Data Scientist

Barcelona, Spain

david.munoztord@mailbox.org

david-munoztord.online

munozt0

david-munoz-tord

Skills

R	9 yrs.
Python	5 yrs.
JavaScript	3 yrs.
Bash	7 yr.
SQL	4 yrs.
Matlab	4 yr.
Data Analysis	9 years
Data Engineering	5 yrs.
Data Visualization	7 years

Libraries

- echarts, ggplot2, plotly, seaborn
- R Shiny, Dash, Streamlit
- tidyverse, data.table, dplyr
- TensorFlow, Keras, PyTorch, Scikit-learn
- Rstan, brms, JAGS
- H2O, XGBoost, LightGBM
- tidymodels, caret, mlr

Tools/Frameworks

- Git, Terminal
- Docker
- CI/CD
- AWS
- ML OPS
- ETL, NIFI

Summary

Results-driven Data Scientist and Engineer with proven experience in pharmaceutical research and financial technology. Expertise in developing scalable data management systems, implementing advanced predictive models using Bayesian methods, and deploying machine learning solutions. Skilled in Python, R, SQL, and AWS cloud infrastructure with a strong background in computational neuroscience and statistical analysis. Demonstrated success in customer behavior classification, fraud detection, and risk modeling.

Experience

Data Science Consultant — **10/2022 - Present**

Freelancer - Pharmaceutical & Tech Sectors

– Collaborated with pharmaceutical and biotech clients on data-driven research projects, leveraging computational modeling and advanced statistical techniques

– Developed data infrastructure solutions for multiple clients, implementing process automation and cloud-based data management systems

– Conducted complex data analysis and predictive modeling across diverse industry domains, including healthcare and financial technology

R Developer / Data Scientist — **04/2022 - 01/2024**

FlowBank

– Architected a machine learning-powered fraud detection system utilizing ensemble methods (neural networks and SVMs), achieving statistically significant reduction in false positive rates

– Developed advanced customer segmentation model using hierarchical Bayesian inference, generating probabilistic clustering with 92% accuracy

– Designed cloud-native data engineering infrastructure on AWS, implementing scalable microservices architecture for data orchestration

– Developed internal R packages to streamline data analysis workflows, enabling standardized data processing across teams

– Created comprehensive R Shiny dashboard for real-time performance monitoring, featuring interactive data visualization and dynamic reporting capabilities

– Implemented probabilistic Marketing Mix Modeling (MMM) using Bayesian statistical techniques to quantify multi-channel marketing effectiveness

– Constructed stochastic risk forecasting models with confidence intervals for Swiss banking regulators, integrating machine learning predictive techniques

Lecturer in Biotatistics — **09/2020 - 01/2022**

University of Geneva

– Conducted lectures on statistics for the graduate program and mentored students in their research projects, providing guidance on statistical analysis and data visualization.

– Created and shared freely available data workshops in R and Python.

Pharmaceutical Research Scientist — **03/2020 - 01/2022**

Novo Nordisk - Campus Biotech, Switzerland

– Pioneered computational neuroscience research on metabolic pharmaceuticals, focusing on GLP-1 receptor agonists' neurological and sensory mechanisms

– Developed advanced data processing pipeline for large-scale clinical trial datasets (terabytes of EEG/fMRI and biochemical data)

– Applied state-of-the-art computational techniques; Advanced Bayesian statistical modeling, Multivariate neuroimaging analysis, Machine learning-enhanced signal denoising

– Collaborated with international research teams (Caltech, EPFL) to implement cutting-edge neuroimaging denoising techniques [ICA Denoising Project]

– Secured research grants and maintained rigorous medical data management protocols

– Produced peer-reviewed publications demonstrating complex data analysis in pharmaceutical research contexts

Education

Ph.D. Candidacy in Computational Neuroscience

03/2020 - 01/2022
Campus Biotech - Switzerland
Topic: Computational modeling of addiction using neuroimaging data

MSc in Neuroscience

09/2018 - 03/2020
University of Geneva - Switzerland
Thesis: Denoising of high resolution fMRI signal using tICA

Complementary Studies in Computer Science

09/2017 - 06/2018
Smith College - USA
Capstone: Machine Learning for neuroimaging data

BSc in Cognitive Science

09/2014 - 06/2017
University of Geneva - Switzerland
Thesis: Early detection of ASD event related potentials

Languages

French	Native
Spanish	Native
English	Fluent
Japanese	Beginner

Technical Skills

- Data Sctuctures and Algorithms
- Data Mining
- Machine Learning
- Linux System Administration
- Visualization for Scientific Data
- Spatial Analysis
- Bayesian Statitics

Extracurricular

- Top 2% on StackOverflow (2024)
- Top 25% on GitHub (2023)
- Global Fellow scholarship (2017)
- Co-founder of Go-Fast, a socially responsible Bike Messenger cooperative in Geneva.

Projects

See my github profile for a more comprehensive list of open source projects.

Lead Developer — **EcoCalc**
R, SQL, Data Engineering, Environmental Analytics

– Developed a comprehensive platform for chemical product environmental impact assessment. Engineered scalable data infrastructure for ecological certification processes. Implemented advanced algorithms for environmental data analysis.

Co-President — **We Data**
R, Python, Education, Training

– We Data is data science organization focusing on knowledge sharing through blog, YouTube, hackatons and workshops. We also host R-Lunches at University of Geneva for community collaboration.

Creator — **DbVieweR**
R, SQL, Full-Stack Development

– Developed a comprehensive Shiny-based database management system [GitHub]. Implemented robust authentication and database manipulation features. Created flexible backend support for PostgreSQL and SQLite.

Collaborator — **hBayesDM**
STAN, R, Python, Neuroeconomics

– Contributed to hierarchical Bayesian modeling library for decision-making tasks [GitHub]. Implemented Q-learning algorithm for probabilistic selection tasks. Developed advanced statistical modeling techniques for neuroeconomics research.

Collaborator/Maintainer — **echarts4r**
R, JavaScript, Data Visualization

– Expanded interactive visualization capabilities in R [Project Website]. Developed high-performance rendering engine for complex data visualizations. Widely adopted in pharmaceutical and research sectors for data reporting.

Collaborator/Maintainer — **fiRebase**
R, JavaScript, Data Engineering

– Created comprehensive Firebase integration package for R [Project Website]. Implemented secure user authentication and file storage mechanisms. Enhanced Shiny application development capabilities.

Collaborator — **AFNI**
R

– Collaborated on the AFNI's functions for 3Dimensional Linear Mixed-Effects Regression (3dLMEr). Fixed residuals output image by adding bottom tolerance.

Publications

See my publications pages for a comprehensive list my scientific publications.

Professional References

Pierre Saouter ✉, **Head of Data Science at FlowBank** *Direct supervisor, can speak to technical skills and project delivery*

David Sander ✉, **Director of the Center for Affective Sciences** *Academic research mentor, validates computational neuroscience expertise*

Eva R. Pool ✉, **Computational Neuroscience Professor at Campus Biotech** *Collaborative research partner, understands advanced data analysis capabilities*