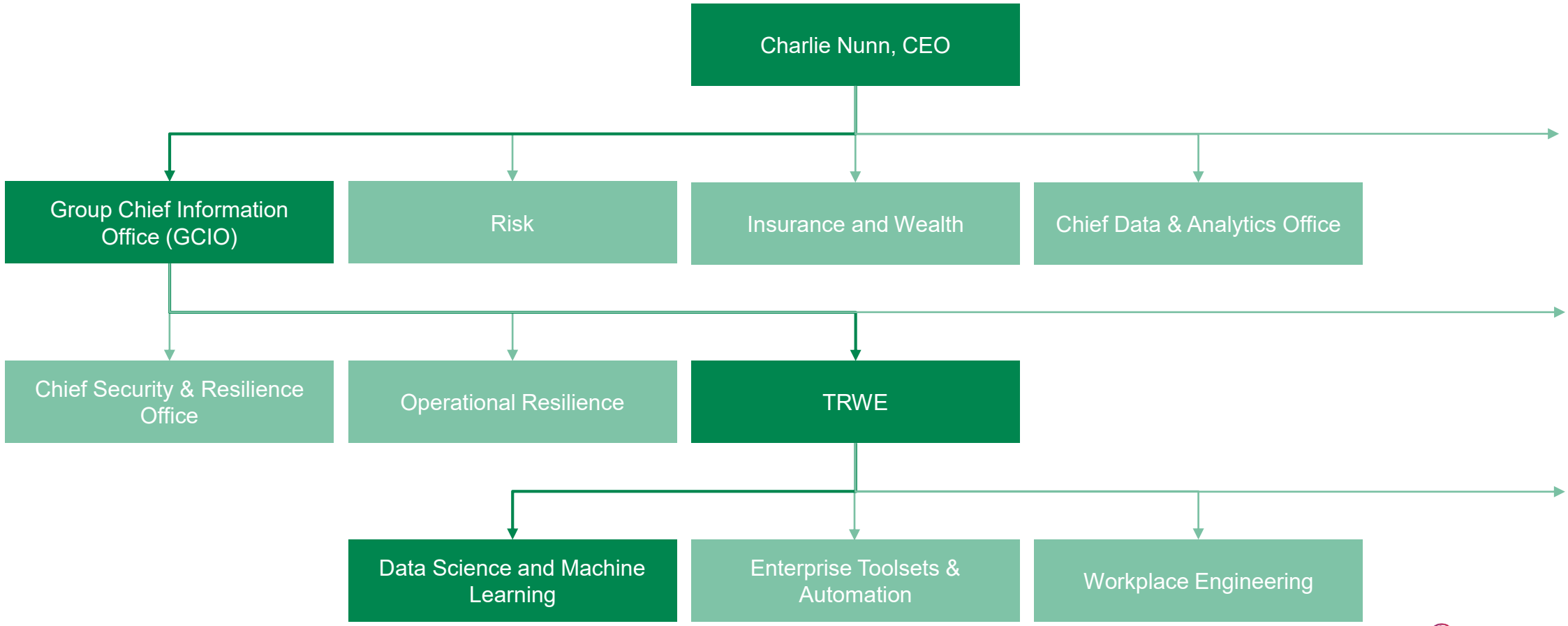
GROUPCIO

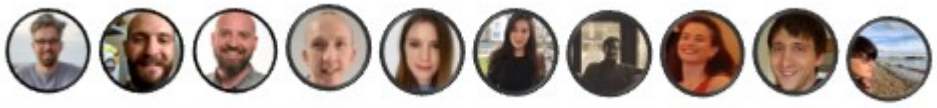
Data Science and Data Engineering for MLOps at Lloyds Banking Group

Gordon Baggott, Katie Adams and Neil Venables
GCIO Data Science and Machine Learning

TECHNOLOGY DATA SCIENCE AT LBG

Where We Sit In Lloyds Banking Group





OUR JOURNEYS INTO DATA SCIENCE

- Data Scientist in LBG for 3 years- customer experience analysis/NLP, analytics, Power BI, optimisation, cloud
- Previously spend 8 years in analytics at LBG in credit and market risk (Commercial Banking)
- Whole career at LBG with various roles in operations, customer service and risk

