

Justin Barton

justindatascience.com | me@justindatascience.com | 0434 857 186 | [Linkedin](#) | [GitHub](#)

SUMMARY

I am currently a Senior Data Engineer at ThoughtWorks with 10 years experience in: finance, technology and startups worldwide. I am interested in the engineering and science of manipulating data to generate a commercial uplift from hard-hitting predictions and insights.

EDUCATION

Springboard (www.springboard.com)

May – Aug 2016

- Data Science Intensive – Project learning with Harvard lectures and industry mentor

University of Adelaide, Australia

Feb 2004 - Nov 2008

- Bachelor of Science – Mathematical & Computer Sciences
- Bachelor of Engineering (honours) – Electrical & Electronic

Proficient with: Python, Pandas, SQL Server, Postgres, SAP HANA, MySQL, Hadoop, Tableau, Java, C#, SQL, VBA, Excel, Access, XML, JSON, jupyter, RedShift, S3, Azure Blob Storage, Azure Table Storage, AWS, git

Experience in using: scikit-learn, R, Spark, Athena, Glue, Airflow, Mathematica, Matlab, Flask, Adobe Analytics, Google Analytics, Javascript, HTML, Jinja, heroku, Jenkins

CAREER SUMMARY

ThoughtWorks - Industry leading agile software consultancy

Jul 2018 - current

- Senior Data Engineer (python, pandas, SQL, AWS, S3, Airflow, EMR, Spark, Redshift, Athena, Glue, Adobe Analytics)

AGL Energy - Helping to shape a sustainable energy future

Aug 2017 - Jul 2018

- Senior Data Engineer - Innovation (python, pandas, SQL Server, SAP HANA, Postgres, Tableau)

Marketplacer - Building marketplaces for e-Commerce companies

Feb 2017 - Jun 2017

- Data Scientist & Engineer (python, pandas, Postgres, MySQL, Tableau)

GrantTree (UK) - startup - R&D Funding for tech companies and corporates

Oct 2014 - Oct 2016

- Data Analyst (python, pandas, R, Visual Basic, Excel)
- Product and Software Lead (building software & new business for high volumes of clients)
- Mentor (leading training workshops, external presentations to create partnerships, sales)

Leyton (UK) - R&D Funding for tech companies and corporates

Dec 2013 - Oct 2014

- Research & Development Tax Consultant (technical & financial analysis for clients)
- Writer (technical reports for non-technical people)

ASC (Australia & Spain) - Submarines & ships – building & maintenance

Jan 2009 – Apr 2013

- Data and Software Engineer (SQL, Visual Basic, Access, Excel)
- Engineer (Project, Controls, Electrical, Logistics & Production Support)

PERSONAL PROJECTS

- Predicting the brownlow - Predicting the winner and 8 of the top 12 players
- Film Recommendation Engine - Finding life altering films for people
- Automating the financial modelling of investments with pandas and flask on AWS

CAREER DATA PROJECTS - HIGHLIGHTS

JetStar (ThoughtWorks)

- Extended the data pipeline (ETL) for 'Next Best Destination' machine learning prediction engine
 - Increased the number of customers in the model from 2 million to 6.5 million through experimentation in AWS with: S3, Redshift, Pandas and Jenkins.
 - Solved the difficult analytical problem of linking high volume Adobe Analytics web traffic data to core customer through various customer IDs from disparate systems.
 - Optimised algorithm to configure the level of parallel and series data processing in Redshift clusters to increase back-fill speed.
 - Migrated data platform from Jenkins scheduling to Airflow for both daily processing and back-filling.
- Built Redshift database to enable data exploration for web analytics
 - Built database and ETL from scratch with: ~ 1000 columns, 4 billion records to enable various functional teams to explore data for modelling that drive personalisation initiatives.
- Drove lean data engineering strategy
 - Through deep diving into the complexities of the data storage and processing of the data platform whilst understanding business needs, outlined and championed detailed strategy to reduce costs by keeping complexity to a minimum.
 - Teaching various commercial and technical teams the use of various cloud patterns, the requirements of a data platform and how we can achieve commercial goals.

