

David John Baker, PhD

Data Scientist | Music Expert | Micro-SaaS Founder

davidjohnbaker1@gmail.com [Website](#) · [Github](#) · [LinkedIn](#)

Data Scientist with **five years** experience developing solutions for small-to-medium sized teams. Expertise in developing automated **micro-services**, statistical **modeling**, and **academic research** (music perception/computation). Experience working in **international** and **academic** contexts, collaborating on **multi-disciplinary** teams, producing **behavior changing** insights.

EXPERIENCE (and SKILLS)

Redgranite Group, LLC

June 2024– Present

Solo Founder and Director

Remote (USA)

- Achieving 95% email-open rate for outreach by developed software (**xml2**, **httr2**) visualizing (**ggplot2**) IRS 990 tax data
- Recruited 3 performing arts organizations with \$1.5 to \$3.5 million operating budget piloting software
- Wrote two R packages (**devtools**) to standardize data pipeline (**SQL**, **dplyr**, **stringr**) and organize UI (**golem**)
- Productionized micro-SaaS **Bolero** (**Shiny**, **Docker**, **AWS**) for non-profits to streamline common data-intensive workflows
- Developing LLM powered RAG chat UI (**Python**) hosted via **Posit Connect** for IRS XML 990 data for tax data query product
- Sold leads dataset of 990 data to marketing agency with scraped custom revenue, contractor, and employee information

Adventum

May 2024– August 2024, February 2025

Data Scientist, Contractor

Remote (USA)

- Reduced 2 hour analysis (**Excel**) to 10 seconds with dashboard (**Quarto**) for fractional CFO service
- Standardized data model/tables used for analysis post-Hubspot export (**data modeling**, **RDBMS**)
- Implemented automated churn and expansion workflow (**lubridate**, **dplyr**), safeguarding \$800,000 MRR calculations

Doran Scott Williams

September 2022– February, 2025

End Point Assessor, Contractor

Remote (UK)

- Evaluated over 80 Master's level assessments in accordance with [UK Government Standard](#) on machine learning topics
 - Administered live interviews on technical aspects (**regression**, **supervised ML**, **clustering**, **model evaluation**)
 - Critiqued communication of findings via both written reports and notebooks (**Python**)

University of Amsterdam

February 2022– March 2024

Postdoc

Amsterdam, The Netherlands

- Developed 3 R packages for ETL pipeline managing data (**git**, **JSON**, **SQL**) created by [Amsterdam Music Lab](#)
- Managed six graduate students across academic requirements of degree (Masters, PhD)
- Organized 40 scientists for international [Musicality and Genomics consortium](#), co-chaired Ethics workgroup
- Third author for article in *Cell* using [Beethoven's DNA](#) and contemporary musical phenotype
- Third author for paper and dataset in *2023 ISMIR Proceedings* on [on Eurovision Song Contest](#)

Goldsmiths, University of London

July 2020– January 2022

Postdoc, Department of Psychology/Computing

London, England

- Modeled semantic and memory effects of audio logos using multi-level models (**lme4**) with industry partner [Soundout](#)
- Second author on peer-reviewed [paper](#), won 2024 Douglas C. West Advertising Creative Article Award via [ICORIA](#)

Flatiron School x WeWork

August 2019– July 2020

Lead Instructor, Data Science

London, England

- Lectured daily on programming (**Python**), statistics (**hypothesis testing**, **AB testing**), and machine learning (**TensorFlow**) to 74 adult learners during 16-week course

EDUCATION

Louisiana State University

Baton Rouge, LA

Doctorate of Philosophy

Goldsmiths, University of London

London, England

Master of Science

Baldwin Wallace University

Berea, OH

Bachelor of Music