



## MARCUS LANNON

Data Engineer

### ABOUT ME

Having read an article on Walmart's use of data analytics to predict skyrocketing Pop-tart sales prior to a hurricane, my eyes were opened to how much data can do. From there, the mathematical and programming skills I learnt at university led me to look further into both data science and big data. As I continued to discover more unique ways in which data has been used, my drive and excitement to work in the industry quickly grew.

Whilst Pop-tarts were my gateway into data, a love for cooking means my diet extends beyond them. I enjoy cuisines from all over the world, with my favourites including Japanese and Mexican.

Outside of work, my passion is cycling, with many weekends spent riding. My greatest achievement on the bike was a 570km, 4-day ride in the Lake District where I climbed some of the steepest roads in the UK.

### CONTACT US

Kubrick Group

T: 020 3866 4620

E: [consultants@kubrickgroup.com](mailto:consultants@kubrickgroup.com)

W: [kubrickgroup.com](http://kubrickgroup.com)

### OVERVIEW

An interest in numbers and problem solving led me to a degree in Mathematics at Durham University, where I graduated with a 1st Class (Hons) in 2018. During the Monte Carlo Modelling module, I was introduced to methods within Python to manipulate large data sets such as NumPy and Matplotlib. We generated large sets of random numbers then ran algorithms to produce and graph probability distributions. These were then applied to increasingly complex queuing models where we observed numbers of 'customers' served and average wait time.

My final year project on 'Simple Lie Algebras and Root Systems' – a complex abstract algebra topic – was written and marked as an introduction to the subject for new readers. This developed my ability to present complicated ideas in clear and understandable ways, both verbally via a presentation and written in a report. I am now focusing on my data engineering at Kubrick, where I look forward to leveraging my skills to solve business challenges.

### WORK

#### DATA ENGINEER – KUBRICK GROUP, LONDON

November 2018 – present

At Kubrick Group I gained knowledge and real-world experience in professional skills, a wide yet in depth array of modern technical and analytic skills. We studied the agile software development lifecycles and gained detailed knowledge of platforms and infrastructure. I specialised in Data Engineering which involved learning and developing advanced skills in Python, Spark, Hadoop, NoSQL and SQL as well as advanced Excel. As a junior data consultant, I applied my skills on real client projects including:

- Sourcing, profiling, enrichment, dimensional modelling and analysis of UK Crime data spanning 10 years. I identified trends and correlations using statistical methods visualised in Tableau. I produced both an interactive Dashboard as well as an Infographic telling story behind the analytics identified.
- For a leading UK insurance syndicate, we produced a robust end to end Python ETL solution to source, engineer and consistently model severity of incidents of fire in UK properties. The data was identified and sourced via a combination of website scraping and API calls. The deliverable of the project was to enable the client to predict frequency and impact of fire in commercial properties given a postcode and building use. The project was completed in a single agile sprint in a team of 18. The code and model were delivered via a git repository complete with full documentation which was then implemented successfully with our client.

### EDUCATION

#### BSC MATHEMATICS, 1ST CLASS (HONS), DURHAM UNIVERSITY

October 2015 - June 2018

Notable modules: Algebra (93), Calculus and Probability (95), Linear Algebra (97), Partial Differential Equations (91), Statistical Analysis (81), Numerical Analysis (85)

### KEY SKILLS

- **Python Libraries:** Pandas, Numpy, SK Learn, NLTK, Plotly
- **Machine Learning Skills:** Statistical Learning, Regression, Classification, Clustering
- **Cloudera Hadoop version 5.4:** HDFS, Yarn, Sqoop, Flume, Avro, Paquet, Impala, Hive and Hue
- **SQL & NoSQL:** MongoDB, Neo4J, Modelling, SSIS, Alteryx, Advanced SQL
- **Apache Spark:** RDD, DataFrames, SparkSQL, PySpark
- **Source Control and IDE:** Git, PyCharm, Visual Studio
- **Other Skills:** Tableau, Linux CentOS 6, SSH, Putty, Agile, LaTeX, Markdown