AGL Energy

- Enabled unstructured customer IoT data to be usable for analytics and reporting in SQL Server
 - Built ETL (Extract, Transform, Load) for 1000s of homes' unstructured solar and load monitoring data (IoT) from Azure Blob and Table Storage, from an unknown and undocumented data model.
- Improved analysis capability for advanced analytics team
 - Built extensions on pandas in python to perform automated data manipulations including: filtering/mining/searching dataframes, counting value combinations, csv handling, split summary statistics.
 - Built object relational mapper (ORM) for extracting and mining data using python, whereby the ORM creates SQL at run time.

- Extended ORM to work in conjunction with: Hadoop, SQL Server, Postgres, SAP HANA and other classes derived from these, for domain specific applications using inheritance.
- Created data strategy and analytics for new energy plan partnered with Amazon
 - Defined data strategy and risks to create the contract between AGL and Amazon to launch AGL Skills (voice control) on Amazon Alexa.
 - Built propensity model to predict which customers would purchase the EDM Smart Home energy plan using data from Customer Analytic Record (feeding Pega) and Adobe Analytics.
 - Built automated daily reports for AGL Skills and Smart Home product suite, drawing in and manipulating disparate sources from: Azure Table Storage, SAP HANA and network folder csvs, running on an Azure virtual machine.
- Increased AGL's technical understanding of our virtual power plant through analysis
 - Conducted a study on spurious electricity grid disconnection issues on 300 battery customers to estimate through experimental modelling the value lost to each customer.
 - Analysed battery data to model and understand the value created during bulk battery control operations (orchestration) on 100s of homes in the AGL battery network.
 - Analysed raw battery data from disparate files to calculate real-life energy efficiency for new Tesla, LG and SolarEdge energy storage devices (battery & inverter)

Marketplacer

- Transformed 100s of tables of data across 15 e-Commerce sites to create insight.
 - Built ETL software running on AWS to manipulate 100s of tables of data across 15 e-Commerce sites to deliver insight reporting in Tableau.
 - Built extensions on pandas in python to perform automated data manipulations including: split summary statistics, SQL join analysis, dataframe transformation class, csv handling, splitting/merging/searching/comparing dataframes, comparing fields.
 - Built object relational mapper (ORM) for extracting and mining data using python, whereby the ORM created SQL at run time.

GrantTree (UK)

- Estimated cost per R&D claim to drive new product development
 - Analysed 5 years of R&D Tax claim data to build a new business case. Cost model for claims drove the decision to invest in a new business model. Developed new product and software for clients.
 - Built and iteratively tested an intuitive web application for clients to upload claim data, known as the portal. This was tested on roughly 30 clients, 7 paid. The information on client's usage of the application was tracked through google analytics.

ASC

- Built data model for 100s of megabytes of disparate control system data for large naval ships. (147m long, 7000 tonnes, 3 mission-critical networks, 19 servers, 1000s of pieces of remote controlled equipment, 10s of 1000s of digital & analogue signals.)

- Built ETL to clean and merge 33 differently formatted and structured cable & equipment databases from our supplier experimentally. The data models for each were unknown.
- Created a model of the control system in hierarchical form - for the interconnection between the: networks, servers, equipment and signals. This included finding all daisy-chained equipment, correcting anomalies and making chains searchable.
- Built search capability of all merged databases, and hierarchical information. Involved creating a model of advanced search criteria with a finite state machine to allow AND/OR expressions and wildcard searching.
- Created a GUI in Access, with bespoke VBA code enabling mouse and key events to create unique views and hierarchical filtering on screen.
- Automated programmatic testing of bespoke PLC source code creation software.
 - Created code to search by regular expressions, dozens of types of PLC code files to find the relevant sections to be checked against the database (that was used to generate code).
 - Created from scratch, a highly customised 'set' data-structure to enable fast checking of incongruences between PLC code against the database.
 - Created a flexible workflow system to: rename, move, and search files, and: display & correct data incongruences. The workflow system of roughly 10 stages allowed the user to have control for adding and removing the stages depending on design requirements.
 - Created a GUI to run the workflow and display the results of each stage. This GUI had a system of stage dependencies coded within, such that prerequisite stages were forced on the user when needed.