- Data Engineering in LBG for 3 years – big data platform, distributed computing, metadata management, cybersecurity
- Previously Computational Biologist at STFC Daresbury Lab and Research Software Engineer at University of Manchester
- BSc Computer Science and MSc Computer Security from University of Birmingham
- MSc/PhD Systems Biology from University of Warwick

- Data Scientist in LBG for 10 years – risk modelling, NLP/regulatory docs, facial recognition/fraud detection, cyber security
- Previously Data Scientist at NatWest Group and Pharmacokinetic Modeller for GlaxoSmithKline
- BSc Hons. Pharmacology & Neuroscience University of Dundee
- MSc Epidemiology from University of Edinburgh

DATA SCIENCE AT LBG

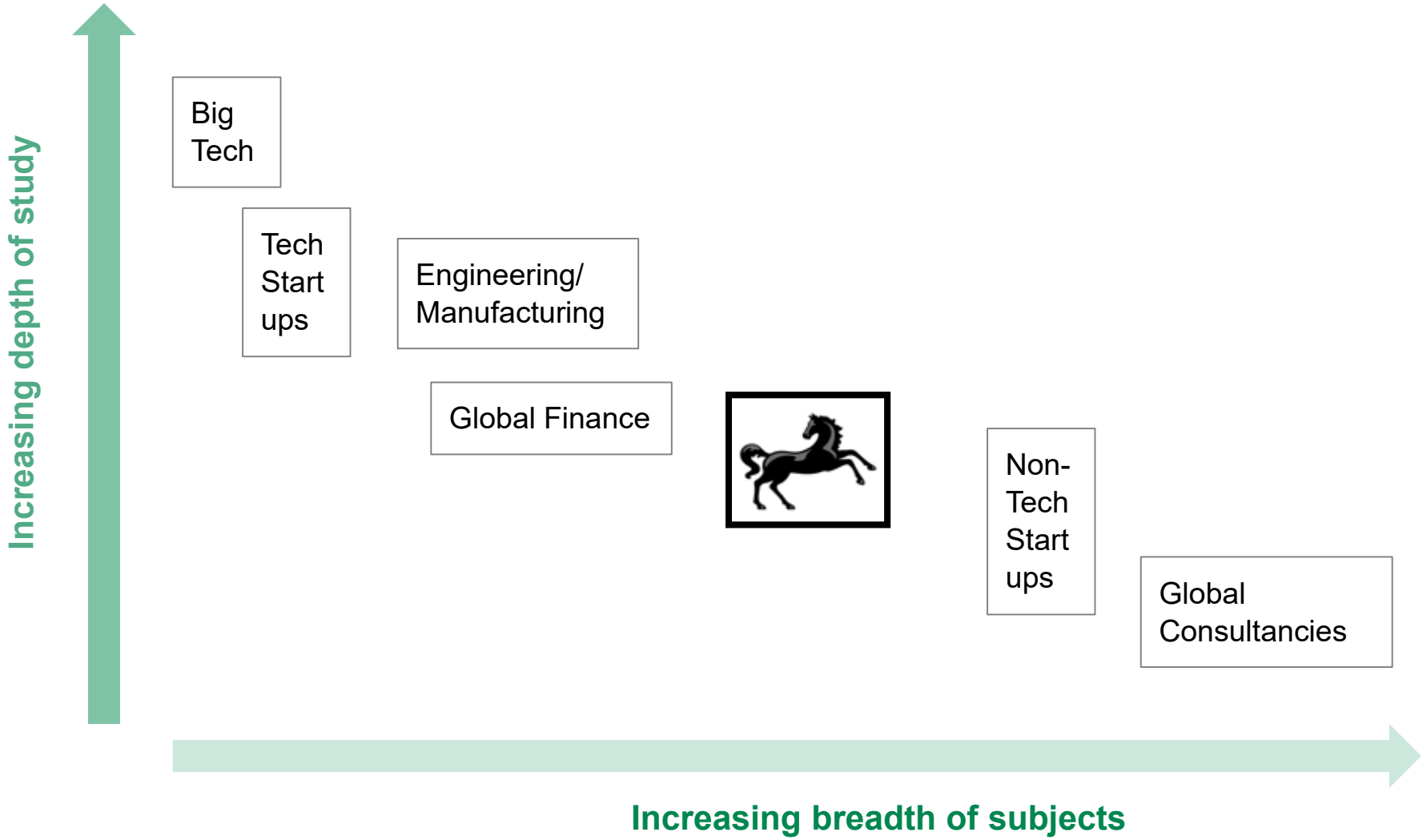
Why Do Data Science At LBG?

