Samuel D'Souza

415 652 8253 samuelrohandsouza@gmail.com Kaggle DevPost LinkedIn

Education

2018-2021: IMPERIAL COLLEGE LONDON, U.K. – BSC. MEDICAL BIOSCIENCES

2012-2017: CHARTERHOUSE SCHOOL, U.K. – INTERNATIONAL BACCALAUREATE

Career

12/2021-PRESENT: LEAD DATA ENGINEER, CHAN ZUCKERBERG BIOHUB, SAN FRANCISCO

07/2019-04/2020: SOFTWARE ENGINEER, AUTOMAT, BERLIN (REMOTE)

03/2018-08/2018: RESEARCH AND DEVELOPMENT ENGINEER, I2X, BERLIN

09/2017-03/2018: RESEARCH AND DEVELOPMENT ENGINEER, LEONYTE BIOSYSTEMS, BERLIN

Publications

Lu, Tzu-Chiao, et al. "Aging Fly Cell Atlas Identifies Exhaustive Aging Features at Cellular Resolution." BioRxiv, Cold Spring Harbor Laboratory, 1 Jan. 2022, https://www.biorxiv.org/content/10.1101/2022.12.06.519355v1.

Skills

Programming languages: Python, JavaScript, C++, Nextflow, Java, Arduino, Rust
Web development: ReactJS, NextJS, NodeJS, HTML
Databases: Neo4J, MySQL, PostreSQL, JSON
Data interpretation and analysis: Machine learning, bioinformatics, data visualisation
DevOps: Docker, Singularity, AWS, Serverless, Git, HPC (on-premise deployments)
Lab Techniques: CRISPR, western blot, qPCR, cell culture, genotyping, phenotyping

Experience

LEAD DATA ENGINEER, CHAN ZUCKERBERG BIOHUB, SAN FRANCISCO (DECEMBER 2021 – PRESENT)

Lead and built a project called Datahub, which aims to structure data to support research. Biomedical research is complex, with highly heterogeneous metadata depending on the sample, modality, or goal. Datahub leverages graph databases to represent highly interconnected metadata. A website and API allows users to easily upload, track, and analyze ongoing research. To support Datahub, I also developed infrastructure and analysis pipelines for the mass spectrometry and genomics platforms. Communicated with leadership, and users, and coordinated software engineers to complete related software projects. Independently researched using graph theory to interrogate multiomic data being generated at the Biohub. Worked closely with the data science platform to analyse and interpret transcriptomics and spatial datasets.

BIOINFORMATICS AND DATA ANALYST INTERN, ILLUMINA, CAMBRIDGE (JULY 2021 – OCTOBER 2021)

Worked as a bioinformatician for Illumina's EMEA office, within their Population Genomics (PopGen) group. Contributed to methods for merging and saving terabytes of variant called files (VCFs).

Developed test software for triaging bugs and identifying algorithmic slowdowns. Worked for the DRAGEN platform in C++ and Python locally and in Illumina's HPC environment.

MACHINE LEARNING RESEARCHER, NATIONAL HEART AND LUNG INSTITUTE, LONDON (OCTOBER 2019 – MARCH 2020)

Developed machine learning models under the guidance of Professor Darrel Francis and Dr. James Howard in the Cardiology Department of Hammersmith Hospital, with the end goal of producing a CMRI (cardiac magnetic resonance imaging) pipeline for automated diagnostic of cardiovascular diseases. Developed a classifier that identified the 'view' of CMRIs, and retrospectively assessed model performance using several metrics. Investigated a potential novel transformer methodology for image data.

BIOINFORMATICS ENGINEER, GLOBAL GENE CORP, CAMBRIDGE (JUNE 2020 – OCTOBER 2020)

GGC collects bespoke exome data from underrepresented populations for investigations into disease genotypes, towards a 1 Million Genome Project. Developed an exome analysis pipeline in nextflow for variant calling and evaluated called variants using bioinformatics tools and Exomiser on a HPC. Also created a PostgreSQL database for storage of the Exomiser analysis, an API backend, and a ReactJS frontend. My project enabled geneticists to review and autogenerate reports on likely causative variants for the subject's disease phenotype.

