compliance activities, improving data governance and process documentation.
- Supported cross-sector projects in defence, engineering and utilities through analysis that improved visibility or project performance and risk.

PROJECTS

Machine Learning & AI Projects

Happening.London (2025-Ongoing)

A data-driven event discovery platform built to improve access to London's cultural and entertainment events.

- Designed and deployed Scrapy spiders and MCP-based crawlers to collect and standardise event listings from sources such as TimeOut, Eventbrite and venue websites.
- Implemented data ingestion pipelines to clean, tag and structure event data for downstream retrieval and analytics.
- Integrated a local LLM via LM Studio to summarise and categorise events by theme and content.
- Developed embedding and vector search pipelines to support personalised event recommendations and thematic discovery.

Marketing ML Lakehouse (2025) | LINK

An end-to-end local data lakehouse for marketing analytics with ML-driven insights.

- Built a DuckDB-based lakehouse architecture implementing bronze→silver→gold data transformation pipelines using pandas and custom ETL workflows.
- Developed XGBoost models for conversion prediction and campaign pacing optimisation, improving forecast accuracy through temporal feature engineering.
- Created an interactive Streamlit dashboard for real-time analytics, enabling stakeholders to explore campaign performance, conversion trends, and ML-driven recommendations.
- Integrated LM Studio for automated insight generation, using local LLMs to summarise performance patterns and generate actionable recommendations from structured data.

Project Brain (2025)

A Neo4j knowledge-graph platform for connecting risks, actions and insights across projects.

- Developed the design of the ontology and core data model in Neo4j, combining structured and unstructured project information.
- Built ingestion and processing workflows that converted documents, notes and artefacts into linked nodes and relationships.
- Implemented semantic indexing and vector search to improve retrieval across long-form project material.
- Integrated with the Marvin assistant to enable natural-language querying over the knowledge graph.
- Contributed to documentation, testing and validation during rollout across defence and engineering clients.

Marvin | GraphRAG Knowledge Graph Assistant (2024-2025)

An assistant that applies retrieval-augmented generation over graph data for project and hackathon knowledge search.

- Contributed to the design of LangChain and GraphRAG pipelines, enabling LLMs to query Neo4j knowledge graphs.
- Supported the creation of Azure ingestion pipelines to extract and summarise challenge and dataset documents into JSON artefacts.
- Assisted with prompt tuning and indexing to improve entity recognition and graph accuracy.
- Integrated Pinecone for semantic retrieval, providing context-aware responses to user queries.

Data Engineering & Analytics Projects

PySpark & Kafka Integration (2025) | LINK

Real-time data streaming and processing pipelines using PySpark and Apache Kafka.

- Built streaming data pipelines consuming from Kafka topics, processing high-volume event streams in real-time using PySpark DataFrames.
- Implemented windowed aggregations and stateful processing for real-time analytics, demonstrating scalable big data processing patterns.
- Designed fault-tolerant streaming architectures with checkpointing and error handling for production-ready data pipelines.

Predictive Schedule Stability & Risk Management (2024-2025) | LINK

A forecasting initiative to identify instability and risk in project schedules using machine learning.

- Developed time-series models (ARIMA, SARIMA) and change-point detection algorithms to identify potential slippage and performance trends.
- Built end-to-end pipelines in Python and Databricks for data preparation, feature engineering, and model training.
- Created a Flask web application enabling teams to upload project data and trigger automated risk analysis workflows.
- Produced Power BI dashboards used by delivery teams to anticipate high-risk milestones, improving forecast reliability and proactive risk management across multi-million-pound projects.

More Projects: Additional ML, NLP, recommendation, analytics, and data-engineering projects available on GitHub HERE

CERTIFICATES

- **Agent Course (Hugging Face)** | CERTIFICATE — November 2025
- **Safeguarding and Safer Recruitment in FE and Training (FutureLearn)** — April 2025
- **Safeguarding and Safer Recruitment in FE and Training (FutureLearn)** | CERTIFICATE — July 2023
- **Leadership and Coaching in Data Analytics (Projecting Success)** | CERTIFICATE — September 2022
- **Data Analysis Concepts (BCS)** | CERTIFICATE — January 2021
- **Certified Neo4j Professional (Neo4j)** | CERTIFICATE — December 2020
- **Data Analysis Tools (BCS)** | CERTIFICATE — December 2020
- **RPA Developer Advanced (UIPath)** | CERTIFICATE — September 2019