Daniel Simpson

simpson.danielb@gmail.com London, NW4 1DJ +44 (0)7384 015 147

PROFILE

Experienced Data Scientist at a global retailer, specializing in leveraging advanced analytics to drive business insights and decision-making. Highly skilled at using DataBricks, Snowflake, Streamlit, and PowerBI to develop and implement a wide range of data-driven solutions, including predictive models (CLTV, Loyalty Churn, Omnichannel Propensity), interactive dashboards (Early Access, Customer Service) and NLP pipelines (Comments Ingestion, Store Summary). Extensive experience deploying and monitoring predictive models by developing MLOps dashboards to track and verify model outputs. With a background in scientific research and education in mathematics, statistics, and data science, I am passionate about utilizing data-driven solutions to drive commercial success within an organization.

KEY SKILLS

- Expert in Python, SQL, R, and MATLAB for data analysis, modeling, and automation.
- Strong hands-on experience using DataBricks, Snowflake, Streamlit and PowerBI.
- Additional experience with C++, JavaScript, HTML, CSS, Java, Git, and cloud platforms (GCP, Azure, AWS).

CAREER HISTORY

TJX Europe, London, *Data Scientist* **Jan 2023 to Present**

- Built and deployed a Customer Lifetime Value model using Databricks and Snowflake, enabling the business to optimize return on ad spend when integrated into Google Ads targeting.
- Developed and delivered an internal Early Access tool using Snowflake and Streamlit, automating a manual process, saving the team up to 30 hours per week, and saving the business £200k on outsourcing costs.
- Leveraged Snowflake, Databricks, PowerBI, and Streamlit to develop data solutions for customer strategy and business decision-making across various teams (Loyalty, Email, Ecomm, Customer Service).
- Planning and leading cross-departmental data-driven hackathons to develop innovative solutions to address large scale business problems.
- Partnering with business leaders (SVPs, AVPs) to champion data literacy by coordinating third-party training and delivering in-house sessions across the business.
- Piloting and evaluating emerging technologies (AI, LLMs) from external vendors to assess potential business value.
- Leading proof-of-concept initiatives to support innovation and R&D within the company.
- Mentoring and upskilling team members to support professional development and technical growth.

Decoded, London, *Senior Data Mentor & Product Developer* **May 2021 to Dec 2022**

- Designed and delivered data centric curriculum for apprenticeships and commercial training across finance and retail sectors.
- Coached professionals at all levels to apply data science techniques to real-world business challenges, guiding impactful project development.
- Led a learner helpdesk, providing technical support for work-based data projects and fostering applied learning.
- Built internal automation tools to streamline client communications and improve operational efficiency.

Bryant High School, Virginia, *Mathematics Teacher* **August 2016 to August 2019**

- Presented and explained mathematical and statistical concepts to diverse learners, consistently meeting district curriculum and assessment deadlines.
- Designed Python-based projects to teach applied mathematics and programming fundamentals.
- Collected, cleaned, and analyzed student performance data, delivering quarterly reports directly to the principal.
- Managed an instructional assistant to support classroom operations.
- Nominated for the 2018 *Outstanding New Teacher Award*.

Various Other Roles - *Teaching, Research & Hospitality* **May 2010 to August 2019**

- Taught math and science in international and domestic classrooms, developing lesson plans and leading extracurricular projects.
- Provided one-on-one and small group academic support, and helped manage classrooms with behavioral challenges.
- Conducted lab-based biological research and data analysis using Excel and MATLAB.
- Led hospitality teams at events, ensuring guest satisfaction and relaying feedback to management.

EDUCATION

Birkbeck, University of London: October 2020
Master of Science in Data Science - **Distinction**
Modules:

- Principles of Programming (Python)
- Big Data Analytics using R
- Computer Systems
- Data Science Techniques and Applications
- Information Systems
- Programming with Data (Python & SQL)
- Fundamentals of Computing
- Machine Learning

Dissertation: Deep learning techniques applied to time-series analysis for stock price predictions. LSTM neural networks were used for modeling, and evolutionary algorithms were used as an optimization technique.

West Virginia University: May 2013
Bachelor of Science in Mathematics
Recipient of The PROMISE Scholarship – merit-based financial aid providing full cost of tuition and fees.

PERSONAL DATA SCIENCE PROJECTS

Portfolio Website - https://danielbsimpson.github.io/

Covid-19 Dashboard

- Designed and created a dashboard app to track covid-19 within the United States.
- Worked with a university supervisor to collect data and calculate the r-rate within each county of the United States.
- Data preparation was done in Pandas and r-rate calculations were performed using the EpiEstim library within R.
- Dashboard created using Dash and Plotly, with the final product deployed using Google Cloud Platform.

Life Expectancy Inference from Global Metrics with OLS

- Taking over 37 different features from countries around the world, the OLS stats model was used to infer the main contributing factors for life expectancy globally.
- OLS summary statistics helped drive the process by identifying statistically insignificant features with high p-values while monitoring the r-squared value to ensure model performance.
- Multiple methods were explored such as including polynomial features and running Lasso and Ridge regressions while monitoring the residual space to evaluate performance of the various models.
- Statsmodels and sklearn were used for modelling, pandas and NumPy were used for data manipulation and matplotlib, seaborn and yellowbrick were used for visualisations.

Facial Recognition and Mask Detection

- Developed a deep neural network to identify whether individuals in an image are wearing a face mask or not.
- Data was sourced from a data repository on Kaggle containing over 4000 images.
- OpenCV was used for facial detection, utilising Haar Cascade and Caffe methods.
- Keras and ImageNet were used to build the convolutional neural network for face mask recognition.
- Matplotlib was used to display the new image containing labelled boxes around individual faces, identifying whether the individual is wearing a face covering or not.