- A wide variety of problems
  - Credit Scoring, Fraud, Marketing Optimisation, Engineering, Cyber Security, People & Workplace, Technology Maintenance, Cloud Consumption
- Many large datasets
  - 30m customers; 70k-100k colleagues; 150k+ personal devices; multiple private, public and 3rd party SaaS cloud tenants
- At the forefront of the hardest problems in data science
  - Explainable AI, AI ethics, data privacy, model security, cyber security ML, ML Ops automation

GROUPCIO

DATA SCIENCE AT LBG

General comparison of the DS role at LBG with other organisation types. What do you get in a single role?

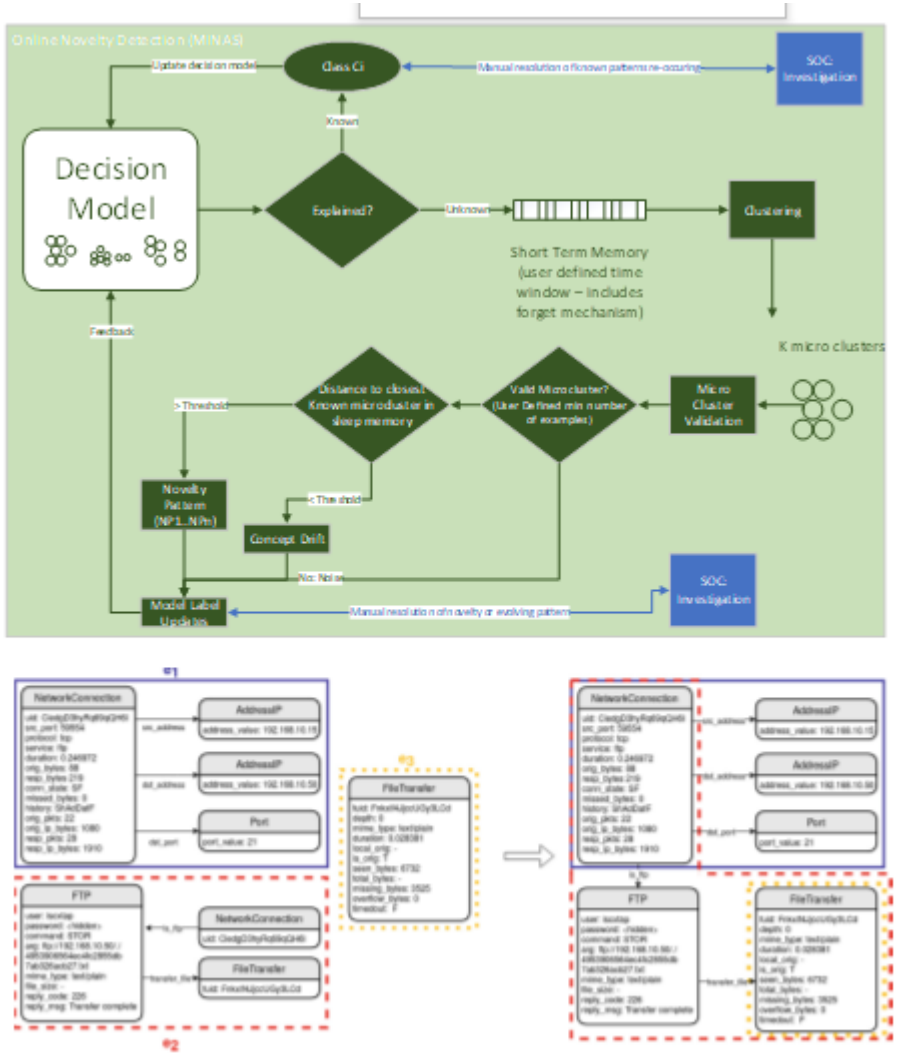


APPLICATIONS OF MACHINE LEARNING

Novelty pattern detection: Harnessing the great unknown in cyber security

Overview

