

# SAMUEL LAWRENCE BROOKS

London, UK | +44 7771 939292 | +1 832 335 8189 | slbrooks20@gmail.com | [LinkedIn](#) | [samuellbrooks.com](http://samuellbrooks.com)  
British Citizen – Eligible to work in the UK with no restrictions

## PROFILE

---

MSc Business Analytics student at University College London with a BBA in Management Information Systems. Hands-on experience building end-to-end data pipelines and warehouses across the modern data stack (dbt, Airflow, Kafka, Spark, BigQuery, AWS, and Terraform) developed through an industry consulting engagement, substantial academic projects, and self-directed work. Professional background in IT governance, data risk, and compliance at a global engineering firm. Bilingual in English and Spanish. Seeking junior data engineer, data analyst, or BI roles.

## EDUCATION

---

**University College London (UCL)** – London, UK Sep 2025 – Sep 2026  
*Master of Science in Business Analytics*

- Modules: Data Engineering, Predictive Analytics, Machine Learning for Business, Programming for Business Analytics (Python)
- Industry project: Data engineering consulting engagement with Guinea Mobile SAC

**C.T. Bauer College of Business, University of Houston** – Houston, TX, USA Aug 2019 – Dec 2024  
*Bachelor of Business Administration in Management Information Systems — Cum Laude (equiv. upper 2:1)*

**NEOMA Business School** – Rouen, France Sep – Dec 2023  
*Study Abroad Programme (University of Houston)*

## PROFESSIONAL EXPERIENCE

---

**Guinea Mobile SAC** – Remote May – Aug 2026  
*Data Engineering Intern (UCL Dissertation Project)*

- Designed and built a reproducible Python pipeline that transforms raw telecom billing system exports into a structured, analytics-ready dataset for a South American MVNO.
- Designed a normalised data model with documented lineage, loading into DuckDB, and ensured data integrity across a full month of daily subscriber snapshots.
- Analysed subscriber growth, product adoption, and package usage patterns, delivering reproducible notebooks and visualisations to company and academic stakeholders across a 15-week engagement.

**McDermott International, Ltd** – Houston, TX, USA May – Sep 2025  
*IT Intern*

- Analysed data from 165+ global file servers using Power BI to visualise trends and identify redundant/obsolete data, contributing to cost optimisation and improved data governance practices.
- Assessed potential vendors for Microsoft Purview implementation, AI compliance training, and data privacy mapping using the company's partner assessment framework. Conducted cybersecurity risk analysis on AI platforms and presented findings to the AI Council.
- Collaborated with cross-functional teams in India and the US on a File Share-to-Azure migration, supporting the creation of an Azure data lake for real-time data visualisation.
- Produced AI-driven compliance training videos and enhanced governance by restructuring SharePoint sites, adding FAQs, and optimising content for clarity and usability.

## KEY SKILLS

---

- **Languages & Querying:** Python, SQL, R, Cypher, Bash
- **Data Engineering & ELT:** dbt, Apache Airflow, Apache Kafka, Spark (PySpark, Structured Streaming), Docker Compose, Parquet, dimensional modelling (Kimball star schema), data lake architecture
- **Cloud & DevOps:** AWS (S3, Lambda, Athena, Glue, EventBridge, IAM, CloudWatch), Azure (Data Lake), Google Cloud (BigQuery), Terraform, GitHub Actions CI/CD, Git
- **Databases:** PostgreSQL, MongoDB, Neo4j, BigQuery, DuckDB, ChromaDB, Oracle
- **Machine Learning & Analytics:** scikit-learn, LightGBM, pandas, NumPy, Bayesian regression, walk-forward validation
- **NLP & AI:** LangChain, spaCy NER, sentence-transformers, RAG pipelines, Anthropic & Gemini APIs
- **Visualisation & BI:** Power BI, Tableau, Microsoft Excel, Plotly, Matplotlib, Seaborn, Streamlit, FastAPI, Leaflet.js