RESEARCHER, IMPERIAL COLLEGE LONDON, LONDON (OCTOBER 2018 – MARCH 2019)

This project was an investigation into the effect of adiponectin on ß-Amyloid secretion in PDK1-Knockdown HEK293 cells, as a potential molecular target for treating Alzheimers. I worked with a group of four other researchers to develop knockdown cells with CRISPR, and evaluate the impact of adiponectin on the ß-Amyloid secretion pathway through protein and genomic expression.

SOFTWARE ENGINEER, AUTOMAT, BERLIN (REMOTE) (JULY 2019 – APRIL 2020)

Automat develops turnkey software solutions. Worked within Automat's 'Workshop Mode' and was responsible for the infrastructure behind the communication software pricing API and a web-scraping tool. Developed and deployed this all on AWS. Worked with Docker, SQL, and Serverless.

RESEARCHER, IMPERIAL COLLEGE LONDON, LONDON (OCTOBER 2019 – MARCH 2020)

Studied the inhibition of paclitaxel-induced apoptosis by epinephrine in a breast cancer cell line using cell viability assays, western blot, and qPCR alongside cell culture techniques.

RESEARCH AND DEVELOPMENT ENGINEER, I2X, BERLIN (MARCH 2018 – AUGUST 2018)

Researched and developed a hardware device for i2x's call assistance service. Developed audio streaming software in python and Arduino, to stream data to i2x's ML infrastructure, and visualise analytics.

RESEARCH AND DEVELOPMENT ENGINEER, LEONYTE BIOSYSTEMS, BERLIN (SEPTEMBER 2017 – MARCH 2018)

Leonyte was an early stage startup aiming to provide real-time testing for pathogens in food. Worked with the engineering team to develop a prototype portable bacterial detection device. This involved reading data sheets from National Instruments, writing summaries, and constructing several prototypes. Visualized biological and system data for presentation and troubleshooting. Smoothed signals and fine-tuned detection algorithms to improve pathogen detection. Helped with administration related to inventory, monthly invoicing, and SCRUM.

Online Courses

2018-2019

- Deep Learning Specialization, Deeplearning.ai (Coursera; 6 module course)
- Build a Modern Computer from First Principles: From Nand to Tetris, Hebrew University of Jerusalem (Coursera)
- Computational Neuroscience, University of Washington (Coursera)

2017-2018

- The Brain and Space, Duke University (Coursera)
- Philosophy and the Sciences: Introduction to the Philosophy of Cognitive Sciences, University of Edinburgh (Coursera)

Hackathons

ALGOTHON VIRTUAL @ ASPECT CAPITAL 2021

Completed several quantitive trading challenges using computational efficient algorithms and neural networks to predict stock movements in several different time series datasets.

IC HEALTHHACK 2020

Category runner-up. Developed a mental health app which uses machine learning to predict sentiment live, and a website to positively affirm positive statements to encourage self-affirmation. Project is on DevPost.

IC HEALTHHACK 2019

Built a wearable posture monitoring device to dynamically analyze posture and therefore reduce the risk of spinal injury. Project is on DevPost.

Extra-Curricular

2022-PRESENT: MUAY THAI (CMA), OPEN WATER SWIMMING (SOUTH END ROWING CLUB - IRISH POD)

2019-2020: IMPERIAL COLLEGE SWIM AND WATERPOLO (ICSWP) TREASURER

2018-2021: BUCS PREMIER LEAGUE WATERPOLO (IMPERIAL COLLEGE LONDON)

2017-2018: BERLIN AQUAHOLICS WATERPOLO (VORSPIELE)