FARAH SHAIR

Data Engineer

ABOUT ME

My love of entertainment, the arts and culture is the reason I found my way into data.

At university I spent a large amount of my free time involved in societies such as Musical Theatre and A Cappella. From singing in a large a cappella competition, to producing and directing a music video, to running publicity for musical theatre shows, I developed a love of the arts which led to my decision to work at Boom Ents, a marketing agency in the theatre industry.

In my marketing role, my passion for analysing, measuring and experimenting meant that I quickly took on the sales reporting and analysis for the company's projects. I implemented the use of Tableau in the company and remodelled the sales reporting processes from manually typing each number into excel, to downloading the files and visualising in excel templates or Tableau workbooks.

Having researched data engineering and analysis to greater depths and deciding to leave the marketing industry to join Kubrick, my interest has particularly grown for projects based on the entertainment industry, customer experience improvement, or marketing problems.

CONTACT US

Kubrick Group

T: 020 3866 4620

E: consultants@kubrickgroup.com

W: kubrickgroup.com

OVERVIEW

Studying a joint honours degree in Biochemistry with Management at Imperial was hugely beneficial, equipping me with both scientific and corporate skills.

I developed a variety of capabilities, such as: analysis and attention to detail when assessing results of practical experiments, high level maths involved in performing statistical tests, as well as communication and leadership in group work. I have gained experience of presenting in front of peers and professors, and gained commercial knowledge in my Management modules.

This combination is one of the main reasons I have been drawn to working in Data. This type of work is born from the fusion of applying scientific and mathematical concepts to a business problem. With my sales and marketing experience, and the technical skills learnt at Kubrick, I am able to draw upon a large toolbox of experiences when applying myself to these projects.

WORK

DATA ENGINEER - BMG - LONDON
Jan 2018 - Current

My time at BMG has allowed me to develop a comprehensive understanding of Tableau Desktop, as I have created many tailored reports for a range of business requirements. I have coded proof-of-concept specs for data manipulations for incoming data feeds in both python (using libraries such as pandas and numpy), and Hive/Impala SQL via Hue for Hadoop, that were then scaled and deployed into production by an out-sourced dev team. I have also written several python scripts to aid with automating some manual work, and for scraping data from websites requiring log ins.

My role involves listening to business users' requirements and designing the best analytics product to meet their needs, across all levels of the business from executives to interns and across territories around the world namely London, LA, Berlin and Sydney. I focused particularly on sales and marketing data, including social media analytics.

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.

MARKETING ASSISTANT – BOOM ENTS, LONDON
February 2017 - September 2017

I worked on the account management of a number of projects, as well as being responsible for the sales reporting and data analysis for all our projects. My responsibilities included executing high volumes of tasks with short deadlines, liaising with clients and holding meetings, and ensuring accuracy and detail in sales analysis for clients. I used primarily excel and Tableau to complete the reports for clients.

PRIVATE TUTOR - LONDON
April 2014 - July 2017

I taught Maths and Sciences to children ranging from 10 to 18 years old.

ASSOCIATE BUSINESS CONSULTANT – WOWCHER, LONDON
September 2016 - November 2016

Sales role involving pitching Wowcher's platform and database to businesses and negotiating deals in order to create offers for customers.

EDUCATION

IMPERIAL COLLEGE LONDON, BIOCHEMISTRY WITH MANAGEMENT, JOINT HONOURS BSC. 2:1
September 2016 - November 2016

The first two years of the course were solely in Biochemistry, exploring the details behind topics such as genes, proteins, enzymes, and macromolecular structure. This involved lectures, tutorials, and lab work.
Notable Modules: Biological Chemistry (1st), Cell Biology (2:1)
My third year was based in the Imperial College Business School, where we covered subjects encompassing many aspects of corporate business; from marketing and HR management, to innovation and strategy, to accounting and finance. The work in this part of my degree had a heavier focus on skills of teamwork, leadership and presentation.
Notable Modules: Finance and Finance Management (1st), Sustainable Business (1st), Global Business Management (1st), Accounting (2:1), Marketing (2:1), Entrepreneurship (2:1)

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: SQL:
MongoDB, Neo4J 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:
Dataiku, Linux CentOS 6, SSH, Putty, Agile