## ACADEMIC PROJECTS

---

### Geopolitical Crisis Narrative Tracker

Mar – Apr 2026

*MSIN0166 Data Engineering – UCL*

- Designed and built a fully containerised system (7 Docker services: PostgreSQL, MongoDB, Neo4j, ChromaDB, FastAPI, Airflow, Marquez) ingesting global news from 19 outlets across 8 geopolitical regions via GDELT API, RSS feeds, and web scraping.
- Implemented NLP processing pipeline: spaCy NER, entity resolution (rapidfuzz), 384-dimensional embeddings (sentence-transformers), and semantic search via ChromaDB with metadata filtering.
- Built a 7-tool LangChain RAG agent with hallucination checking, served via FastAPI with an interactive Leaflet.js conflict map. Full data lineage tracked with OpenLineage/Marquez.

### Premier League Match Outcome Prediction

Jan – Mar 2026

*MSIN0097 Predictive Analytics – UCL*

- Built a multiclass classification pipeline (Logistic Regression with temporal decay, LightGBM, MLP) with walk-forward expanding window validation over 4,800+ out-of-sample matches. Bootstrap significance testing confirmed results ( $p = 0.007$ ).
- Developed a probability blending method combining model predictions with bookmaker implied probabilities, outperforming standalone models on log-loss. Extended with a Dixon-Coles Monte Carlo season simulator (10,000 simulations).

### The Indie Game Survival Guide

Jan – Mar 2026

*MSIN0166 Data Engineering – UCL*

- Built an end-to-end PySpark pipeline ingesting data from 6 sources (Steam API, SteamSpy, Reddit, IGDB, Google Trends, Steam Charts via Selenium) into PostgreSQL, MongoDB, DuckDB, and Parquet. Led Spark pipeline development, debugging, and data visualisation.

## PERSONAL PROJECTS

---

### Competitive Pokémon Meta Warehouse

Apr – May 2026

*dbt – BigQuery – Airflow – Python*

- Building a modern ELT data warehouse in BigQuery with dbt, ingesting data from PokéAPI and monthly Smogon usage statistics into a Kimball-style star schema (fact + dimension tables across staging, intermediate, and mart layers).
- Implementing dbt tests (uniqueness, not null, accepted values, referential integrity) and auto-generated documentation with lineage graphs; orchestrating monthly refresh runs via Apache Airflow.

### Live Football Match Event Tracker

May 2026

*Apache Kafka – Spark Structured Streaming – PostgreSQL – Docker*

- Developing a real-time streaming pipeline that ingests live Premier League match events via API into Apache Kafka, processes them through Spark Structured Streaming with watermarked windows, and sinks enriched events to PostgreSQL.
- Joining live scores against pre-match predictions from the Predictive Analytics project to produce a Streamlit dashboard tracking model accuracy and live league-table recalculations in real time.

### Cloud-Native Stock Portfolio Tracker

May – Jun 2026

*AWS – Terraform – GitHub Actions – Python*

- Designing an AWS data lake (S3 raw → processed → curated layers) with event-driven Lambda ingestion from the Finnhub API, JSON-to-Parquet transformation, and Athena SQL querying over Glue Catalog tables.
- Defining all infrastructure as code in Terraform and deploying via a GitHub Actions CI/CD pipeline (lint → validate → plan → apply), with least-privilege IAM roles and CloudWatch monitoring.

## ADDITIONAL EXPERIENCE & INFORMATION

---

### Various Restaurants – Houston, TX, USA

Jun 2017 – Dec 2024

*Customer Service Roles*

- Delivered high-quality service in fast-paced environments over 7+ years, developing strong communication, problem-solving, and teamwork skills. Trained and mentored new staff, improving team efficiency and service standards

**Languages:** English (Native), Spanish (Native)

**Work Eligibility:** Eligible to work in the UK and USA with no restrictions.

**References:** Available on request