- Cyber threat detection is traditionally done by looking for known, malicious patterns of activity or by identifying extremes of abnormality (anomaly detection) in specific data sources
- To find new or evolved threats and find attacks trying to look normal we must use novelty analysis
- Data and compute intensive, low probability of success, huge benefit if successful.



2-stage modelling process

Filter out known event sequences in streaming network activity logs.

Identify valid patterns (clustering, NN, 1C-SVM), compare to sleep memory of known patterns to identify concept drift or true novel patterns.

Graph feature generation

For data where interactions are key then extracting features from a graph is very powerful.

GROUPCIO

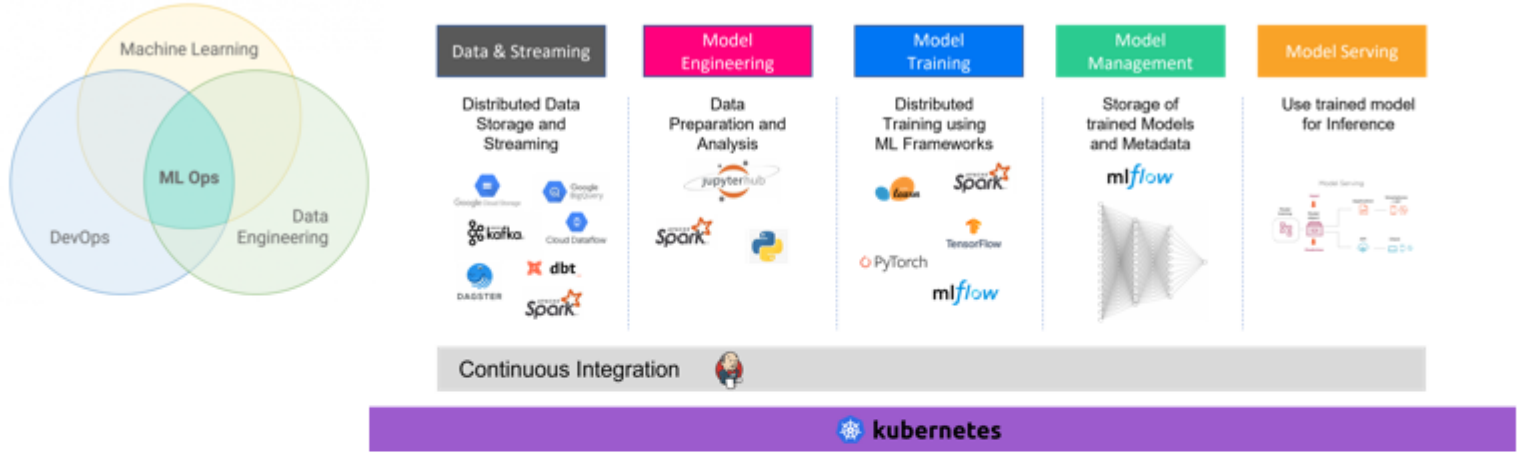
Node level, graph level and overlap metrics.

