Work Experience

Name: Gareth Edwards

Email: gareth.reeve50@gmail.com

Data Scientist _____

PEP Stores

May 2025-Present

- Led the successful migration of a core Customer Lifetime Value (CLV) segmentation model between Google Cloud Platform (GCP) environments, ensuring strict data integrity and operational continuity across business units.
- Architected a large-scale, complex ML feature store in BigQuery by aggregating the migrated CLVM data with disparate transactional, demographic, and marketing-interaction datasets to fuel advanced predictive analytics.
- Developed and tuned an XGBoost based predictive model to accurately forecast customer activity states and lifecycle transitions for the upcoming financial quarter, directly enabling targeted retention and growth strategies.
- Partnered with ML Engineers to operationalise the end-to-end pipeline on Vertex AI, leveraging a custom Python scaffold framework to ensure a highly modular, scalable, and tracked (MLflow) deployment.

Researcher/Data Scientist_____

Council for Scientific and Industrial Research (CSIR)

April 2022-2025

- Partnered with senior leadership and operational teams to create data products to aid decision making in various industries (politics, education, microbiology, urban planning, disaster management and crime prevention)
- Managed and developed the MS Fabric data pipeline and visualisation for the national election prediction project presented by the CSIR that visualized CSIR's proprietary Elections Prediction model on national broadcasts.
- Determined learner migration and learner population projections for all primary and high school learners in the Western Cape by integrating multiple data sources from the Western Cape Department of Education (WCED) using ARIMA time series models.
- Developed advanced predictive models leveraging machine learning techniques—including random forests, neural networks, and XGBoost—to enhance a building simulation framework for the City of Johannesburg. Integrated simulation outputs to generate forecasts of future municipal revenue, supporting data-driven urban planning and financial decision-making.
- Designed, developed, and deployed a cross-platform mobile application using Node.js and React Native to support crime monitoring and community-driven data collection. Successfully piloted the solution with the Mitchells Plain Neighborhood Watch, enabling more efficient incident reporting and improved local safety insights.

- Applied text summarization and topic modeling techniques to call center data, improving customer service analytics.

- Supported a disaster management project by developing an object detection model (using YOLO) capable of identifying human presence in drone imagery to enabling faster situational awareness and more effective emergency response.

- Applied multiple large language models—including OpenAI models, PEGASUS, T5, and BART—using Python to develop a proof-of-concept system for summarizing large volumes of text. The solution was designed to support analysis workflows related to state capture proceedings by improving the efficiency of document review.

Freelance

November 2023 - March 2025

- Wrote blog posts for a UK based company called Data Bear, outlining how to leverage the MS Fabric Lakehouse architecture with Power BI. (https://databear.com/microsoft-fabric-connecting-power-bi-to-fabric-lakehouse/). (Article's author name has been changed on article since I have left)

- I assist learners who are doing an AI bootcamp created by the US based company, edX. The assistance includes one on one tutor sessions, handling student questions regarding machine learning in Python and grading assignments.

- Partnered with senior leadership and operational teams to build dashboards in Power BI to highlight candidate recruitment pipelines and business KPI metrics for Loftus Bradford recruitment company based in Barcelona, Spain.

- Create and maintain a data-driven blog focused on football analytics, producing engaging content that explores insights, trends, and narratives at the intersection of sport and data. (https://Gareth1995.github.io/Footy-Analytics/)

Masters in Data Science Dissertation

University of Cape Town
January 2021–August 2022

- Developed Bayesian state-space time series models to track waterbird population sizes over time using JAGS and implemented in R.

- Developed bioindicators to track waterbird biodiversity over time in a pilot wetland site called Barberspan.

- Created visualisations for Time Series analysis and Index values using R.

R Programming Tutor

University of Cape Town
August 2021

- The primary task was to tutor students who were new to the R programming language over the duration of a 2-week crash course.

Computer Science Tutor

University of Cape Town

February 2017 – June 2017

- The primary task was to mark tests and assignments. Also held weekly meetings with the students for them to ask any questions about the work content.