JAMES LEWIS

Data Scientist

James066lewis@gmail.com | linkedin.com/in/james-lewis3/ | github.com/KetchupJL

Data Scientist (MSc Data Science & Statistics, Distinction) with strong expertise in statistics, machine learning, and applied data engineering. Experienced in designing end-to-end data solutions, from automated pipelines and feature engineering to predictive modelling and deployment. Proven track record in applied ML, fintech, and blockchain analytics, with a focus on balancing statistical rigour and scalability to deliver business impact. Skilled in Python, SQL, TypeScript, and modern data science tooling.

EDUCATION

University of Exeter | MSc Data Science and Statistics (Distinction, 78% avg)| 2024-25

- Key methods: supervised & unsupervised learning, Bayesian inference, time-series forecasting, hierarchical & spatial–temporal modelling.

Loughborough University | BSc (Hons) Economics | 2019-23

SKILLS & TECH STACK

Languages & Tools: Python | R | SQL | Git | Bash

Data Science & ML: Feature Engineering | EDA |Data Cleaning | Model Selection | Supervised & Unsupervised Learning | Cross-Validation | Time Series Analysis | Statistical Modelling | A/B Testing

Libraries: scikit-learn | LightGBM | XGBoost | CatBoost | statsmodels | TensorFlow | PyTorch | Keras

Databases & Cloud: PostgreSQL, BigQuery, Redshift, SQLite, Google Cloud Platform (BigQuery, Cloud Functions)

Visualisation: ggplot2 | Plotly | Tableau | Seaborn | Matplotlib | Dash | Streamlit

RELEVANT EXPERIENCE

Independent Quantitative Researcher & Systems Developer| 2023–Present

- Designed and deployed automated data pipelines (Python, TypeScript, SQL) to support real-time ML models.
- Developed and validated models achieving 15–20% simulated risk-adjusted ROI, demonstrating the ability to translate advanced ML into measurable outcomes.
- Delivered low-latency infrastructure with sub-second signal distribution and automated retraining workflows, showcasing scalable, production-ready ML systems.

Hudl | Intern, Revenue Operations | Oct 2023–Apr 2024

- Delivered sales insights by aggregating 30+ product features in SQL, enabling data-driven go-to-market strategy.
- Designed a live intranet dashboard that reduced data retrieval time by 8 hrs/week, improving operational efficiency.

Project Highlights — Full Data Science & ML Portfolio on GitHub

MSc Dissertation: Risk-Adjusted Interval Forecasting of 72-Hour Returns in Mid-Cap Solana Tokens

- Developed an adapted Quantile Regression Forest (QRF) framework that outperformed Linear QR and LightGBM baselines, achieving 0.88 coverage at 90% intervals and lowest tail pinball loss.
- Built and engineered 100+ features from OHLCV, on-chain, and cross-asset data (6,400 observations) to forecast 72h token returns.

Podcast Listening Time Prediction (Kaggle) | Python, LightGBM, Feature Engineering

- Ranked Top 20% globally by engineering time-series and user features and tuning a LightGBM model (RMSE 18.5). Delivered a fully reproducible pipeline for competition submission and post-hoc analysis.

Phytoplankton Disruption from the Deepwater Horizon Spill | R, forecast, tidyr, ggplot2

- Conducted a long-term ecological study of the 2010 oil spill's impact on phytoplankton populations using satellite remote sensing data.
- Applied zonal time-series and anomaly detection techniques with STL decomposition on NetCDF datasets.
- Produced geospatial visualisations and temporal trend maps in R to quantify ecosystem disruption and recovery dynamics.

INTERESTS

Clubs & Leadership: University of Exeter Google Student Developer Club (Advanced Python workshops); Computer Science Society (Hackathons & AI projects); Finance & Investment Society