APPLICATIONS OF ML OPS

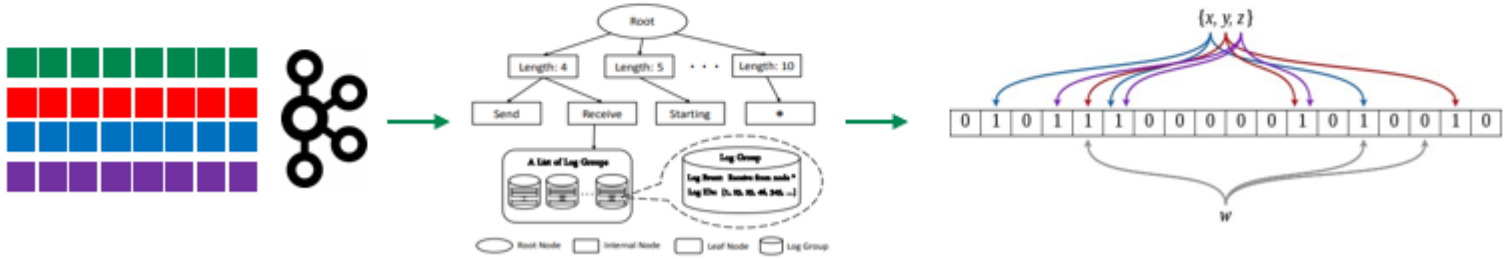
Cyber Security on GCP

Overview

- MLOps is a set of practices that aims to maintain ML models in production.
- The real challenge isn't building an ML model but building a system around its operation.
- We need to develop tools and techniques to reduce this burden and push towards automation.
- Requires skilled people from quantitative fields.



Example: Processing machine logs to count unique visitors.



Stream of logs (3Vs of Big Data)

Automated Log Parsing

Bloom Filter

ANALYTICS FOR TECHNOLOGY

Data Engineering, Analytics, NLP and Data Visualisation of technology data

Overview

- Analytics has a vast amount of uses in financial services, and here we focus on understanding the role technology plays in enabling our colleagues to do their jobs
- Data is brought together from many sources, including Cloud platforms, 3rd party systems & tools and data collected about colleague behaviour and experiences
- Data Engineering and analytical techniques (e.g. NLP) transform the data into insight that drive future focus and investment or support risk management

Example 1: New Technology Roll Out

- Tracking usage of new tech
- Monitoring & alerting- faster response to incidents
- Sentiment analysis & topic modelling for colleague feedback

Example 2: Moving To The Cloud

- Cloud platform APIs provide granular out of the box capability, we combine the data for a high-level view relevant for senior managers
- Monitoring and alerting- performance, stability and process effectiveness

IF YOU ARE INTERESTED….

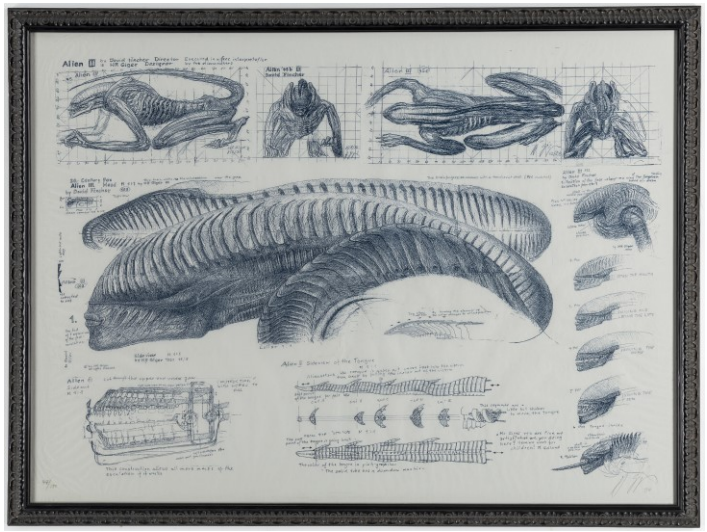
What we need you to show us

Complex problem solving:

- Solve a tough problem elegantly
- Personal or work project
- How did you get to the solution?
- Why?

