# Matthew Bain

### Data Scientist | Tableau, Databricks | Neuroscience Researcher Turned Al Developer

Data Scientist with 5+ years experience and certifications in Tableau and Databricks. My work spans neuroscience and assistive tech, with a focus on robust, interpretable ML. I help organizations understand their data and automate the hard stuff. I am seeking a role involving analytics and pipeline development.









## Professional Experience

Present lan 2024

## Open Source Developer

Independent

Toronto, ON

- Developed a software package [datopy] simplifying data model design and ETL workflows (Python, Pandas)
- Built a data management package [autocv] automating job application workflows (R, Excel)
- · Implemented a data app [zip-explorer] displaying geospatial KPIs in space and time (Streamlit, Plotly)

# Contact

**Ontario** 

**Personal** Website

GitHub

in LinkedIn

Oct 2023 Jun 2023

### **Data Scientist**

VocaLinks Inc.

Toronto, ON

- Created and communicated market breakdowns and technical tutorials (R Markdown, Python) on 4 Al-assisted assistive technology products, facilitating data-driven discussions among industry leaders
- · Distilled key statistical insights in hearing research into visually compelling reports (ggplot (R), Seaborn), driving a 100% increase in sales of an assistive listening product
- Designed the data layer of an Azure speech recognition data center, enabling rapid delivery of meaningful analytics to 200+ customers via the cloud

### Core tools

SQL	
Python	
Apache Spark	
Vertex AI	
R	

Aug 2022 Sep 2021

### **Research Analyst**

University of Western Ontario

London, ON

- Designed scalable online experiments (AWS, Amazon MTurk, JavaScript) to expedite data collection for 3 auditory neuroscience projects, yielding 2 publications and a robust cloud
- Provided multivariate statistical modeling support for research on neural signatures of epilepsy, laying the groundwork for automated labeling of 10+ hours of multimedia stimuli (Python, GCP)

# **Platforms**

TensorFlow
Tableau
GCP
Databricks
GitHub Actions

Aug 2021 Jul 2018

### **Computational Cognitive Neuroscientist**

University of Western Ontario

- London, ON
- Designed and deployed 11 experiments assessing neural speech processing in 300+ research subjects
- Developed a time series analysis toolbox (R, MATLAB) to measure hearing impairment through brain imaging, enhancing diagnostic sensitivity by a factor of 2

# Key competencies

Model Tuning
ML Pipelines
GenAl
CI/CD
Bayesian Statistics

## **Graduate Teaching Assistant**

University of Western Ontario

London, ON

• Directed 4 weekly research methods and statistics labs for 100+ undergraduate neuroscience students, leading to a nomination recognizing excellence in teaching and mentorship

Sep 2018

Apr 2021

### **Head of Finance** Apr 2018 Mississauga, ON Lorne Park Landscaping Apr 2015 Cofounded and managed payroll, client invoicing, and government remittances for a successful partnership specializing in property maintenance, servicing 80+ clients and leading a team of 5 employees Education BSc (Honours), Mathematics and Statistics Dec 2023 Hamilton, ON McMaster University Sep 2021 MSc, Neuroscience [Thesis] Aug 2020 London, ON University of Western Ontario Sep 2018 BSc (Honours), Neuroscience Apr 2018 Hamilton, ON McMaster University Sep 2014 Certifications Databricks Certified Data Engineer Associate [Certificate] Mar 2027 **Databricks** Mar 2025 Tableau Certified Data Analyst [Certificate] Feb 2027 Tableau Feb 2025 **Professional Machine Learning Engineer** Expected May 2025 Google Cloud Platform (GCP) Machine Learning Specialization [Certificate] Dec 2024 Stanford University Selected Projects datopy (Data Management Python Package) [Documentation] Present · Implemented and maintain a package for handling unstructured data, providing a simple Jan 2024 interface for data modeling, extraction, validation, and building ETL pipelines (Pandas, PyTest, Pydantic, GitHub Actions) mlvizz (Machine Learning Data Visualization Package) Present · Build Python interfaces for efficient, intuitive model selection, tuning, inspection, and ML Apr 2024 pipelines through modular, object-oriented designs with built-in data validation (Scikitlearn, TensorFlow, ArviZ, PyMC, SciPy) statvizz (Statistical Data Visualization Package) Present Build Python extensions unifying Pandas data summarization with Seaborn statistical Apr 2024 plotting functionality for a unified, intuitive, fully transparent statistical plotting interface (Pandas, Seaborn, SciPy, Matplotlib)

mathvizz (Mathematical Visualization Package)

Seaborn, Matplotlib)

 Build Python interfaces for exploring the geometry of machine learning math, including linear maps, derivatives, statistical distributions, and series approximations (SymPy, SciPy,

Present

Jul 2024



### One-stop Superstore Dashboard [Tableau]

• Designed an interactive dashboard displaying sales in a convenient, unified view, with key KPIs and sales logs visualized in both space and time (SQL, Excel)

### Dec 2024 | Nov 2024

### Zip Explorer (Geospatial Data Analysis Streamlit App) [Web App]

• Implemented a data application with interactive visualizations of geospatial data in space and time, across various KPIs and at state and county levels (Streamlit, Pandas, Plotly, Altair, GeoPandas)

### Jul 2024 | May 2024

### autocv (Automated CV-generation R Package) [Documentation]

• Built a data management package automating job application workflows using validated data stores and macro-integrated spreadsheets, resulting in an efficient, extensible workflow (dplyr (R), pkgdown, Excel)