Ed Berry

DATA SCIENTIST

Sky Betting & Gaming, Leeds, UK

✉ eddjberry@gmail.com | ⌂ eddjberry.com | ⬛ eddjberry | ⬛ eddjberry | 🐦 ed_berry

Education

University of Leeds   *Leeds, UK*

PHD IN PSYCHOLOGICAL SCIENCES   *Oct. 2014 - Sept. 2017*

- Investigated attention and short-term memory in children using techniques from Experimental Psychology.

University of Cambridge   *Cambridge, UK*

BA (HONS) IN NATURAL SCIENCES (PSYCHOLOGY)   *Oct. 2011 - Jun. 2014*

- First two year in Philosophy with a second-year module in Experimental Psychology. I then studied Psychology full-time in my third year.

Experience

Sky Betting & Gaming   *Leeds, UK*

DATA SCIENTIST   *Apr. 2018 - Present*

- Lead the Identification working group for Safer Gambling. In this role I am required to meet external stakeholders, including regulators, as well as contribute to internal strategy.
- Rebuilt a key model to identify customers at risk of gambling-related harm using Apache Spark, delivering a 150% increase in recall performance.
- Led on a project to collect survey data to estimate characteristics of the customer base and build new models, collaborating with Research and CRM teams. Made use of advanced survey modelling techniques to model these data.
- Developed a methodology to combine important risk markers into a single score for use by operational teams.
- Led internal tutorials and workshops on R for data analysts, as well as providing 1-to-1 mentorship.
- Mentored junior team members in programming, machine learning, and data science.
- Use Agile development practices, such as Kanban and JIRA.

Leeds Institute for Data Analytics, University of Leeds   *Leeds, UK*

DATA SCIENTIST INTERN   *Oct. 2017 - Apr. 2018*

- 6-month project working with Sky Betting & Gaming to develop models to identify problem gamblers. As part of this role I worked with a statistician and academics from the Business School.
- Developed a methodology to estimate disposable income at a customer-level using open data.
- Further developed my skills in R, and learned Big Data technologies such as Spark and Impala.
- Presented the work carried out for this project at the Advances in Data Science 2018 conference in Manchester, UK.

School of Psychology, University of Leeds   *Leeds, UK*

POSTGRADUATE RESEARCHER   *Oct. 2014 - Sept. 2017*

- Designed and programmed novel experiments in Python.
- Analysed and visualised large datasets in R using generalised linear models.
- Organised data collection in primary schools, managing teams of undergraduates and research assistants.

Department of Psychology, University of Edinburgh   *Edinburgh, UK*

VISITING RESEARCHER   *Sept. 2016 - Oct. 2016*

- Won a competitive scholarship to do a 1-month study visit to the Department of Psychology.
- Programmed a novel experimental task in Python and designed analysis in R.
- Translated code to fit a mixture model to the data from Matlab into R.

Skills & Tools

Statistical Data Analysis & Machine Learning

ADVANCED (4+ YEARS)

- Multilevel/mixed effect models
- Survey analysis
- Bayesian statistics (Bayes Factors & estimation)
- Generalised linear models
- Machine learning, including regression, classification, decision trees, and autoencoder neural networks

Python

INTERMEDIATE (2+ YEARS)

- Wrote a battery of cognitive task in Python during my PhD.
- Used pyspark to process and model data in a production pipeline.

R
ADVANCED (4+ YEARS)

- Extensive experience using R intensively for data manipulation, visualisation, and analysis.
- Substantial experience interacting with Spark from R using both SparkR and sparklyr, having written multiple blogposts on the topic.
- Contributed to open source R packages as well as writing an R package for internal use in my organisation.
- Particular expertise with the tidyverse eco-system of packages on which I've led multiple tutorials.

Research
ADVANCED (4+ YEARS)

- Designed a number of experiments during my PhD, publishing results in peer-reviewed journals.
- Collaborated with internal Research teams at Sky Betting & Gaming to collect and model survey data.

SQL
INTERMEDIATE (1+ YEARS)

- Impala for ad-hoc queries and within production pipelines.

Testing
INTERMEDIATE (1+ YEARS)

- Unit testing R packages with testthat.
- Input validation of R functions with assertthat.
- Manual testing of data using SQL.

Version control
INTERMEDIATE (2+ YEARS)

- Git for versioning code and collaboration with branches, pull requests, and code review.
- Bitbucket and Github to manage code.

Visualisation
ADVANCED (4+ YEARS)

- Used Shiny to build an interactive visualisation of the results of an academic publication.
- Built dashboards using Shiny and flexdashboard to display business data.
- Used ggplot2 extensively to plot data in reports and academic publications. I have also run tutorials covering ggplot2 and mentored colleagues learning to use it.

Publications

JOURNAL ARTICLES

1. Atkinson, Berry, Waterman, Baddeley, Hitch, and Allen (2018). Are there multiple ways to direct attention in working memory? *Annals of the New York Academy of Sciences* **1424**(1), 115–126.
2. Berry, Waterman, Baddeley, Hitch, and Allen (2018). The limits of visual working memory in children: Exploring prioritization and recency effects with sequential presentation. *Developmental psychology* **54**(2), 240.
3. Berry, Allen, Waterman, and Logie (2019). The effect of a verbal concurrent task on visual precision in working memory. *Experimental psychology* **66**(1), 77.

BLOG POSTS

1. Berry (2017). Machine learning and k-fold cross validation with sparklyr. *https://www.eddjberry.com/post/2017-11-02-sparklyr/*.
2. Berry (2017). SparkR vs sparklyr for interacting with Spark from R. *https://www.eddjberry.com/post/2017-12-05-sparkr-vs-sparklyr/*.
3. Berry (2017). Writing your thesis with bookdown. *https://www.eddjberry.com/post/writing-your-thesis-with-bookdown/*.
4. Berry (2018). Cross-validation with sparklyr 2: Electric Boogaloo. *https://www.eddjberry.com/post/2018-11-23-sparklyr-cross-validation/*.
5. Berry (2018). Feature selection by cross validation with sparklyr. *https://www.eddjberry.com/post/2018-12-12-sparklyr-feature-selection/*.