Show you love to learn more:

- ‘Fast’ R&D
- Build your capability/specialism



Technical brilliance:

- Understand the tools you use
- Why are you different?
- Know when DS is not the answer

HOW TO APPLY

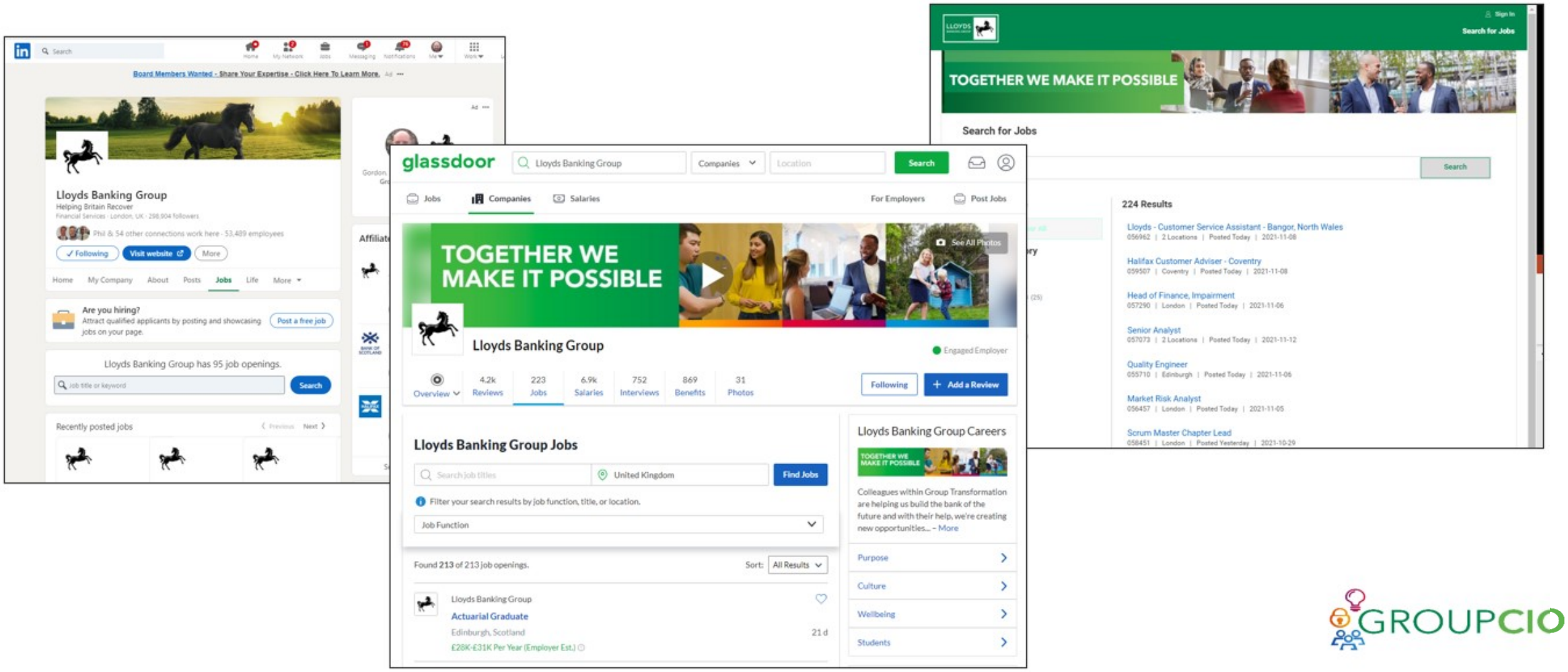
Data Science Graduate Scheme

"Use your coding skills and love for maths to transform the way we use data and tech to solve business problems."

https://www.lloydsbankinggrouptalent.com/graduates/our-graduate-schemes/data-science-graduate-scheme/

HOW TO APPLY

Careers Sites



PAY & BENEFITS

What do you get for a data science job at LBG?

Level	Typical Entry Qualifications	Min Pay (outside London)	Max Pay (outside London)
Applied Scientist/Data Engineer	BSc in quantitative subject	31k	57k
Senior Applied Scientist/Data Engineer	+MSc/PhD or evidence of relevant experience	46k	85k
Lead Applied Scientist/Data Engineer	+Significant practical experience, proven technical expertise and demonstrated deliveries across multiple roles/companies/industries	64k	130k

- Group Profit Share (0-30%)
- Flexible benefit (4%)
- Pension (up to 15%)
- Private Medical
- 30 Days holiday (+Public)
- Agile working
- Health & Wellbeing
- Diversity & Inclusion
- Recognition