THOMAS CLEMENTS

Big Data Engineer

ABOUT ME

From a very young age, I have been mathematical minded and enjoyed tackling problems in a logically and analytical way. Nothing is more evident of this than my fascination for playing chess. At the age of 13 I secured a place on the England National U18 where I played across the globe for five years. During this time, I developed a passion for technology and computers where I began creating and developing small robots and first got introduced to computer programming.

At university I chose to take a year out to work in reactor physics in a nuclear power station. Here I was exposed to the power of Excel and data analysis, and I taught myself VBA to further the department's work.

In my final year of university, I continued my data analysis work by deciding to independently work alongside my degree part time analysing transactional data for local businesses.

CONTACT US

Kubrick Group

T: 020 3866 4620

E: consultants@kubrickgroup.com

W: kubrickgroup.com

kubrick
GROWING DATA EXPERTS

OVERVIEW

In 2017 I graduated with a first class MSci Natural Sciences from the University of Bath. While studying Physics and Chemistry simultaneously to a Master's level as well as working part time in my final year, I was continuously flexible and disciplined. Whilst also requiring vast independent study around the course to gain the necessary skills and knowledge in both subjects.

In my project 'Modelling Cardiac Action Potential: The Propagation of Cardiac Action Potentials' I chose to use python to model, analyse and visualise the data. This required self-teaching of python adding extra complexity to the project. The majority of models are complex and very computationally expensive. My model demonstrates that reduced models show the same macro effects. Meaning the results from this model are lightweight enough to have valuable future utility in targeted drug dosing.

WORK

DATA ENGINEER - KUBRICK GROUP - LONDON
September 2017 - Current
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 - these lead me to work towards accreditation as a Cloudera Hadoop CCA Developer.

As a junior data consultant I applied my skills on real client projects including:
- Seasonality correlation to sales for a large retail client.
- Feature analysis for a large financial services client providing HR with churn rate modelling against different employee profiles.

The projects were developed in agile teams using Git and Kanban boards following agile processes in the presence of an agile coach. Delivery of the projects used virtual environments and Docker. The professional skills I have acquired and matured were invaluable in these projects.

BUSINESS INTELLIGENCE ANALYST – CONTRACT, BATH
November 2016 - August 2017
Designing and developing spreadsheet databases and dashboards to analyse trends of business turnover and customer transactions. An example of value added to a business include an eatery changed opening times on certain days due to trends of customer transaction value.
Skills involved: Excel with VBA and Google Sheets with JavaScript.

NUCLEAR SAFETY ENGINEER – EDF ENERGY, HEYSHAM
September 2014 – September 2015
Analysed data and produced reports for technical areas of the station.
Designed and produced spreadsheets to improve efficiency and reduce errors within the department.
Simulated reactor core conditions on UNIX systems.

EDUCATION

MSCI NATURAL SCIENCES, FIRST CLASS, UNIVERSITY OF BATH
September 2012 – June 2017
Physics Major, Chemistry Minor.

Modules:
Quantum Physics, Simulation Techniques, Mathematical Methods

Master's Project:
Computationally modelled the propagation of atrial action potentials using Python. The project was managed using Git and written up using LaTeX.

KEY SKILLS

Python Libraries:
Pandas, Numpy, SK Learn, NLTK, PyMongo and Plotly
Cloudera Hadoop version 5.4:
HDFS, Yarn, Sqoop, Flume, Avro, Parquet, Impala, Hive and Hue
NoSQL:
MongoDB, Neo4J
SQL:
SSIS, Lavastorm, Advanced SQL
Apache Spark:
RDD, DataFrames, SparkSQL, PySpark with some exposure to and understanding of Scala
Source Control and IDE:
Git, SourceTree and Git Extensions, PyCharm
Other Skills:
Datalku, Linux CentOS 6, SSH, Putty, Agile