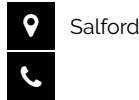


ANDREW FERGUSSON



Salford

@ aim_any.Oj@icloud.com

Lead Data Scientist

WHO AM I?

I'm a lead data scientist, sorting out metadata and championing linked data over data linkage. 10 years experience as a data analyst and engineer; recently completed a undergraduate degree in computing science and statistics.

EXPERIENCE

May 2022 -
current

Lead Data Scientist (Full-time)

Office for National Statistics

Promoting using open standards to speed up data dissemination across government is my jam. Engaging with other government departments and academics to adopt a cube-styled CSV-W where all the data exchange is provided using common vocabularies and ontologies with the data structure definition rolled in. I've been solving the thorny problems within ONS and outside by tweaking pre-post ELT and RAP activities to get data in a better and standardised shape. From this we get faster analysis, better portability, and consistency. Throw a CSV-W at a configured plotting tool and your y-axis comes pre-labelled. It's almost magic. By pushing open standards, I get to say I'm helping deliver the promise of the open web as envisioned by Tim Berners-Lee.

November 2020 -
April 2022

Senior Data Engineer (Full-time)

Office for National Statistics

As Senior Data Engineer in the Integrated Data Platform, I am responsible for transforming public data in various formats from government departments to high quality structured 5* Data. As part of a software engineering team, I work with other engineers using a developer stack of Python, Jenkins, Docker, and using common Python data libraries like Pandas, Numpy, Scipy, rdflib, and Databaker. In addition to transformations, I am programming a new OSS package to convert CSV to CSV-W called csvcubed.

- Developed new mechanism to build 5* Data from OData API
- Co-developed new python library to convert CSV to CSV-W (csvcubed)
- Engineered new processes to chunk large datasets to improve performance, be good API consumers
- Delivered peer upskilling sessions on DevOps practise, and pair-programming
- Managed junior staff, supervised experience contractors, and peer mentored to upskill

July 2019 -
October 2020

Data Engineer and Lead Analyst (Full-time)

One Manchester

Working within the Business Insights team, I was responsible for the roadmap of automating internal reporting from SQL Server Reporting Server/Excel to SQL/Power BI. Generated stakeholder buy-in to new process by creating tailored dashboard prototypes in Power BI using existing internal data sources and external sources under investigation for suitability. Championed Agile methodology for development of insight solutions, enabling bandwidth for valuable investigations of longstanding but not urgent questions.

- Developed new reporting requirements through conversations with data users
- Built Power BI dashboards using appropriate visualisations to convey a complete picture
- Transformed and shaped data from multiple sources to allow end-users and other analysts to build stable and accurate reports
- Investigated transactional and time series data for breaks, and periodicity using Python, Jupyter Notebooks, Pandas, SciKit, and Seaborn
- Developed risk model for accounts in good standing including factors such as in-person payment of rent due to COVID-19 lockdown
- Trained senior stakeholders on the use of Power BI on the web and on mobile to ensure that access to data never prevented a critical and timely decision
- Developed map of all reports generated detailing users, data sources, dependencies, and technologies prior to automation project
- Architected data warehouse to capture near real time insights on One Manchester's customers, transactions, and work

December 2013 -
June 2019

Lead Developer (Data Engineer) (Full-time)

BNY Mellon

Working within Risk Reporting and Analysis, but aligned to European ICAAP production, I owned the data solution for Economic Capital for two years. The responsibility included collection, cleansing, and analysis of credit, market, and operational risk data using various storage repositories and toolsets. I transitioned Excel-based data activities to SQL to improve repeatability, traceability, and accuracy of data in line with user needs.

- Gathered user stories around data and delivered a cohesive target data solutions
- Worked with users to collect feedback on ideas for improvement to overall solution set to feed into future releases and increase overall satisfaction
- Pulled together the user stories, ideas, and best practises to design a roadmap for data which delivered the needed product enhancement at the right time
- Developed multiple SQL, Python, and VBA toolsets to ingest new datasets, dynamically cleanse data at runtime and generate service health reports to meet end user needs
- Managed the collection of 200+ individual datasets collected on various schedules covering all Basel risk categories, data types, and formats for 7 legal entity consolidations
- Wrote, championed, and implemented five service level descriptions for the recipient ICAAP & Economic Capital with other BNY Mellon affiliates covering data receipt, cleansing activates, escalation processes, quality thresholds, and data definitions
- Produced several MI packs for the MI reporting teams within risk, covering EMEA Senior Management Risk Committee, and various stress testing forums covering operational risk metrics

EDUCATION

October 2015 -
September 2021
Grade: 2:1

Computing & IT (Computer Science with Statistics (Bachelors of Science))

Open University

PROJECTS

January 2021 -
current

csvcubed: turn CSV files into linked data

<https://gss-cogs.github.io/csvcubed-docs/external/>

csvcubed is a python library and CLI which helps data publishers disseminate their data sets using CSV-W and Linked Data.

CERTIFICATIONS

September 2022 -
No expiry

Coach in Running Fitness

England Athletics